ASSESSING MENTAL MODELS IN

MULTIDISCIPLINARY OPERATING ROOM TEAMS

Ivana Nakarada-Kordic

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Health Sciences, The University of Auckland, 2016.

ABSTRACT

The similarity of team members' mental models regarding clinical tasks is likely to influence teamwork effectiveness. There are currently a number of approaches to measuring similarity. However, they have not been applied in the complex environment of the operating room (OR), where professionals of different backgrounds must work together to achieve optimal outcomes for patients. This thesis had three objectives: 1) to develop a new empirical method for assessing the similarity of mental models in surgery, focusing on laparotomy; 2) to begin the process of validation of the new approach; and 3) to demonstrate how the new approach could be used in clinical practice.

The first objective was achieved by developing a software application (Momento) to sort key tasks in order to capture the information on mental models regarding task sequence and responsibility. Momento was developed through an iterative process including literature review, exploratory observation and expert opinion.

The second objective was achieved by examining the specific assumptions underlying the validity of the Momento approach. Twenty six-person OR teams, each comprising three subteams (anaesthesia, surgery and nursing) completed Momento prior to two simulated emergency laparotomies. Participants sorted 20 cards depicting key tasks, according to when in the procedure each task should be performed, and which subteam was primarily responsible for each task. The following assumptions were tested: a) similarity scores for mental models would be positively related to team familiarity scores or how familiar team members are with each other; b) similarity scores for mental models would be greater within OR subteams than between members of different subteams; c) different statistical measures used to calculate the similarity scores would yield similar results. The data provided support for all but the first validity assumption.

The third objective was achieved by separately analysing data for each key task. Differences were identified in team members' mental models for specific tasks for both responsibility and the order in which they should be performed. This may have implications for teamwork and patient safety.

The Momento approach could help elucidate and align the mental models of OR team members and potentially improve teamwork and patient outcomes.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors Jennifer Weller, Alan Merry, and Craig Webster for all their support and feedback throughout the thesis process. I appreciate their comments and the time taken to improve the quality of this thesis, and especially their patience in unpacking long, mind-bending sentences. They have taught me the value of clarity and conciseness that did not come naturally to a non-native English speaker trying to explain some quite difficult conceptual things. I believe my academic writing has improved as a result.

I was very fortunate to receive guidance in the early stages of my research from two renowned experts on teamwork and communication in healthcare teams: my advisor Tanja Manser and Lorelei Lingard. I am very grateful to my supervisors for making this collaboration possible. I also wish to thank Professor Chris Frampton for his statistical guidance, David Cumin for his invaluable help throughout various stages of my research, and Dejan Nakarada-Kordic for his patience with multiple iterations in the process of designing the Momento software application. I also thank Jan Hamon for proofreading the final version of my thesis.

I am very grateful to Health Workforce New Zealand for providing a stipend to enable my project. I am immensely grateful to the members of the MORSim research group for making my research possible and for making me feel part of the team: Jane Torrie, Jennifer Weller, Alan Merry, David Cumin, Ian Civil, Alexander Garden, Andrew MacCormick, Wai Leap Ng, Nishanthi Gurusinghe, Wendy Guthrie, Sharee Johnston, Brenda Knowles, Tracey Lee, Simon Mitchell, Michael Crossan, Abbey Gundesen, Craig Webster, and Matt Boyd. Having a background in psychology, I especially appreciated the input from the group on the medical matters relevant to my research.

Data collection would not have been possible without the clinicians who attended the MORSim study days. I am also very grateful to the staff working at the Auckland City Hospital level 8 operating rooms for their openness to my presence in their ORs during the exploratory in-theatre observation phase of the project. Special thanks are due to Wendy Guthrie and Sharee Johnston for making the in-theatre observations possible.

Most of all, I would like to thank all of my friends and family who supported me through this process. I dedicate this thesis to Mia and Max, who are yet to understand that mum is not aiming to become 'that kind of doctor' and to Dejan for his unwavering support, in sickness and in health.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Co-Authorship Form

Graduate Centre
The ClockTower – East Wing
22 Princes Street, Auckland
Phone: +64 9 373 7599 ext 81321
Fax: +64 9 373 7610
Email: postgraduate@auckland.ac.nz
www.postgrad.auckland.ac.nz

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Nakarada-Kordic I, Weller JM, Webster CS, Cumin D, Frampton C, Boyd M, and Merry AF. Assessing the similarity of mental models of operating room team members and implications for patient safety: a prospective, replicated study. BMC Medical Education. 2016; 16(1).

Content of this publication is included in Chapter 4 of this thesis.

Nature of contribution by PhD candidate	Conception and design, acquisition of data, analysis and interpretation of data, drafting and revision of the manuscript
Extent of contribution by PhD candidate (%)	80%

CO-AUTHORS

Name	Nature of Contribution
Jennifer M. Weller	Supervised the research process, made suggestions for change and improvement and reviewed and suggested changes to the manuscript.
Craig Webster	Supervised the research process, made suggestions for change and improvement and reviewed and suggested changes to the manuscript.
David Cumin	Provided suggestions for design, analysis and interpretation of data, and reviewed and suggested changes to the manuscript.
Christopher Frampton	Provided advice on the statistical analysis and interpretation of data and reviewed the manuscript.
Matt Boyd	Provided suggestions for design and analysis of data, and reviewed the manuscript.
Alan F. Merry	Supervised the research process, made suggestions for change and improvement, and reviewed and suggested changes to the manuscript.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- that the candidate wrote all or the majority of the text.

Name	Signature	Date
Alan F. Merry	[signature]	15/4/2016
Jennifer M. Weller	[signature]	15/4/2016
Craig Webster	[signature]	16/04/2016
David Cumin	[signature]	18 April 2016
Christopher Frampton	[signature]	17 April 17, 2016
Matt Boyd	[signature]	18/4/2016

SUMMARY

Statement of the problem

The increasing complexity of patient care and service delivery has meant that a multitude of health professionals and healthcare teams are now likely to be involved in the various stages of a patient's care. Teamwork is therefore central to patient safety. Recent research has revealed that inadequate or problematic interactions between health professionals, rather than a lack of clinical skills, are likely to influence performance effectiveness and failures in medical treatment[1-5]. The Quality in Australian Healthcare Study[6] found that 16.6% of all hospital admissions were associated with an adverse event that was caused as part of their medical treatment, which resulted in disability or a longer hospital stay. More than half of these were deemed preventable. The resulting disability was permanent in 13.7% of the adverse events, and in 4.9% of cases, the patient died. In New Zealand, 12.9% of screened inpatient records from 13 hospitals providing acute care revealed an adverse event which caused an average of 9.3 extra days of hospital stay[7, 8]. Similar findings have been reported elsewhere around the world[9-12]. Critical events and inadvertent injuries caused by or resulting from medical treatment are often related to breakdowns in the quality of teamwork, such as in communication and coordination[1, 13-16], particularly information loss[17], and difficulties in discussing errors[5]. The results of studies on the incidence of adverse events in hospitals show that the operating room (OR) is the most common hospital site where adverse events occur, often with serious consequences[18, 19]. It is a complex and dynamic setting, where poor information sharing and other aspects of teamwork have been found to negatively affect patient outcomes[20-22].

Healthcare teams, including the ones working in the OR, are typically multidisciplinary, characterised by interactions between professionals of different training backgrounds, expertise and experience, and demanding a high level of coordination within and between team members[23, 24]. Professional groups (for example, doctors and nurses), as well as specialty groups within professions, appear to favour different approaches to teamwork and these differences can create inter-professional barriers to communication[25]. As a consequence of their different educational and socialisation experiences, different professional disciplines are thought to have each adopted a different internalised set of basic concepts, approaches to their work, methods of enquiry, observational categories, and general ideas of what constitutes a discipline[26]. This internal representation of reality is also known as the person's "mental model"[27]. Within a team, each member will have a set of his or her own mental models. For example, each individual may have a mental model of the situation and the plan, a model of when various tasks should be done, and a model of who should be responsible for each task[28]. These individual preconceptions can get in the way of understanding the opposing discipline's point of view and hinder interdisciplinary work and communication. To be able to truly engage in interdisciplinary teamwork, team members need to share, at least to some degree, their mental models[26]. There is likely to be a core body of information that must be shared by all key players if teamwork is to be consistently effective[29]. The extent to which the component parts of mental models of individual members of a team overlap (like the common intersection of several circles on a Venn diagram) has been called the team's "shared mental model"[30, 31]. Shared mental models have been defined as team members' similar or shared understanding of relevant knowledge, such as knowledge about aspects of their common work or each other's tasks[30, 31].

The focus of this thesis was on shared mental models of multidisciplinary OR team members in relation to some of the team's key tasks and responsibilities related to surgery. Specifically, the aim was to capture the information on the extent to which members of OR teams agree on their key tasks in laparotomy – both in terms of the sequence in which they should be undertaken, and who should be primarily responsible for each task. Having compatible mental models means that all team members are ‘‘on the same page’’ with respect to how common tasks are to be performed, and are able to anticipate and predict each other’s needs[32]. This is thought to be the cognitive basis of the smooth and effortless coordination observed in many expert teams in high-intensity environments[32].

Theoretical models of teamwork such as the input-process-output (IPO) model describe shared mental models as prerequisites for successful teamwork. They are seen as driving team processes that lead to effective team performance, with more recent models depicting shared mental models as being shaped by team processes during and after performance[33-35]. An expanded version of the IPO model includes antecedents of shared mental models, such as characteristics of individual team members. However, these factors are often neglected when the relationships between shared mental models and the outcomes of teamwork are examined. In particular, there are contradictory findings on the relationship of one such antecedent and shared mental models, the extent to which team members are familiar with each other, and team performance. The relationship between team familiarity scores and scores indicative of shared mental models in OR teams was thus examined in this project.

The IPO model is applicable to the healthcare setting, including the OR. In the OR, the differing views of the situation and differences in understanding of key roles and responsibilities of different team members may be particularly detrimental during crises, when the time for planning a coordinated approach is limited[36]. Time pressures, ambiguous or incomplete clinical information, and changing team composition, can all add to the demands of each role. This can affect the OR team’s ability to adapt to unexpected challenges and may ultimately affect patient safety[5, 17]. Thus, it seems logical that teams in healthcare would benefit from sharing to some extent a mental model of the key tasks and of responsibilities. However, it is less clear just how similar the individual mental models of the OR team members should be, in order to yield optimal teamwork and performance. I propose that the degree of similarity of mental models in a team would depend on the type of team. For teams with less specialised roles whose every member can perform every other team member's tasks, such as in aviation teams, the team would benefit most from having a large overlap of their individual mental models. For teams such as OR teams, however, roles of individual professionals are mostly not interchangeable (for example, the roles of anaesthetists and of surgeons), and successful completion of the team's tasks would not be possible without relying to some degree on individuals' specialised knowledge and expertise. Consequently, we would not anticipate a complete overlap between mental models of individuals in a multidisciplinary team related to a particular aspect of their work. Because of their ability to coordinate their actions based on both overlapping and unique professional knowledge such multidisciplinary teams can cover a wider skill domain and complete more complex tasks than teams consisting of members with less specialised roles.

In addition, each team member will have multiple mental models regarding different aspects of their work. The Venn diagram analogy used previously in the context of shared knowledge in sports teams[35] can be used to illustrate the concept of a shared mental model of a specific aspect of a team's work. In Figure 1, each circle in the diagram would represent a mental model regarding that single aspect held by a team member. The area of overlap (or intersect area) between all the circles would represent the team's shared mental model about that aspect of the team's work. In teams with less specialised roles, we can expect a large overlap between the mental model of team members. For teams with highly specialised roles, however, the overlap between the mental model of each team member would be smaller.

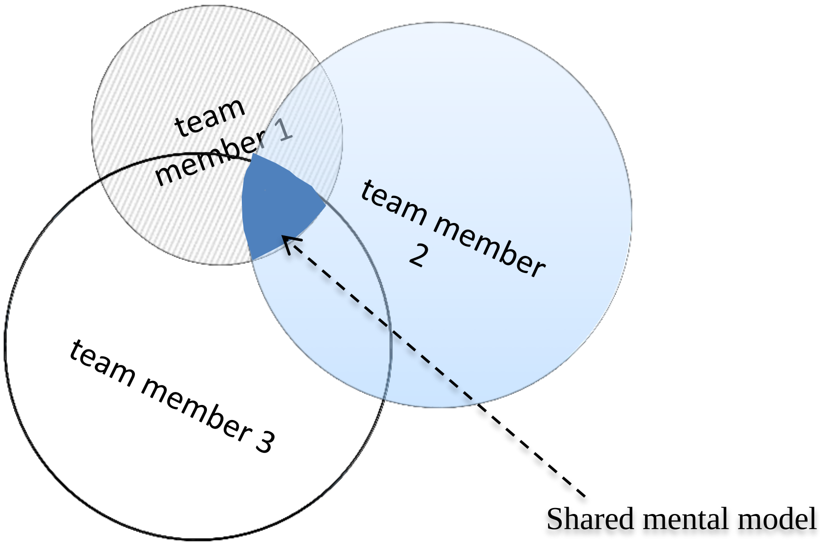


Figure 1. A conceptual diagram illustrating the concept of a team's shared mental model about a specific aspect of their work (see text for a full explanation).

Professionals in an OR team have different backgrounds and areas of expertise that shape each of their mental models representing a specific aspect of the team's work. These differences are represented in the diagram by the areas of each individual circle that do not overlap. The size of each circle will vary depending on the breadth of expertise of the professional that shapes the mental model the circle represents. For OR teams, it may be optimal that the mental models of individual experts regarding an aspect of the team's work be only partially shared, rather than identical, as tackling a highly complex task such as surgery would not be possible without relying on multidisciplinary expertise.

Empirical research in fields other than healthcare has confirmed the importance of shared mental models for the performance and processes of teams[37-40]. Research on shared mental models in healthcare teams, however, is scarce[41]. Most prior investigations outside of healthcare have been conducted within the military domain[42-45] or on team dyads, often comprised of students exposed to computer simulations of military missions[46, 47]. Studies in the OR tend to focus on individual professional groups, such as anaesthetists[48-50]. This lack of empirical research on mental models in healthcare has been attributed to the lack of measurement procedures appropriate for settings as complex and dynamic as, for example, the OR[41]. In the context of OR teams specifically, there has been calls for more research on shared mental models, to describe them systematically and to test the suggested link between shared mental models and surgical outcomes[51].

A technique to elicit and represent mental models is a central requirement for mental model research. The biggest hurdle to the progress of research on shared mental models has been the lack of validated measurement tools[32, 40, 52]. At present, it is possible to capture a representation of an individual's mental model, as for example, a series of steps a person understands are needed to complete a specific task, or a hierarchy of roles in a team as perceived by an individual. There is however no standard way to compare the individual outputs of several people – for example, the members of an OR team.

Contemporary conceptualisation of validity of measurement sees the process of validation as an ongoing matter, reliant on the accumulation of various types of evidence that would support validity. Consequently, another aim of this project was to begin the process of validation of the new tool for assessing mental models in teams.

Research on shared mental models could help in pinpointing areas of disagreement in teams, and in tailoring training strategies to promote shared understanding of the key steps in the processes that take place in the OR[36]. This would be expected to improve cooperation between members of inter-teams and thereby improve the teams' performance. This in turn may contribute to reducing error in surgery[53].

Therefore, the overall aim of the research presented in this thesis was to develop and validate a new approach to the assessment of shared mental models in multidisciplinary OR teams.

Research objectives, methodology, and key findings

This project was conducted in the context of the **M**ultidisciplinary **O**perating **R**oom **Sim**ulation (MORSim) study, a larger body of research aimed at examining various aspects of teamwork in the OR (Australia and New Zealand Clinical Trials Registry ID 12612001088831). As part of the MORSim project, 20 complete OR teams (comprising 120 healthcare professionals in total) from general surgical ORs at two large teaching hospitals in the Auckland region were recruited. Each team included: a consultant and a junior surgeon (surgical subteam), a consultant anaesthetist or senior anaesthetic fellow and an anaesthetic technician (anaesthetic subteam), and two nurses (nursing subteam). In each case these were team members who worked together in the same professional roles from time to time at local hospitals. Each team participated in a series of simulated emergency scenarios in a simulated operating room at the University of Auckland's Simulation Centre for Patient Safety. Two scenarios served as a vehicle for my project. Scenario 1 was a laparotomy for an abdominal stab wound and scenario 2 was a laparotomy for a perforated viscus. These were presented in random order to control for time of day and order effects.

My project had three specific objectives. These included:

1) Tool Development - developing a new empirical method for assessing shared mental models in surgery,

2) Tool Validation - validating the new method by examining the specific assumptions that would support it, and

3) Tool Application - demonstrating how the new assessment method can be used to support clinical practice.

Tool Development

The first objective was achieved through a step-by-step approach that led to the development of a task sorting tool for capturing individual mental models of OR team members of the sequence of key tasks related to the two clinical scenarios in the MORSim study. The tool was also designed to capture the information on who is primarily responsible for those key tasks. In addition, the development process involved an exploration of the different ways to calculate the degree of similarity of mental models of task sequence and responsibility for tasks in the OR team. The approach was informed by several studies[40, 54-56]. A number of tasks to be considered for possible inclusion in the task sort were identified based on exploratory observations of tasks, roles, and dynamics of OR teams during 10 relevant clinical cases. Subject matter experts (SMEs) selected tasks relevant to the two MORSim study scenarios. These were then narrowed down through an iterative process to a list of 21 tasks for each scenario. The findings of the exploratory observations were also used in the development of the layout of the form in which to sort tasks, and subsequently, the computer version of the task sort. Based on the feedback from the pilot of the computer version of the task sort using members of the MORSim research team and a group of OR clinicians, adjustments were made to the software to optimise its ease of use and functionality. The final version of the task sorting tool, named the Momento tool, was designed to be easily customised for other clinical scenarios. It consists of 18 generic tasks common to all laparotomies, and 2 scenario-specific tasks. When using the Momento tool, participants are provided with a summary of an upcoming clinical scenario and are then presented with the relevant set of electronic cards. They are asked to arrange these in the order in which they should be done and to identify the subteam (anaesthesia, surgery, or nursing) primarily responsible for ensuring that they are done. Participants can sort cards in parallel if they think some of the tasks need to be done at the same time. They can exclude a card, if they believe the associated task is not required for the successful management of the patient in the scenario.

As part of the tool development objective, I also explored the different ways in which the information on mental models held by individual team members captured by the Momento tool can be compared between team members to assess the extent to which mental models are shared within that team. At present, there is no consensus on how to calculate the degree of similarity of mental models within a team. Past researchers have used different ways to calculate the degree of similarity of mental models tailored specifically to their studies. In my thesis, I considered several well-established statistical measures of similarity and chose three to apply to the data gathered in the subsequent tool validation study. For calculating the degree of similarity of mental model of task sequence within OR teams I used: a) Euclidean distance; and b) Krippendorff's alpha for ordinal data. For the degree of similarity of mental model of responsibility for tasks, I used: a) percentage agreement; b) Fleiss' kappa; and c) Krippendorff's alpha for nominal data. I have named this new approach to exploring and mapping mental models about specific aspects of surgical operations "the Momento approach".

Tool Validation

The second objective was to begin the process of validation of the new Momento approach. This included testing the link between the similarity scores for mental model and team familiarity scores, and between similarity scores for mental model and subteam groupings. The assumption for the former was that OR teams whose members are more familiar with each other or have worked with each other more frequently in the past are likely to have more similar mental models. The assumption related to the latter was that team members from the same OR subteams (i.e., anaesthesia, surgical, and nursing) would have more similar mental models of task sequence and responsibility for tasks with teammates from their own subteam than with members of different subteams. The final assumption that would support validity was that different metrics used to calculate the similarity scores for mental models would reveal similar relationships between mental models and team familiarity, and professional groupings, if they are in fact capturing the same phenomena.

The validation study was conducted during MORSim study days. For each of the 20 participating six-member teams, participants were given written case briefings and time to read them before starting each scenario. They were then asked to use the Momento tool to sort the items related to that case.

The findings of the validation study provided partial support for concurrent validity of the Momento approach. While the link between team familiarity scores and the similarity scores for mental model was not demonstrated in this study, the findings suggest that overall, there may be more shared understanding of task order and responsibilities within professional subteams than between them, regardless of the metric used to calculate similarity.

Tool application

The third objective was to demonstrate how the new approach to assessment of mental models could be used in clinical practice. For this purpose, the data on mental models obtained in the validation study were analysed for each individual task in the surgical procedure rather than for the overall sequence of tasks in each of the MORSim surgical scenarios. The results of this analysis suggest that while members of multidisciplinary OR teams had a relatively good shared understanding of the order of tasks in the surgical procedure, there were substantial differences in their understanding of who should be primarily responsible for most crucial tasks in the surgical procedure. Members of the three OR subteams believed their own subteam was primarily responsible for around half the tasks in each procedure.

Implications

This research presents the first step towards a comprehensive and flexible approach to the assessment of mental models in multidisciplinary OR teams, with potential to be extended to other areas dependent on inter-professional teamwork. The findings of the differences in the mental models of OR team members have implications for effective team function and patient safety. If team members do not have a shared understanding of who should be primarily responsible for important clinical tasks during surgery, the result may be that the tasks are not done or that time is wasted with duplication of tasks. Both situations can potentially lead to compromised care, especially in a crisis where time and resources are limited.

The new approach to assessment of shared mental models can be used to identify dissimilarities in clinical practice. Used in this way the Momento has the potential to reduce harm to patients. The Momento approach is however more likely to be used as part of the team training process, where key surgical tasks with low level of agreement could serve to initiate discussion to ensure all members are “on the same page” as to the order of tasks and distribution of responsibilities prior to the team working together on the case. In addition, identifying dissimilarities of mental models in surgical teams could drive interventions at an institutional level to improve the degree of similarity of mental models within healthcare teams. Providing time for team members to agree, perhaps through a briefing session or in relation to the WHO Surgical Safety Checklist ahead of a procedure, on the order of crucial tasks and on who should be responsible for each task, may help to clarify potential ambiguities and better align mental models[36,37].

The new Momento approach could subsequently be used to check if the interventions have been effective at aligning the mental models for these crucial tasks within surgical teams. There is also potential for the Momento approach to be adapted for use in domains other than healthcare that are dependent on multidisciplinary teamwork. Finally, the Momento approach could also be used in research, to further test the proposal that shared mental models are a crucial dimension of effective teamwork in conjunction with measures of team process and outcome.

1. BACKGROUND

1.1. Teams and teamwork

A team has been defined as a "distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal, who have each been assigned specific roles or functions to perform, and who have a limited life-span membership" (p.4)[57]. Thus, not just any group of individuals constitutes a team and teamwork cannot be reduced to a simple aggregation of individual behaviours[58]. Teams differ from small groups through having common goals, intensive communication and coordination among team members, task-relevant knowledge and interdependencies, multiple sources of information, and adaptive strategies to respond to change (p. 1053)[58]. Apart from the technical skills required to do their job, team members must possess specific "non-technical" - or not vocation-specific - knowledge, skills and attitudes (KSAs) in order to work effectively together[59]. This can, for example, include the skill of monitoring each other's performance, knowledge of their own and teammates' responsibilities related to the task, and a positive disposition toward working in a team[60]. A team requires interdependency between its members. Typically, no single individual can accomplish the team's task alone, without working with teammates[59].

Various perspectives on teamwork have emerged since the first serious attempts to study team processes more than half a century ago. Common to the existing models of teamwork is an attempt to describe the factors that lead to effective outcomes[61]. Most of these models are based on the principles of the input-process-output (IPO) systematic framework, where, in its simplest form, team function is considered in terms of input, team processes, and team outputs, or outcomes of team processes[58, 61, 62]. Inputs are prerequisites of teamwork. They are conditions that exist prior to performance and facilitate or constrain team interaction.

These can be: individual characteristics, such as expertise and motivation; team characteristics, such as size and shared cognition (including shared mental models); and organisational characteristics, such as complexity and policy. Team processes describe how team inputs are transformed into team outputs - the results of team activity. Team processes include teamwork coordination and communication, while examples of team outputs include task achievement, error rate, and ultimately, a patient outcome in the case of healthcare teams.

The IPO model applies across different industries and has also been adopted for research on teamwork in healthcare[63]. In recent variations of the IPO model, the relationships between inputs, processes and outputs are more flexible and nonlinear, reflecting the complex and dynamic nature of teamwork. The nonlinear IPO model takes into account the effect of feedback loops where traditional outcomes, such as team performance, are considered as inputs shaping future team processes, or where two of the three components of the IPO model interact to influence the third[64]. An example of the former is the outcomes of team performance realigning the team's shared understanding of the roles and tasks of team members in order to improve the team's coordination.

Research in other high-risk domains such as aviation[65] and the nuclear industry[66] has long confirmed the importance of teamwork to improve safety and has identified shared mental models within a team as one of the key mechanisms for effective teamwork[30, 38]. However, despite the importance of teamwork in this field, so far, research in healthcare has mainly focused on relationships between processes and outcomes. Factors considered in the traditional IPO model as inputs, shared mental models in particular, have only recently become the subject of empirical investigation[41].

1.1.1. Teams and teamwork in healthcare

Teams in high-risk domains such as healthcare consist of individuals with complementary, interlinked roles, who conduct complex, time-limited tasks under changing conditions and in challenging environments[67]. To be successful, team members must coordinate their actions and work as an interdependent unit[68]. The OR is a good example of such a complex, high-risk environment. OR teams consist of professionals from different specialties, who are often assembled ad hoc to work interdependently on complex tasks and under changing conditions, where unexpected critical events can occur at any time.

Traditionally, teamwork skills have been emphasised less than medical knowledge and procedural skills in medical training[5]. Research on teamwork in the OR is relatively new. It began by adopting the approaches from the aviation industry in order to study the influence of non-technical skills in preventing adverse events[51, 69]. However, flying a commercial aircraft tends to be a more predictable process than treating a patient[69]. Successfully managing patients with their unique physical characteristics and ailments usually requires a personalised and unified approach for each patient and expertise from several health professionals of various backgrounds.

It follows that teamwork in the OR may be quite different to that in the cockpit. Both in healthcare and in aviation, predefined protocols or standard operating procedures are used to minimise uncertainties. However, due to the unique physiology of each patient and the potential risk of unforeseen complications during surgery, teams in the OR have to rely to a greater degree than pilots on the expertise of their individual members. In one study, for example, surgical teams appeared more reliant on individual interpretation of the situation and expectations of team members than on predefined protocols[70]. Consequently, the extent to which members of OR teams have a common understanding of a situation and its demands in the team may be lower than in aviation teams.

Studies have also shown that professionals within multidisciplinary healthcare teams often have differing perceptions of their teamwork. For instance, professional perspective was found to strongly influence interpretations of one's own role and the roles of other team members in the OR[71, 72]. Team members' understandings of other professions' roles, values and motivations were often in contrast with their views of themselves. For example, surgeons, nurses and anaesthetists in one of the studies independently rated their own profession as being less responsible for creating and resolving tension[71]. Other studies also revealed differences in the perceptions of professional groups of the quality of teamwork. Thus, nurses rated the quality of teamwork less favourably than doctors[22, 73-76], as did trainee doctors compared to senior doctors[77-79]. Entrants to undergraduate medicine, nursing and pharmacy programmes had different attitudes and beliefs as to whether clinical work should be the responsibility of individuals, or if there should be collective responsibility, even before they started their training[80]. Different perceptions and attitudes related to teamwork were also found among different disciplines involved in the resuscitation process in the emergency room[79], and among surgeons, anaesthetists, nurse anaesthetists and nurses in the OR[81]. The perceptions and attitudes individuals have about their teammates, their tasks, and their environment will influence individual behaviours and interpersonal interactions, as well as form a basis for their mental model of how a system functions[33].

In the healthcare context, having differing perceptions of teamwork can be problematic, as studies have shown that they can influence patient outcomes. For example, in the intensive care context, having a unified positive attitude to teamwork was found to be related to objective measures of quality and safety of patient care[82]. A high degree of openness of communication, mutual respect, and strongly shared goals among team members were related to reduced post-operative pain and improved post-operative functioning, and decreased length of hospital stay for patients undergoing elective surgery[16].

A recent review of teamwork[83] summarised the evidence supporting teamwork as critical to improving patient outcomes and staff wellbeing. This review also reported on the empirical evidence that system improvements (such as formal practices to strengthen communication among healthcare providers and specific team training interventions) have the potential to raise clinicians' awareness of these issues and to support effective teamwork. However, these factors are yet to be fully appreciated by healthcare organisations. The concepts that can help our understanding of effective coordination, such as team situation awareness and shared mental models, have rarely been studied systematically in this context[83]. In particular, having a shared understanding (or a shared mental model) of tasks and roles in the team enables team leadership, mutual performance monitoring, backup behaviour, adaptability, and team orientation. These have been identified as the "big five" teamwork skills necessary for effective team performance[29]. Methodological challenges to research on these factors, however, remain to be addressed[84].

My particular focus in teamwork research is on shared mental models. My research contributes to this work through the development and validation of a new approach to measuring the degree of similarity of mental models of task sequence and responsibility for tasks in OR teams in emergency surgery. The following section provides a detailed overview of the nature of mental models.

1.2. Mental models

This section describes how individual and shared mental models have been conceptualised in cognitive psychology and human factors literature.

1.2.1. **Definition and conceptualisation**

In cognitive psychology literature mental models were first conceptualised at the individual level of analysis[27]. Various terms in the literature refer to very similar concepts. Some of these include the person's "cognitive map"[26], "schema"[85], "script"[86], "frame"[87], "scene"[88], and "cultural model"[89, 90].

Mental models were envisioned as internal symbolic and dynamic representations of external events, or simulations of the world, representing a set of logical rules, that human beings manipulate in order to reason[91, 92]. The central idea behind the mental model construct is that its structure is analogous to the structure of an actual situation[93]. Humans use mental models as heuristics to describe, explain, and predict system behaviour, by rapidly retrieving the related information stored in memory[30].

The concept of mental models has been adopted in the field of human factors in an effort to explain human performance within complex systems. In this context, the notion of mental models has been applied to the design of systems (machines, procedures and tasks) that are compatible with operator mental models of how the system functions[94, 95]. Mental models enable individuals to describe what the system is for and what it looks like, how it works and what it is currently doing, as well as to form expectations about what the system is likely to do next[96]. A related practical implication is that people can be trained to develop and use specific mental models on the basis of given information, for example, to operate a device[97].

Cannon-Bowers and Salas[28] were the first to use mental models to explain differences in performance across teams. They summarised earlier characterisations of mental models in a team setting, and defined them as knowledge structures held by members of a team that enable them to form accurate explanations and expectations for the task, and in turn, to coordinate their actions and adapt their behavior to demands of the task and other team members (p. 228)[30].

The degree of overlap or the extent to which mental models are shared within a team can influence the team's performance[28, 30]. Such shared knowledge is seen to be constructed prior to team performance, through team members' previous interactions and experiences in a particular context and is regarded as relatively stable[98].

The terms "shared" and "team" mental models have been used alternately in literature to describe the same construct. Mohammed and Dumville[99] defined team mental models as team members' "shared, organized understanding and mental representation of knowledge about key elements of the team's relevant environment"(p. 90). Thus, people organise their knowledge of concepts, situations, objects, other people, and their environment, as well as the relationships between them, into structured patterns stored in their memory according to their most salient and important features[30, 100]. The degree to which these organised mental representations overlap or are common among team members represents their shared mental model.

1.2.1.1. *Mental models in OR teams*

Having a shared knowledge about "what should be done" as well as "who should do what" is essential for understanding the allocation of tasks and smooth collaboration within a team[101]. However, not having a clear understanding of one's own role and the roles of teammates is quite common in teams in general and may impair team coordination[51]. In the context of the OR, this may have consequences for patients. Unlike commercial aviation teams, where everyone in the cockpit can fly a plane and team members' roles are highly interchangeable, roles and responsibilities of OR team members are highly specialised, which enables them to undertake more complex and unpredictable tasks[102]. Thus, for example (and in general), an anaesthetist cannot perform an instrument count, a surgeon cannot perform an induction of anaesthesia, and a nurse cannot perform surgery. To some extent this is a matter of credentialing rather than necessarily a matter of expertise and ability – older generations of surgeons often did have some knowledge of anaesthesia, and conversely, and anyone could count a set of instruments. However, within the framework of processes and expected standards that apply within hospitals, staff are usually accredited to undertake the work of only one discipline, and indeed it is increasingly the case that individuals have very little expert knowledge about the other relevant disciplines, except to the extent that issues impact on their own responsibilities. An anaesthetist, for example, is very likely to know that applying pressure is one way a surgeon can temporarily stop uncontrolled bleeding, and in some circumstances may need to ask the surgeon to do this. Abdominal surgeons generally know enough about neuromuscular blockers used by anaesthetists to notice when these drugs are wearing off, and the abdomen is becoming too tight for the operation to continue, and will ask for this problem to be addressed. The key point though, is that a surgeon, an anaesthetist and a nurse could not usually simply change roles and expect to be able to complete the required tasks successfully and safely. As noted, this is in contrast to the pilot and co-pilot in the cockpit of an aeroplane. Moreover, OR teams work in a context that has more "failure modes" than an aircraft, where various factors such as a patient's unique physiology, the use of high-technology equipment and an impact of an invasive procedure on the patient have to be considered[102]. If teammates are not aware of some team members having a different perspective to their own on what needs to be done and when, there is a risk that those team members will be regarded as unpredictable by their teammates. Further, this may influence whether or not, when, and by whom the team's tasks get done. The expectations that team members have about what their tasks should be and what their team members' roles and responsibilities are will vary depending on the situation and the type of team. In highly interdependent and highly unpredictable situations, such as during surgery, not having a similar understanding can lead to miscommunication and poorer team performance[33]. In less specialised teams, it is reasonable to expect the team members to have a highly similar or even identical mental model of what needs to be done and a highly similar mental model of who should do what. In teams consisting of members with more specialised knowledge and roles, however, instead of requiring team members to have identical mental models, to successfully complete highly complex tasks that require interdependence, it would be optimal to have enough overlap between mental models of individual team members to allow them to form common expectations for their task and teamwork[30]. This would in turn enable team members to successfully monitor the situation and provide corrective input if things are not going as expected.

Inspired by a Venn diagram analogy previously proposed in the context of sports teams[35], the conceptual diagram in Figure 2 illustrates the concept of a team's shared mental model regarding one specific aspect of the team's work that can apply to OR teams. The figure depicts an example of a team consisting of three subteams of two members each: anaesthesia (A1 and A2), surgery (S1 and S2), and nursing (N1 and N2). Although each subteam or individual within the subteam can have multiple mental models of various aspects of their work, for the purpose of clarity, the diagram relates to a mental model that team members have regarding a single aspect of their work, for example, the sequence of the team's tasks in a laparotomy. Each circle thus represents a team member's mental model of the sequence of the team's tasks in a laparotomy. The mental model of each individual is influenced by his or her training and expertise. In addition, the mental model of each subteam is influenced by the domain knowledge of that subteam. Both encompass a large body of information that is needed to conduct a laparotomy. Individuals with more training and expertise in their field would be expected to have a larger circle that would correspond to the breadth of their mental model (such as the individual S1 in the surgical subteam) than those with less training and expertise in their field, for example, team member A2 in the anaesthesia subteam. Just like the common intersection of several circles on a Venn diagram, the extent to which mental models regarding the sequence of tasks in laparotomy of all individual members overlap represents that team's shared mental model of the task sequence. At the same time as sharing a mental model with all their teammates, an individual can have a mental model that is similar to a greater degree to a mental model held by some team members but not others. For example, the overlap of team member A1's and team member N1's circles represents the degree to which they share an understanding of the task sequence that is not shared to the same degree with team member N2. If we focus on subteams (identified by colour in Figure 2), for each subteam, the intersection of the two circles representing the mental model of each of the two members of that subteam (for example, A1 and A2) represents the shared understanding between those members. This shared understanding is based on the subject area unique to that particular subteam. If mental models are dependent on previous knowledge and training background, we would expect members of OR subteams that consist of professionals of similar backgrounds (for example, two surgeons) to have more compatible mental models than two professionals from different subteams, for example, nursing and anaesthesia. Unlike in teams where the roles of team members are not highly specialised, in teams of individuals with specific skill sets, such as the OR team, not having an identical mental model regarding an aspect of the team's work makes it more likely that the specialised knowledge necessary for undertaking complex and unpredictable tasks is used. Consequently, the common shared area on the diagram in Figure 2 will be bigger for the former type of teams, where individual circles will almost completely overlap, than for the latter ones, where roles of different professionals are more distinct from one another.

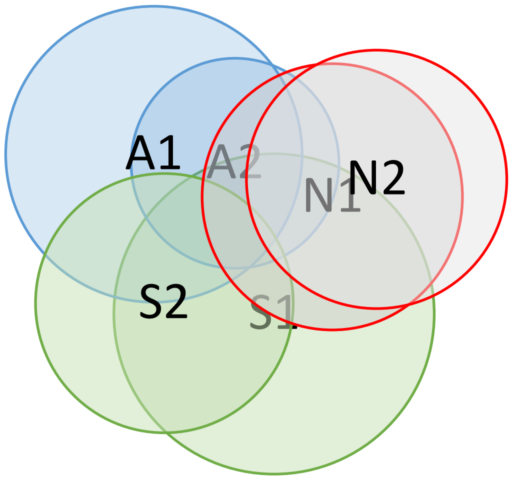


Figure 2. A conceptual diagram illustrating the concept of an OR team's shared mental model of one aspect of the team's work. In a six-member team consisting of three subteams (anaesthesia: A1 and A2; surgery: S1 and S2; and nursing: N1 and N2), each individual will have a mental model (represented with a circle) that he or she may share to a degree with one or more of his or her teammates (indicated by overlaps between circles) both within their subteam (for example, A1 and A2), and inter-team with members of the other two subteams (for example, A1 and N2). The area where all the circles intersect represents the shared mental model of the team.

Studying mental models in teams has practical implications for team training. It is likely to be impossible to train teams to have specific expectations for every possible situation that may arise. Instead, it is more useful to provide teams with strategies or tools that will enable them to quickly access their knowledge of the system in which the team will operate, so that they can form expectations appropriate to the current situation[30]. This is particularly the case in highly unpredictable situations and where team members' roles are highly specialised, such as in OR teams.

1.2.1.2. ***Types of mental models***

Cannon-Bowers, Salas and Converse[30] proposed four types of mental models, according to the things they represent, that may be shared within teams: 1) equipment - this, for example, may include an understanding of equipment functions, limitations of the system; 2) task – including understanding of procedures, actions and strategies to perform a task; 3) team interaction - or an understanding of how team members should interact within a given task, awareness of members' responsibilities, and role interdependencies; and 4) team members – including the knowledge of other team members' knowledge and skills. These four categories have been combined into two broad types of content to form taskwork mental models, which are a combination of equipment and task mental models, and teamwork mental models, which are a combination of team interaction and team member mental models[37, 47]. This type of aggregation facilitates the assessment of the extent to which multiple types of mental models are shared in a single study[47]. This is relevant to my thesis, where both taskwork and teamwork mental models within OR teams were captured simultaneously, using a newly developed tool designed for this purpose. These included a mental model of a task sequence, which is a type of task mental model, and a mental model of responsibility for tasks - a type of team interaction model.

Cooke et al.[103] conceptualised mental models as representing different types of knowledge. First, mental models can represent factual or declarative knowledge in that they can contain information about the relevant concepts and elements and relationships between them – or "what" needs to be done and by whom. Next, mental models can represent procedural knowledge or the knowledge of steps, sequences, procedures and actions involved in accomplishing a task. Lastly, mental models can also represent strategic knowledge, in that they can represent the information on the link between the team's capability to perform the task and the external environmental requirements. Strategic knowledge captures the situation component of mental models, or context, and it provides the basis for problem solving.

As part of my thesis, I aimed to develop a tool that would elicit the information on both the declarative and the procedural aspects of mental models in OR teams. In the case of the declarative knowledge, this involved capturing the information on the knowledge of which tasks need to be done in the surgical procedure and by whom. In the case of the procedural knowledge, the information on understanding of the sequence in which tasks should be performed in the surgical procedure was captured.

1.2.1.3. ***Properties of mental models – similarity and accuracy***

In addition to their content and form, mental models can also be classified according to their properties – similarity and accuracy. Similarity (also referred to in literature as sharedness[32, 47, 104], agreement, convergence, compatibility, commonality, consensus, consistency, and overlap[32]) represents the degree to which mental models regarding a specific content area converge among team members[30], represented by the common area on the Venn diagram depicted in Figure 1. Accuracy, also referred to as quality[37], is "the correctness of the knowledge structures maintained by team members" (p. 973)[40]. Accuracy is related to how correct a mental model is, compared to that of a subject matter expert[40]. Highly similar and highly accurate models are expected to have the greatest team performance benefits[37, 105]. However, shared mental models can be similar, but inaccurate, and vice versa. Thus, a team might have a similar but wrong view of a situation. The above definition of accuracy as "correctness" has been disputed. Mathieu et al.[37] argued that within a single content domain, there may exist a multitude of equally good, although different models. Thus, mental model accuracy may be difficult to evaluate in field studies as it may be hard to determine a single correct model. In the healthcare context in particular, there may be many approaches to patient care, and instead of choosing a single "correct" one, expertise will strongly constrain the selection of preferred models from the set of poorer ones.

Accuracy of mental models was not the focus of my project. However, I did address it to a degree by enlisting subject matter experts to agree, through an iterative process, on the criteria for the selection and the inclusion of the crucial tasks in the surgical procedure to be used in the task sorting tool. The issue of determining the degree of similarity of mental model was one of the goals of the present research and is discussed in detail in Chapter 2.

1.2.1.4. ***Shared mental models and models of teamwork***

Salas, Sims, and Burke[29] reviewed 138 models of teamwork from the team literature and proposed the "big five" factors that promote team effectiveness: team leadership, mutual performance monitoring, backup behaviour, adaptability, and team orientation. According to this model, shared mental models are the supporting and coordinating mechanism that blends together the value of each of the five factors for effective teamwork. In particular, Salas et al. proposed that mutual performance monitoring, backup behaviour, and adaptability occur more often in teams where mental models related to the team's work are shared. Highly effective teams share a mental model of the task environment. This includes a shared understanding of how the equipment functions and limitations of the system in which the task is performed. They also share a mental model of the task itself, and a mental model of the tasks and abilities of interacting team members. This helps them to adapt to non-routine situations and predict the behaviours of other team members[65, 106]. In the OR specifically, due to their interdependent nature, team members are required to have a high level of shared understanding of each other's roles, tasks and goals of the procedure throughout the surgical process (p. 182)[51]. The dynamics of the OR is such that team members "must be aware at all times who is in charge of the task underway and how their individual role fits into the group effort….each member must understand the other team members' level of competency, style of working, and knowledge of the current task" (p. 1008)[107]. Evaluating the degree of shared understanding within surgical teams is important for identifying problems in procedures and in team dynamics that may negatively impact on team performance[51].

Theoretical models of teamwork such as the input-process-output (IPO) model describe shared mental models as prerequisites for team processes that lead to effective team performance, with more recent models depicting shared mental models as being shaped by team processes during and post-performance[33-35]. For example, having a similar mental model about the task, the environment, and about each other is thought to be a prerequisite for implicit coordination within teams, characteristic of highly effective teams. This type of coordination takes place when team members anticipate the actions and needs of their colleagues and task demands, and dynamically adjust their own behaviour accordingly, without having to communicate directly with each other or plan the activity[108, 109]. Implicit coordination typically occurs during non-routine situations characterised by high levels of uncertainty and time pressure that might require but not allow for explicitness[110]. It develops over time and is less time-consuming and labour intensive than explicit coordination which is expressed in an unequivocal manner that is usually plain and easy to understand[109, 111]. Unlike explicit coordination, implicit coordination relies more heavily on the pre-existing common understanding of the situation. It can involve members communicating verbally as a result of anticipation of other team members' needs rather than in response to a request[112]. This can, for example, involve asking for, or giving an update, or verbally offering unsolicited assistance. Implicit coordination has been reported anecdotally in the context of the OR environment, where surgical team members commented that while working in an established team they often felt as if they were "reading each other's mind" (p. 88)[36]. Mutual performance monitoring, backup behaviour, adaptability, and unsolicited information sharing have been observed in the OR setting as manifestations of implicit coordination[3, 23, 48, 56, 113, 114].

The IPO model of teamwork can be expanded to consider the antecedents to shared mental models to further clarify their relationship with processes and outcomes. These antecedents include some individual characteristics of team members, prior team interventions, and contextual factors[32]. Antecedents of shared mental models have not been studied as extensively as team processes and outputs.

In addition to the view of team coordination being reliant on shared mental models, and consistent with the nonlinear IPO model of teamwork, some authors also conceptualised the reverse in an attempt to explain how shared understanding develops within teams. Eccles and Tenenbaum[35], for example, conceptualised the development of shared understanding as a vital component of team coordination that develops through team processes. The pre-process phase is characterised by activities that are thought to facilitate shared understanding through information sharing[31, 33]. These include setting collective goals, establishing roles and norms, and developing collaborative planning. In-process shared understanding takes place during performance of the task, along with verbal and nonverbal communication that leads to coordinated actions[35]. Shared understanding can develop in the post-process phase through verbal reflection after the task has been completed, for example, through debriefing, discussion, and experience sharing[34, 35], so that conclusions regarding performance can be drawn, and self-corrected in future performances[33].

Figure 3 depicts the examples of factors that have traditionally been considered as inputs, including the ones considered to also be prerequisites for shared mental models, processes and outputs relevant to teamwork in the healthcare setting.

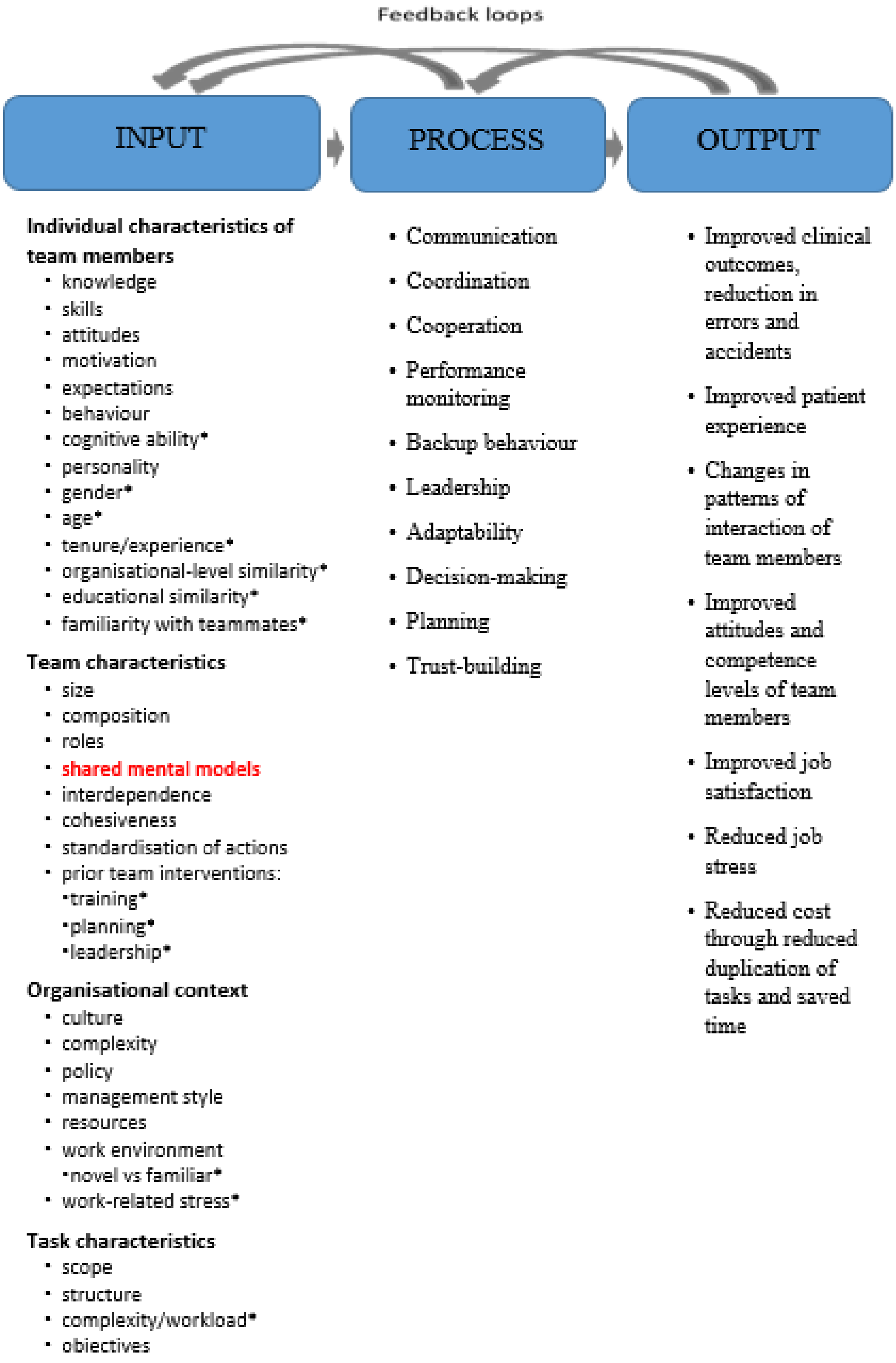


Figure 3. A schematic representation of the nonlinear input-process-output (IPO) model of teamwork, with some examples of what is traditionally considered as individual inputs, processes and outputs applicable in the healthcare setting (adapted from Mohammed et al.[32], Rosen et al.[115], Ilgen et al.[64], Rousseau, Aubé, and Savoie[61], and Healey et al.[116]). The inputs denoted with an asterisk have been studied empirically as antecedents of shared mental models in contexts other than healthcare[32].

Researchers suggest that prior experiences in team settings shape individual's expectations about future teamwork that are then generalised to similar settings[30, 117]. Examples would include how to behave in a team setting and which behaviours are important for team success. Unlike short-lived teams, teams who remain intact for a prolonged period of time, such as several months or years, have the opportunity to develop a high degree of shared understanding of their work through repeated interaction, communication and discussion[118, 119]. Past experiences in a team or the degree to which team members are familiar with each other should help develop a degree of shared expectations of teammate-specific roles and responsibilities[120], and a common frame of reference that can be used during task completion[121]. Having positive team experiences, such as good coordination, communication and team building that are shared by cohesive teams should strengthen one's expectations for teamwork[122]. Thus, it is expected that the longer team members work together, the greater the extent of their shared knowledge[109, 119].

In this thesis, I explored the relationship between team familiarity scores and the scores for the degree of similarity of mental models as a step towards validation of the new approach to the assessment of shared mental models. Familiarity with teammates was determined by asking each member of an OR team to indicate on a 4-point scale how often they worked with each of their teammates in the past. The familiarity score for two team members was calculated as the mean of their ratings of each other. The team familiarity score was an average of all possible pairwise ratings within the OR team.

In summary, mental models are mental representations of the outside world that help individuals make sense of it. Shared mental models have been conceptualised as a degree of overlap between individual mental models within a team. In teams with interchangeable roles, and in highly predictable situations, having a greater overlap between mental models of individual team members may be better for teamwork and team performance. In teams consisting of subteams of experts with non-interchangeable roles, such as in OR teams, it is not necessary for members to have identical mental models, but only compatible ones. The degree of sharing of individual mental models in such teams can be improved through training. A broad classification of mental models would include taskwork and teamwork mental models. Both types were of interest to the present research. A distinction is also made between two properties of shared mental models – similarity and accuracy of the mental model. Models of teamwork link the degree of similarity and accuracy of mental models within a team with how successful teams are in coordinating their activities and performing effectively. The focus of this thesis was on capturing the degree of similarity of mental models within multidisciplinary OR teams. “Similarity” and “sharedness” are used interchangeably throughout the text. Accuracy was outside of the scope of this project. Prior experiences of team members, such as how familiar team members are with each other, are seen as variables contributing to sharedness of mental models. The relationship between the extent to which team members had prior experience with working with each other (captured here via familiarity scores) and the degree to which they shared their mental models was also examined in the present research. The next section describes the empirical evidence for the relationship between shared mental models and related variables.

1.2.2. Empirical evidence of shared mental models, their antecedents and outcomes

According to the traditional IPO model of teamwork, shared mental models are inputs that facilitate team processes and outcomes. Several comprehensive literature reviews summarise the recent empirical evidence for the link between shared mental models and team processes and performance in different settings[31, 32, 39, 41, 123]. Shared mental models have been found to contribute to improved team performance in many studies[40, 47, 54, 56, 105, 124, 125]. Similarity of mental models has also been related to improved team processes[37, 47] such as coordination[126], communication[40, 66], backup behaviour quality and quantity[126], team monitoring in anaesthesia teams[56], collective efficacy[127], strategy implementation[128], engagement[129], team viability and member growth[130]. In non-routine situations, focusing less on the time to complete a task was found to lead to more face-to-face communication that enabled information sharing necessary to form or update a shared mental model[66]. A positive relationship was also demonstrated between similarity of mental models and skill acquisition[131], and decision quality[132]. Several studies have shown that the relationship between shared mental models and performance can be mediated by team processes. When increased sharing of mental models occurs within a team, team members' interactions are more effective, leading to better performance[37, 47, 126, 128]. The apparent relationship between the degree to which mental models are shared and team processes and performance suggests that evaluations of this degree of sharing of mental models could be used in supporting training needs analysis and training evaluation[125]. Specifically, the former could involve administering a tool to individual team members that would capture the information on their mental model related to a specific aspect of the team's work, in order to identify the potential differences within the team. This information could then be used to tailor the follow-up training to improve the degree of similarity of their mental models accordingly. Such practical implications are discussed in Chapters 4 and 5 of this thesis.

In the relatively small amount of work conducted on the antecedents of shared mental models, researchers have focused on stress[54], planning[38], rank and length of time in service[133], demography, team experience, team member recruitment, team size[130], and team training[112, 126, 134].

The empirical evidence regarding the relationship between the extent to which team members are familiar with each other and other team-related variables is conflicting. Studies conducted on coal mine workers found better average production and fewer accidents within teams with members who are familiar with each other due to a decrease in poor coordination of teamwork[135]. In studies involving information technology teams, having more similar mental models have been found to have a positive impact on team coordination among members of software design teams. Prior experience of team members with the same software and projects had a positive impact on project performance, especially in geographically distributed teams where their familiarity compensated for problems in communication and coordination[136]. In sports teams, researchers have found that the more prior experience players had with their teammates in tennis doubles teams, the more similar their knowledge was regarding their team's responses to match situations[120]. This in turn led to the team engaging in more implicit coordination. Groups of friends were found to be better than groups of strangers at applying conflict management strategies to their current task[137]. A longitudinal study showed that teams consisting of friends and acquaintances perform significantly better than teams composed of strangers, but that the effect of the extent to which team members were familiar with each other is most beneficial during the team formation stage and decreases over time[138]. For example, team-level cognitive ability and openness to experience were found to be positively related to the similarity of a task-focused mental model, regardless of the extent to which team members were familiar with each other, which in turn was positively related to the perception of team members of how coordinated the team's activities were[139]. In research in the commercial sector analysing company performance, no significant relationship was found between the extent to which team members were familiar with each other and their company's performance[140]. Some findings in this field suggest the extent to which team members were familiar with each other may even negatively impact on performance. The negative impact on performance was thought to be due to team members who were more familiar with each other dedicating more time to social interactions, which diverted their energy from team goals[141]. It was also thought to be due to low diversity of information, where all team members possess identical or common information – which can lead to groupthink[142].

In summary, there is now compelling empirical evidence showing the positive link between the degree of similarity of mental models in a team and the quality of team processes and performance. However, there are still contradictory findings on the relationship between the extent to which team members are familiar with each other and the degree of similarity of mental models in different domains. It is possible that this is partly due to the research previously not distinguishing between teams consisting of members whose roles are not highly specialised and teams of specialists, as is the case of OR teams. There is also the absence of similar studies in the healthcare context. Thus, the latter relationship was explored in this thesis in multidisciplinary OR teams. The theoretical assumption that past experience with teammates or how familiar team members are with each other is positively related to the degree of similarity of mental models was tested as a step towards validation of the proposed approach for the assessment of mental models in surgery.

The next section addresses assessment issues in mental model research and provides an overview of existing methods for the assessment of mental models.

1.3. Assessing mental models

This section outlines the existing approaches to the assessment of mental models in teams. Card sorting and concept mapping approaches relevant to this research are discussed in detail, including the empirical studies that have used a combination of the two approaches.

Mental models in teams have been studied in many contexts, using various assessment approaches. Previous researchers have not only used different methods of eliciting, scoring, and representing shared mental model data, but they have also used different conceptual definitions[125]. The diversity, complexity, and number of methods used in studies published in peer-reviewed journals attests to the difficulty of assessment[32, 52, 134]. There is presently no evidence to support a single method as superior to others in its theoretical value, reliability, and validity[32, 143]. Although researchers have emphasised that choosing the correct technique to assess knowledge for a task is essential[32, 39], as well as using it correctly, there is surprisingly little practical guidance on these matters[144].

Several literature reviews and meta-analyses[32, 39, 41, 52, 143, 145] summarise the breadth of approaches used to capture mental models and assess the degree to which they are shared. What seems to be common to the broad range of techniques for assessing the similarity of mental models is elicitation of individual team members' mental models of a specific component relevant for the team, followed by some sort of aggregation analysis of individual scores to show a level of sharedness[146].

In order to gain a comprehensive insight into the nature of a mental model, both content and structure of a mental model should be investigated[52]. Content refers to the concepts the mental model represents, such as task, equipment, or team members' roles. Structure of a mental model refers to how concepts are represented in the individual's mind, for example, the sequence in which tasks in a procedure should be performed. A recent meta-analysis of methods for assessing mental models in teams[39] found that when the selected method did not capture the structure of the mental model, there was no observed relationship between the degree of sharedness of mental models and team process. Mohammed et al.[145] distinguish between elicitation techniques that capture the content of a mental model and representation techniques that reveal its structure. The final step in the assessment of shared mental models involves examining the relationships between concepts in a whole team. Here, it is the level of similarity of models within the team that is of interest[39]. Techniques that capture both content and structure of mental models include paired comparisons ratings, qualitative techniques, concept mapping, and card sorting. These are discussed in the context of finding the most suitable method for revealing the information on the content and structure of mental models in OR teams recorded in this project.

1.3.1. Paired comparisons ratings

Also known as similarity ratings, or relatedness judgements, paired comparisons ratings are the most commonly used technique, while qualitative methods are the least popular[32]. The former involves the participants rating the similarity between pairs of task or team-related concepts or statements supplied by the researcher, usually on a Likert-type scale. In order to determine how these concepts are organised in the shared part of the mental model of the individual members of the team, a quantitative index of similarity, or overlap of the team members' models, is then calculated using computerised network analysis algorithms such as coefficients S in Multidimensional Scaling (MDS), Pathfinder's C or UCINET's QAP correlation[32, 39]. A study by Lim and Klein[124] provides a typical example of how the paired comparisons technique can be used to calculate the degree to which individual mental models are shared in a team. These researchers asked individual members of military combat teams to judge the relatedness of pairs of different combinations of 14 statements describing team procedures, equipment and tasks (to capture their mental model of taskwork), and 14 statements describing team interaction processes and the characteristics of team members (to capture their mental model of teamwork). Team members rated the relatedness of all pairs of statements (91 pairs) using a 7-point response scale (1= unrelated to 7 = highly related). Examples making up these 14 statements were "Team members conduct routine maintenance of their equipment and weapons in the field", "The team is highly effective", and "Team members trust each other". The statements were chosen in consultation with subject matter experts. Pathfinder, a structural assessment technique that creates a model or network based on individual's ratings of similarity between each pair of statements, was used to generate each team member's mental model of taskwork and teamwork. Here, the closeness of the link between statements is represented by a numerical weight, where the statements rated as high in similarity are more closely linked to the rater's model than those rated as low in similarity. The proportion of common links in relation to the total number of links present in two team members' individual networks is the similarity measure for that pair of individuals in a team. The average of similarity scores for all the dyads within a team is the average similarity of mental models for that team. Lim and Klein's study is a rare example of research conducted in the field, using real groups of professionals. Most other studies[37, 38, 47, 105, 126] that focused on capturing shared mental models in general were conducted on undergraduate university students, assembled ad hoc, who were arbitrarily assigned specialist roles and performed specialist computer tasks taught specifically for the purpose of the study.

Mental models in a team captured using paired comparisons ratings, as opposed to other approaches, have been found in a meta-analysis[39] to be most predictive of team processes, such as coordination and communication, Mohammed et al.[32] argued that an advantage of this technique is the use of computerised scaling algorithms that calculate the quantitative indices of similarity and graphic representation of how concepts are linked. However, they criticised the approach for being highly reliant on researchers in determining cognitive content and often using a large number of relatedness statements that may be cumbersome for the raters. The data are uninterpretable in its raw form, unless transformed by a specialised software algorithm that requires a certain level of researcher expertise[145]. Such analysis can be time consuming and extensive, and researchers may find it difficult to define the criteria for when to stop the analysis[147]. This is largely due to the software offering a choice of various statistics in order to arrive at the most interpretable network, which is the one that will allow for the maximum variability in the data with a minimum number of parameters. This may result in different conclusions being reached by different researchers about the structure of a given shared mental model.

1.3.2. Qualitative techniques

Qualitative techniques of assessment of similarity of mental models require that the researcher elicits concepts from participants' statements and extracts relationships between concepts from documents and/or observed and often videotaped team interactions. The advantages of this approach are that the participants provide what is often rich content themselves, and that no aggregation of scores is required, as data are collected at team level. One disadvantage is the need to rely on the researcher's interpretation of the data to deduce the cognitive structure indirectly from behaviour and/or documents, assuming that they do indeed reveal the said cognitive content[32]. Coding of the data can often be time consuming and complicated and comparisons across cognitive structures may be difficult due to different terminology used by different team members[145]. This technique was used in a study of nuclear power plant control room crews working in various simulated nuclear power plant scenarios, where independent coders noted down the behaviours indicative of the crew members developing a shared understanding of the current situation[66].

1.3.3. Concept mapping

In concept mapping, participants place predefined concepts in a hierarchical or chronological structure or map. Concept mapping has been used to capture the sequencing of team actions, which reflects the organisation of knowledge[39]. For example, a concept map can graphically represent chronological steps needed to complete a mission successfully. A point system developed by a researcher is used to code the overlap in individual concept maps of team members, represented by the number of shared linked concepts (which can be steps, events, actions or tasks), and the information on how concepts are linked can be represented graphically[32]. Mohammed et al. [32] noted that concept maps are useful for capturing both the declarative knowledge (the knowledge of "what"), and procedural knowledge, or the step-by-step knowledge, of how the task should be completed. However, they criticised the approach for relying heavily on the researcher's choice of concepts, which can often assume the existence of a rigid linear relationship between concepts when relationships may be represented differently in team members' minds. In addition, determining similarity and accuracy may be cumbersome in large teams and complex concept maps.

1.3.4. Card sorting

Card sorting involves participants sorting concepts, usually events or tasks, written on cards supplied by researchers into meaningful piles that visually represent how the concepts are organised[32]. Usually, the number and labelling of piles is determined by the participants themselves, although a completely structured version exists, where the number of cards and categories and the criteria for the sort can be imposed by the researcher[143]. Each pairing of cards is typically given a score of "1" if they were placed in the same pile, or "0" if not. Each team member's string of "0s" and "1s" is then correlated with every other member's to arrive at a team similarity score. This technique was used in a study where members of submarine attack centre teams sorted 33 cards depicting examples of either effective or ineffective teamwork (developed by subject matter experts) into any number of categories according to how they thought the examples were related[35]. Data analysis involved creating a 33x33 matrix depicting all possible combinations of the cards. A score of "1" was assigned to each cell where the corresponding cards were placed together in a single category. Similarity of the model was then assessed using the Phi coefficient – the Pearson correlation coefficient between two dichotomous variables.

An advantage of the card sorting technique over paired comparisons ratings is that all concepts can be compared simultaneously, rather than pair-by-pair, which provides an insight into how the sorter organises the related information as a whole[143]. Mohammed et al.[32] criticise this technique as being difficult to score if there are a large number of concepts to sort. The analysis can be even more challenging if team members generate a different number of piles of concepts, using different criteria. This critique, however, does not apply to the structured card sort.

Unlike the other two assessment approaches described above, both concept mapping and card sorting have an advantage of being quick, easy, and intuitive to administer and use, making them suitable for contexts where time is precious, such as in the OR. Wildman et al.[143] argued that practicality should take precedence “in some industries and types of teams, [where] there is simply no time to stop and fill out self-report surveys of any sort, even if that has been determined to be the most theoretically appropriate type of measurement”(p. 934). The next section of this thesis describes the combined approach which served as a basis for the development of the new mental model tool.

1.3.5. A combined assessment approach

A combined approach to capturing mental models requires the participants to choose from a variety of prelabelled concepts in the form of cards and place them in a prespecified chronological structure representing their mental model of the issue in question. Unlike traditional card sorting, instead of sorting cards in predefined piles, cards depicting concepts are sorted chronologically. Unlike traditional concept mapping that simply shows linear sequencing of concepts, this approach includes a way to capture the information on the ordering of concepts across various team members. The advantage of this technique is the resulting clear graphic representation of how concepts are linked and how they are related to different team members. Another advantage is the ease of administration and scoring compared to other assessment approaches, such as pairwise comparisons of concepts and qualitative techniques[55]. In the research described in this thesis, I made use of the positive features of card sorting and concept mapping by combining these two techniques, with some modification to allow for greater flexibility in capturing mental models of members of OR teams. The resulting technique was used to simultaneously capture mental models of task sequence and of responsibility for task of individual team members within multidisciplinary OR teams.

1.3.5.1. *A review of the use of the combined approach*

Despite its advantages, the combined approach is not widely used in research into mental models. A multidatabase search of academic literature was conducted using electronic search engines (PsychINFO, MEDLINE, Google Scholar, PubMed, Scopus, ERIC, CINAHL Plus, ProQuest, Sage, EBSCOhost, JSTOR, ScienceDirect, PsycTESTS, Wiley Online Library, SpringerLink, and Web of Science), and a manual search of bibliographies identified from the literature was conducted in reviewing the extent to which the combined approach had been used in previous studies. The following keywords were used: "mental" "model", "shared mental model", "shared knowledge", "shared cognition", "card sort", "task sort", "concept map", "measurement", "technique", "medical", "health", "team members", and "team". This search produced only three studies published in peer-reviewed journals that have combined the ease of use of the card sorting technique and the ability to simultaneously capture declarative and procedural knowledge characteristic of concept mapping. In all three, the concept maps were constructed specifically for the study. The details of the studies are summarised in Table 1.

In a study by Marks et al.[40], mental models of team members were measured using dynamic forms of concept maps called "team interaction concept maps" (p. 975), where participants assigned various task activities to different team members along a timeline. Thus, mental models of task (i.e., the sequence of activities) and of teamwork (i.e., who was thought to do what) could be assessed simultaneously. Marks et al. found that more accurate and similar mental models positively influenced team communication processes. They also found that mental models predicted performance more strongly in novel than in routine environments, and that teams who shared higher quality mental models generally outperformed teams that exhibited less similarity or accuracy.

The assessment approaches used in the other two published studies[54, 56] were modelled on the Marks et al.'s team interaction concept mapping. Using the same scoring methodology as Marks et al., as well as ad hoc assembled teams of university students, Ellis[54] found that acute stress was related to less similar mental models which in turn led to poorer performance.

The study by Burtscher et al.[56] is the only empirical study attempting to directly capture shared mental models of healthcare practitioners published to date. It addressed the relationship between the similarity and accuracy of mental models, and team processes (specifically, monitoring behaviours) and team performance in anaesthesia teams consisting of an anaesthetist and an anaesthesia nurse. The study involved team members sorting 30 cards depicting relevant tasks for anaesthesia induction, identified by subject matter experts, into a chronological order for each member of the team dyad, following a simulated scenario involving induction of anaesthesia. This study showed that, in the absence of a '"sufficiently similar" (p. 262) mental model (which the researchers defined as being one standard deviation below the mean similarity score), monitoring behaviour, such as observing the actions of teammates and the patient's vital signs, disrupts performance of the two-person anaesthesia team. Based on this finding, the authors concluded that similar mental models are a prerequisite for effective mutual team monitoring in anaesthesia teams. This study, however, only focused on one professional team dyad in the OR. This is consistent with many other studies in the OR context where the focus was on individual professional groups, predominantly anaesthesia[3, 23, 50, 113, 148, 149]. The study also captured mental models related to only the first stage of general anaesthesia (i.e., anaesthesia induction), as opposed to all the stages of the operation, and including the whole multidisciplinary OR team that I focused on in my thesis.

Table 1. A summary of peer-reviewed studies that used the combined concept mapping/card sorting approach to capturing mental models in teams

Author (Year)	Sample/team composition and size	Team task	Type of shared mental model captured	Assessment technique	Methodology for scoring the similarity of mental models	Findings
Marks et al. (2000)[40]	3-person teams assembled ad hoc 79 undergraduate university students, randomly assigned to roles and teams	Computer-based simulation of a tank war-game	Team (which team member does what) and task (sequence of activities for each team member)	Team interaction concept map with 24 concepts/actions Team members picked 8 concepts that best represented the actions of each of the 3 team members during mission accomplishment and placed them in the appropriate rows on the concept map (capturing action sequence for all 3 team members) Columns represent a cross section of what all 3 team members should be doing at once	1 point given for each linked set of concepts (A-B) that was shared between 2 team members' mental models; when all 3 members mentioned a certain pairing of concepts, 3 points were given; 3 linked concepts (A –B- C) that appeared in 2 members' concept maps counted as 2 points, and 3 linked concepts that appeared in all 3 team members' maps would add 6 points to the overall score. The same logic applies to shared linkages that were composed of 4 or more concepts Scores could range from 0 (no similarity among any of the three member concept maps) to 111 (3 identical concept maps)	Shared mental models positively influenced team communication processes and team performance. Shared mental models predicted performance more strongly in novel than in routine environments; teams who shared higher quality mental models outperformed teams that exhibited less similarity or accuracy
Ellis (2006)[54]	4-person teams assembled ad hoc 97 university students; team members were not assigned specific areas of expertise	Computer-based war-game simulation	Teammates' roles and responsibilities and team interaction patterns (captured separately)	Modelled on Marks et al.'s (2000) team interaction concept map - 8 blank spaces (2 per team member) that needed to be filled with one of 10 concepts that represented different aspects of the task domain, where participants placed concepts that best represented the actions of each team member on the map	1 point when 2 team members shared 2 linked concepts (A-B), 3 points when 3 team members shared 2 linked concepts, and 9 points when all 4 team members shared 2 linked concepts Scores for the two concept maps were then averaged to form a score for shared mental model of team interaction	Acute stress negatively affected team interaction mental models, which mediated the relationship between acute stress and performance
Burtscher et al. (2011)[56]	2-person teams assembled ad hoc 31 Anaesthesia residents and 31 anaesthesia nurses	Simulation scenario – intubation of a 25-year old male "patient" scheduled for right knee arthroscopy	Team (which team member does what) and task (sequence of activities for each team member)	Modelled on Marks et al., but including 30 concepts – sorting which team member should perform which task and in what order (15 blank spaces for each team member)	Individual maps compared within each team; 1 point for similarity given if a concept was assigned to the same person; 1 additional point if (a) concepts were placed in the same position and (b) had the same previous concept	In the absence of a similar mental model, a higher level of team monitoring had a negative effect on team performance Similarity and accuracy of mental model interacted to predict performance

All three studies using the combined approach used different scoring methods that may be adequate for calculating sharedness in smaller teams, but become cumbersome as the number of team members increases. Also, in two out of three of these studies, the relative positions of a subsequence of ranked tasks in the overall sequence of all tasks are ignored when determining the similarity score between two individuals. Thus, for example, two team members may have the same three tasks ordered in the same way sequentially in their sequence of tasks, but positioned at the top of the entire sequence of tasks for one member of the pair, and at the bottom of the sequence for the other member. However, these two team members will generate the same similarity score as a pair of sorters who both sort the same sequence of tasks at the top or the bottom of the overall task sequence, as in Figure 4. In contrast, the approach used by Burtscher et al.[56] described in Table 1 penalises heavily those cases where the overall sequence may be off by one or more ranks. The approaches to scoring the degree of similarity used in the studies described in Table 1 also penalise those pairs of raters who have assigned the same rank to a task but who have not assigned the same rank to the preceding task.

Sequence	Team member 1	Team member 2
1	C	E
2	A	D
3	B	C
4	D	A
5	E	B

Example 1

Sequence	Team member 3	Team member 4
1	C	C
2	A	A
3	B	B
4	D	E
5	E	D

Example 2

Figure 4. Two examples of how two members of a team ranked five tasks (represented by letters) in a sequence from 1 to 5. If we use the scoring methodology of Marks et al.[40], and Ellis[54] described in Table 1 in both Example 1 and 2, we would assign 2 points to the 3 linked tasks (in this case C-A-B) that appeared in the concept maps of the two teammates, regardless of their position in the overall sequence of tasks. Using the scoring system devised by Burtscher et al.[56], however, in Example 1, we would assign a score of 0, as a) no two tasks were placed in the same position by the two team members, nor b) were they preceded by the same task. In Example 2, however, we would assign a score of 5, as three tasks C-A-B were ranked in the same order, and there were two occurrences where the equally ranked tasks were also preceded by a same task (i.e., task A by task C, and task B by task A). In Example 2, although 1 pair of concepts was preceded by the same task (i.e., task B, in case of the fourth pair in the sequence – D and E for one team member, and E and D for the other), no score is assigned for having the same preceding task.

In addition, Burtscher et al.[56]'s approach does not distinguish between mental models of task sequence and those of responsibility for the task. Instead, the method generates a combined score that includes an extra point if a pair of raters have assigned the same concept to the same person, in addition to sharing the same pair of concepts in a sequence.

The existing combined approaches also restrict the participants to sort tasks sequentially for each team member, without allowing for the possibility that some tasks can be done in parallel, as is often the case in teams. In addition, none of the existing approaches allowed for tasks to be entirely excluded from the sort if the sorter did not think the depicted task was required for the overall task completion. Using existing methods, such tasks would have to be excluded entirely from the calculation process. However, this would reduce the variance in terms of the combined score and make those pairs of ratings where one had the “not required” category seem more similar than they are in reality.

I addressed the above issues by developing a new approach to calculating similarity of mental models that employs several established statistical measures to provide more comprehensive similarity scores (see Chapters 2 and 3).

1.4. Validation of methods of assessment of shared mental models

The paucity of objective validated assessment tools has been identified as a major limitation to improving teamwork and ultimately reducing sentinel events in the healthcare setting[150]. To date, very little is known about the psychometric properties of the existing methods for eliciting mental models and calculating the similarity of mental models in general[99]. For example, it is questionable whether the findings of studies involving undergraduate students can be generalised to members of naturally occurring work teams who perform complex tasks in field settings, over prolonged periods of time and have in-depth skills and expertise[130]. Undergraduates do not have the expertise that individuals in complex domains do, hence their mental models are likely to be different from those of specialists in a given field.

Related to the problem of how best to capture the similarity of mental models is the issue of “interrater” consistency, or reliability, versus agreement, or consensus[151], which is rarely mentioned in previous research on shared mental models. Similarity of mental models seen as interrater consistency implies that team members with similar mental models share the same pattern of responses, regardless of whether the responses are the same. For example, the ratings on a measure of mental model of two teammates would have high consistency if one team member used ratings 1 through to 3 on a 5-point rating scale (for example, 1, 1, 2, 3,3), while the other team member used ratings 3 through to 5 on the same scale (for example, 3, 3, 4, 5, 5). By contrast, similarity seen as interrater agreement relates to the degree to which team members have the same responses in the absolute sense, such that they make essentially the same ratings[152]. For example, there would be high agreement between two teammates where one of them used ratings 1, 1, 2, 3, 3, and the other 1, 1, 2, 2, 3, on the same 5-point rating scale. This distinction has implications for the metric used to arrive at shared mental models. As noted by Webber[151], "… agreeing on one set of strategic actions (as measured by interrater agreement) is more important for interdependent team performance than simply having correlated strategic knowledge and expectations (as measured by interrater reliability)" (p. 313). Thus, the focus in the present study was on similarity as the degree of agreement between team members, or the extent to which they have the same responses.

Unlike in the past, validity is today regarded as a unified concept, where construct validity is the whole of validity. This encompasses content and criterion evidence, reliability, and employing various methods for theory testing[153, 154]. Validity conceptualised in this way requires that relevant evidence is accumulated from multiple sources to support or refute meaningful interpretation of scores. Thus, validation is not an all-or-nothing matter. Rather, it is conceptualised as an ongoing process which relates theory, predicted relationships and empirical evidence at a particular time period in which the validity evidence was collected, to suggest "which particular interpretative meanings are reasonable and which are not reasonable for a specific assessment use or application" (p. 831)[154]. In the early approaches to validity, if the test measured something it was presumed to be designed to measure, the test was deemed to be validated. In the current approach, however, it is the proposed interpretation of test scores that is validated, and not the test itself [155]. Thus, validity involves an evaluation of the overall plausibility of a proposed interpretation or use of test scores[155].

When developing a new measure, it is essential to address content validity[156]. Content validity is the extent to which a measure reflects a particular content domain it was intended to measure[157]. To assess construct validity, the variables of interest should "behave as expected" (p. 103)[158] in relation to other variables representing related constructs[145]. The evidence needs to be built up in different ways to demonstrate that a given method actually captures cognitive structures of interest[153, 154]. This may include, for example, studying expert-novice differences in mental models, where experts are expected to have a higher similarity score on the measure of mental models than novices, and team member versus non-team member differences, where members of the same team are expected to have a higher similarity score for mental models than individuals who are not members of the team. The ability to distinguish between groups that the measured construct should theoretically be able to distinguish between is also referred to as concurrent validity[159]. The evidence that would support validity could also include convergence of the results from multiple techniques, where the same relationships between variables could be obtained using different scoring methods[145].

Validity should be investigated every time a cognitive assessment technique is used in a new context or with different stimuli[160]. The new approach to assessment of similarity of mental models in OR teams in this thesis was evaluated using the unified approach to validity while focusing on content validity and concurrent validity. Content validity was addressed in the tool development stage and involved consulting a panel of subject matter experts in the selection of tasks to be sorted using the new tool. It also involved an interpretation of a series of statistical tests to examine the postulated relationships between relevant variables. Different approaches to calculating similarity of mental models were used to test the assumption that would support validity to establish whether the same construct can be captured using multiple techniques of score aggregation.

1.5. Purpose of the thesis

This project contributes to a wider programme of work[161-166] focusing on teamwork and error prevention in critical care settings, with the overarching hypothesis that harm from human error in surgery and anaesthesia can be reduced through systematic analysis of its causes and through implementation of appropriate strategies to counter these.

The overall aim of this project was to contribute to reducing harm to patients arising from failures in communication by advancing our understanding of the sharing of mental models in the OR.

Related to the overall aim, this thesis had three **specific objectives**.

The **first objective** was to develop a new empirical method for assessing the similarity of mental models in surgery, focusing on laparotomy. This involved developing a task sorting software application for capturing individual mental models of OR team members about particular aspects of laparotomies, followed by developing a method for calculating the degree of similarity of these mental models between individuals. The purpose was to capture the degree to which OR team members agree on the order of tasks necessary for the successful completion of routine laparotomies and who should be primarily responsible for these tasks. It was intended that the basic principles of the new approach to eliciting mental models of individual team members and assessing the degree of their similarity be applicable to different situations, in order to provide a standard method for assessment of these types of mental models in general surgery. I address this objective in Chapter 2 - Tool development.

The **second objective** of this project was to begin the process of validation of the new approach to assessment of mental models in surgery. To provide the initial evidence to support validity of the new approach, the following relationships were examined:

- The link between the scores for the degree of similarity of mental models and team familiarity scores or the extent to which team members are familiar with each other. The assumption that would support validity here is that OR teams whose members have worked with each other more frequently in the past are likely to have more similar mental models of the current task sequence and responsibility for tasks than those whose members have worked with each other less frequently in the past.
- The relationship between the degree of similarity of mental model and subteams – the assumption being that team members from the same OR subteams (i.e., nursing, surgical and anaesthesia) would have higher similarity scores for mental models of task sequence and responsibility for tasks within their own subteam (i.e. intra-team), compared to mental models with members of different subteams (i.e. inter-team).
- The convergence of findings using different approaches to calculating the degree of similarity of mental models – the assumption being that different approaches used to calculate similarity scores for mental models should demonstrate similar results if they are in fact capturing the same real phenomena. Thus, the assumed relationship between the similarity scores for mental models and team familiarity scores, and similarity scores for mental models and professional groups should be able to be demonstrated when the degree of similarity is calculated using different methods of score aggregation.

I address the second objective in Chapter 3 – Tool validation.

The **third objective** of this research was to demonstrate how the new approach to assessment of the similarity of mental models in OR teams could be extended for use in clinical practice. This involved analysing the data from the validation study at a task level, where the similarity scores for mental models of task sequence and responsibility for task within an OR team were calculated for each individual task. This analysis was done at the level of the whole OR team and within the subteams of anaesthesia, surgery, and nursing. Chapter 4 – Tool application addresses this objective.

The next chapter addresses the first objective of the project. In it, I describe the step-by-step process of development of the new approach to assessing mental models in OR teams.

2. TOOL DEVELOPMENT

The need for better tools for assessing mental models in teams is clear. In the concluding comments of their meta-analysis paper, DeChurch et al.[39] called for new approaches to enable the assessment of relevant aspects of mental models in a variety of dynamic field settings such as healthcare, where the existing cumbersome methods are not feasible. In this chapter I describe the steps in the development of a task sorting tool for capturing mental models of task sequence and responsibility for task in multidisciplinary OR teams. The larger context in which my doctoral thesis took place is described first. I then go on to describe a pilot study testing the functionality of the new assessment approach. Finally, I describe the process that led to the selection of the methods for calculating the scores for the degree of similarity of mental models within OR teams used in the subsequent validation study.

2.1. The larger context – the MORSim project

This thesis was undertaken in the context of the **M**ultidisciplinary **O**perating **R**oom **Sim**ulation (MORSim) project, a larger body of interdisciplinary research aimed at examining various aspects of teamwork in the OR (Australia and New Zealand Clinical Trials Registry ID 12612001088831). The overall objective of the MORSim project was to develop a simulation-based course for multidisciplinary OR teams to improve communication and information sharing among team members[167].

The simulation scenarios in the MORSim project provided a context for the development, testing and application of the new task sorting tool for the assessment of mental models of OR team members. Scenarios were developed from real cases encountered by members of the MORSim research group, which included academic and clinical leaders from surgery, anaesthesia, and nursing. The scenarios were designed so as to provide challenges to participants from all of the participating disciplines.

The MORSim simulations were run over the course of a day for each team of participants. The simulation days took place at the Simulation Centre for Patient Safety (SCPS), University of Auckland. Each study day of MORSim involved a team of surgeons, nurses, an anaesthetist and an anaesthetic technician taking part in three highly realistic simulations of critical OR events, where the whole team could work together to manage the patient. Two scenarios (the first, and the last) ran without interruption and provided the vehicle for collecting the mental model data for this thesis. In order to provide a meaningful clinical context for the participants to relate to, the tasks in the task sorting tool were developed from the clinical requirements of these two simulated cases. The two scenarios involved acute abdominal pathology: scenario 1 was appendicitis complicated by sepsis and subsequent allergic reaction; and scenario 2, a stab wound with lacerated inferior vena cava (IVC) complicated by cardiovascular collapse. Details of the scenarios are outlined in Appendix I. In consultation with several specialist surgeons, a surgical model was created to physically fit with existing anaesthetic simulators such that the combined simulator was capable of undergoing anaesthesia and surgery.

Each simulation required the whole team to work together (and in particular coordinate information and activities) to manage the patient effectively. Scenarios were presented in random order to control for time-of-day and order effects. The main study day is described in more detail in Chapter 3 – Tool validation.

2.2. Overview of steps in the development of the new task sorting tool

Langan-Fox et al.[52] proposed a set of steps to be used when developing a tool intended to capture the information on mental models in teams. These steps were considered when developing the new approach to assessment of mental models in this project. First, the researcher should gain an overview of the nature of teamwork within the organisation. This involves becoming familiar with the type of competencies and skills team members need to complete their tasks, the scale and complexity of the team's tasks, and the possible content of the shared mental model. The latter could, for example, include getting a general sense of what does and does not constitute a shared mental model of tasks and a shared mental model of responsibilities. Second, the researcher should choose a method that is time efficient and simple due to the research time constraints. Third, the boundaries of the mental model to be studied should be defined. This may involve such decisions as, should the researcher select the tasks or should this be left to the participants, and how many critical tasks should be used in order to adequately represent a mental model of tasks that need to be completed, in this case, during surgery.

Figure 5 summarises how I achieved the first three steps in my thesis. First, to gain an overview of the nature of teamwork within the organisation, I conducted exploratory observations of ten surgical cases in the clinical setting. The field notes from these exploratory observations, as well as a comprehensive literature review of teamwork and mental models in teams, helped in becoming familiarised with the dynamics of OR teams and in identifying the key clinical tasks and responsibilities of OR teams. This information subsequently served in the selection of content relevant to mental models to be explored using the new tool and how this content was to be presented and organised using the new tool. Second, the literature review on teamwork and the nature of mental models and their measurement described in Chapter 1 helped identify the types of mental models to investigate and the existing methods for assessing mental models in teams. Both the in-theatre observations and the literature review helped with the selection of the optimal approach for assessing mental models of task sequence and responsibility for task in OR teams in my thesis. Third, the boundaries of mental models to be studied were determined with help from subject matter experts, a group of OR clinicians who, based on specific criteria, engaged in an iterative process of defining a list of key tasks in a laparotomy.

Steps four and five proposed by Langan-Fox et al. represent steps towards content validation. The fourth step involves generating a pool of tasks that represent the content of the mental model of interest. In the context of development of the new task sorting tool, this involved generating a list of crucial tasks necessary to complete a surgical procedure, piloting of the tool, getting feedback on the tool, and adjusting of the tool on the basis of feedback. Subject matter experts (SMEs) were consulted in selecting the key tasks to be sorted chronologically. They also helped design the layout of the form in which to sort the key tasks, which would represent a team member's concept map. The tool was then pilot-tested on a group of clinicians from the research team and participants of the MORSim practice run study.

The fifth step involves deriving the final task pool. In my thesis, this step was based on the feedback from researchers and OR clinicians participating in the MORSim practice run day. It involved generating a final list of key tasks necessary for a successful management of a patient undergoing laparotomy and refining the layout of the computer-based version of the new tool. It also involved selecting the optimal methods for calculating the degree of similarity of mental models of OR team members.

The sixth and final step described by Langan-Fox et al. includes eliciting, representing, and analysing the shared mental model in the main study. In my thesis, this was achieved by using the new task sorting tool in the main study designed as a further step in the initial tool validation, and using the selected metrics to calculate the degree of similarity of mental models.

Figure 5 provides a summary of the entire step-by-step process in the development of the task sorting tool for capturing mental models in OR teams.

In the sections that follow, I describe in more detail the preliminary in-theatre observations and how the information gained through this exercise helped in the development of the new task sorting tool. The iterative process of selection of tasks for the task sort is described in more detail in section 2.4. I also describe how the development of the layout of the form in which to sort tasks progressed from a paper version to a computer-based one (section 2.5).

Steps in the development of the new tool for the assessment of mental models	Source/participants	Outcome
1. **Familiarisation** with the nature of mental models, teamwork, tasks, and roles in the OR	Literature review In-theatre observations (10 cases)	Choosing to focus on mental models of task sequence and of primary responsibilities of OR teams for key tasks in general surgery
2. **Choosing a method of assessment** of mental models in OR teams	Literature review In-theatre observations	Chosing a combined card sorting and concept mapping approach for the assessment of mental models of OR team members
3. **Defining the boundaries of mental models** to be studied	In-theatre observations Literature review Subject matter experts (SMEs) - 7 health professionals (3 anaesthetists, 2 surgeons, 2 nurses, all members of the MORSim research team)	Key tasks for the two surgical scenarios to be selected by SMEs Key tasks that can be performed by only one OR subteam to be selected No more than 24 tasks to be selected per scenario (as informed by previous research and to make the task sorting quick and easy)
4. **Content validation - iterations to generate a list of tasks; layout prototypes; piloting and refinement**: a) Generating a large pool of tasks characteristic of OR teams in general surgery b) Refining the list/reducing the number of tasks c) Administering a paper version of the tool (a layout with 3 columns, labelled 'nursing team', 'surgical team', and 'anaesthesia team, and a set of 24 cards depicting individual tasks to be sorted in chronological order, with 8 cards down each column, representing tasks to be completed by each of the 3 subteams) d) **Piloting** the computer version of the new tool	a)In-theatre observations; SMEs from step 3 b) 8 health professionals - 3 anaesthetists (2 from previous iterations and 1 new), 3 nurses (1 new), 1 anaesthetic technician, and 1 surgeon (as above) - all members of the MORSim research team c) 2 human factors researchers (members of the MORSim research team); a software specialist d) 10 research team clinicians - 2 anaesthetists from step 2 above and 1 new; 1 surgeon from steps 1-3 above, 1 from step 1, and 1 new; 1 nurse from step 2 and 3 above, and 1 new, 1 anaesthetic tech as in step 3, and 1 simulation technician Pilot study day - a team of 6 OR health professionals, assembled ad hoc (2 surgeons, 2 nurses, an anaesthetist, and an anaesthetic technician)	a) Selection of 36 scenario-specific tasks from an initial pool of 56 laparotomy-related tasks characteristic of the three OR subteams b) 24 (8 per each OR subteam) salient, discrete, key chronological tasks relevant for a successful completion of 2 simulated surgical scenarios, that could potentially be performed predominantly by one subteam, to be used in the task sort -Number of tasks further reduced to 21 in a subsequent iteration c) Change of form layout to allow for tasks to be sorted in parallel as well as in sequence, and Assigning tasks to one of the 3 OR subteams to be achieved by right clicking on a task in a subsequent computer version to bring up a menu and choose one of the 3 options/teams d) Further reducing the number of tasks to 20 with exclusion of ambiguous tasks. Allowing for unsorted tasks where a task may be positively dismissed as "not required" for the procedure
5. Content validation - **deriving the final task pool; selecting metrics** for calculating the degree of similarity of mental models of OR team members	Feedback from those participating in the piloting of the tool in step 4; literature review	A generalised tool, consisting of 20 cards to be sorted in a predefined form, in order to capture mental models of chronological tasks that occur in all laparotomies through: 12. Having 3 generic "anchor" tasks for timing into phases that define the sequence of the procedure, with 3-4 generic tasks that precede and follow each anchor (a total of 15), and 2 tasks that could be scenario-specific Methods to calculate the similarity of mental model selected
6. **Main study** /tool validation - administering the tool during the MORSim simulation study days	20 established, 6-member OR teams (2 surgeons, 2 nurses, 1 anaesthetist, and 1 anaesthetic technician)	Comparisons of similarity intra- and inter- team, using different methods to calculate the degree of similarity Analysis of the relationship between shared mental models and team familiarity

Figure 5. Summary of steps, participants, and outcomes in the development of the new task sorting tool.

2.3. Exploratory in-theatre observations

2.3.1. Purpose

The purpose of preliminary in-theatre observations was exploratory and the resulting observations descriptive in nature. Specifically, the goal of these observations was to learn about the procedures and processes in the OR, tasks and roles of individual team members and subteams (i.e., anaesthesia, surgery, and nursing), as well as the whole team. The goal was also to observe the sequence in which actions are performed in various stages of a surgical case. This information then assisted in identifying the key actions or tasks for the task sorting exercise, developed to capture mental models of task sequence and of responsibility for task. In addition, the goal was to observe how members of subteams (anaesthesia, surgery, and nursing) coordinate their activities, as well as how they coordinate their activities with the other OR subteams (for example, between surgical and anaesthesia subteams, and nursing and surgical subteams).

2.3.2. Sample and method

The observations were conducted in Level 8 operating theatres of Auckland City Hospital. Ethics and institutional approvals (NTX/12/EXP/067, A+5462, CMDHB#1267) were obtained for the collection of clinical observations data as part of the MORSim project. A total of 10 procedures (three laparotomies, five vascular cases, and two urology cases) were observed over a period of two months, resulting in a total of 24 hours of surgery observed. The same number of cases was observed in a qualitative study on the nature of coordination of surgical teams[168]. The observation of procedures lasted between 35 minutes and 8.5 hours. The selection of cases was one of convenience. However, observing a variety of surgical cases allowed me to familiarise myself with the tasks, roles and responsibilities of the three

OR subteams (anaesthesia, surgery, and nursing) and gain a general insight into the nature of team dynamics between different surgical subspecialties. Watching a variety of cases also enabled me to identify the characteristics of roles and tasks potentially unique to laparotomy that I would subsequently explore in the simulated setting. The details of the specific cases observed are summarised in Table 2. As in an earlier study[169], field notes were taken, but no formal method in structuring the observations was used, consistent with naturalistic observation.

Table 2. Details of the observed surgical cases

Specialty	Procedure	No of cases	Duration (min.)
Vascular	Right portacath insertion	4	52-55
Vascular	Bilateral long saphenous vein ligation + varicose ligation (defx1)	1	100
General (laparotomy)	Excision of choledochal cyst with hepatojejunostomy	1	247 (estimated 100)
	Laparotomy for splenectomy [incorrectly recorded in the operative schedule as porto-caval shunt]	1	175 (estimated 260)
	Liver resection (left hepatectomy) and biliary reconstruction + cholecystectomy	1	510 (double the estimated time)
Urology	Bilateral ureteroscopy +- laser fragmentation +- stent insertion +SPC insertion [it was decided not to proceed with the later due to a scrotal abscess [also repaired here]	1	135 (estimated 165)
	Cystoscopy +SPC insertion	1	37 (estimated 75)

2.3.3. **Summary of observations**

Detailed description of the observations can be found in Appendix II. The main points can be summarised as follows:

- Team coordination in the OR occurs both in serial and parallel fashion. Concurrent activities are often performed by different individuals or subteams, or require inter-team collaboration. Successful coordination of these activities appears to rely on the members' understanding and expectations regarding each other's tasks and preferences, and the knowledge of the steps necessary for the successful completion of the case. This finding helped in the selection of the initial pool of general tasks in surgical procedures to be considered for the task sorting tool. As a result of the observation that certain tasks are conducted in parallel and by different members of the OR team, various designs of layout in which to perform the task sort were tested.
- In general, team members appeared more comfortable engaging in intra-team coordination, with members of their own subteam, than in inter-team coordination, with members of the other two subteams. This information provided further support for the need to test the assumption that mental models may be more similar intra- than inter-team to be tested in the tool validation phase of my project.
- Coordination breakdowns appeared to be more common inter-team, although episodes of miscommunication were observed intra-team, and especially where an outsider to the team or a trainee was involved. Established teams, whose members regularly work with each other in a team, were more likely to coordinate their information and activities smoothly and they appeared to have a better understanding of each other's roles, expectations and tasks. This observation provided further justification for testing the assumption that the extent to which team members are familiar to each other may

be positively related to similarity of mental models in OR teams to be tested in the validation phase of my project.

2.4. Selecting tasks for the task sorting tool – content validation

Guided by the review of relevant literature, the findings from the in-theatre observations, and SMEs, I generated a pool of 56 potential tasks for the new task sorting tool. Task selection was an iterative process guided by previous research[40, 54-56]. This involved narrowing down the initial large pool of potentially relevant tasks to a final list of key tasks through a series of suggestions, discussions and refinements by SMEs. Expert OR health professionals, including at least two representatives from each OR professional group (i.e., anaesthesia, surgery, and nursing) assisted in selecting the tasks for the task sort, based on the scenarios in the MORSim study. In addition, feedback was sought from a convenience sample of six OR clinicians (two surgeons – a consultant and a registrar, two nurses, an anaesthetist, and an anaesthetic technician), assembled for the purpose of the MORSim practice run day, who completed the task sorting exercise. The following criteria were used in the selection of tasks:

- Each task should represent salient, general-purpose, chronological steps or actions required during all laparotomies, as they would be performed routinely and without complications.
- The task should be considered clinically important and likely to make a difference to patient safety if the team did not have an agreed understanding of if, when, and by whom it should be done.
- It should be a task that could be performed predominantly by one OR subteam, as opposed to by the team as a whole, or concurrently by multiple subteams.

The final list of tasks was not intended to be comprehensive, or necessarily consist of tasks that are of equal importance. Rather, the list was to simply represent clearly identifiable steps required for completion of a laparotomy in two study scenarios. Guided by previous research (Chapter 1, Table 1), it was agreed that the final list of tasks for the task sort was not to exceed 24 items. The aim here was to maximise the ease of use of the new tool while minimising the time required to engage in the task sorting activity – factors that may be important for potential future use of the tool in the clinical setting. Prior to piloting the software version of the tool, and following multiple iterations (described in Figure 5), the initial pool of 56 laparotomy tasks was refined and scaled down to a list of 21 tasks per scenario in the MORSim study (Table 3).

Table 3. Lists of tasks used in the task sorting exercise in the pilot study for the two MORSim study scenarios.

Laparotomy for an abdominal stab wound (scenario 1)	Laparotomy for a suspected perforated viscus (scenario 2)
1. Check blood availability	1. Take brief anaesthetic history and verbal consent
2. Check patient positioning on table is satisfactory for access to abdomen	2. Prepare drugs and intubation set
3. Assess and manage cardiovascular and fluid status	3. Ensure IV lines are patent
4. Lead sign in	4. Check the position of the endotracheal tube and ventilate
5. Pre-oxygenate patient	5. Ensure patient is covered for dignity and kept warm
6. Check monitoring is attached (ECG, SpO2, A-line)	6. Check for scars/signs of previous abdominal surgery
7. Perform a rapid sequence induction	7. Decide if compression stocking or calf compressors are required
8. Perform urinary catheter insertion	8. Attach and check suction and diathermy
9. Prepare surgical trolley	9. Confirm whether new antibiotic agents are required
10. Make incision	10. Lead time out
11. Suction blood out	11. Make incision
12. Remove stab knife	12. Examine intra-abdominal organs
13. Ensure safe management of stab knife	13. Locate site of perforation
14. Pack abdomen	14. Repair perforated viscus
15. Decide if massive transfusion protocol should be activated	15. Wash out abdomen
16. Phone blood bank to request blood if required	16. Decide if blood should be transfused
17. Serially remove packs off abdomen	17. Place a drain
18. Assess damage, locate and repair injury(ies)	18. Close the incision and dress wound
19. Decide whether to leave abdomen open or close up	19. Ensure bed space for patient post-op
20. Lead sign out	20. Give main handover to PACU or ICU staff
21. Liaise with ICU regarding post-op placement if required	21. Confirm estimated blood loss

2.5. Developing layout for the task sorting tool

Two human factors researchers and a software specialist were consulted regarding the layout of the form in which to perform the task sort. The first version of the layout matched the one used by Marks et al.[40] described earlier (Chapter 1, Table 1). It involved SMEs sorting cards, depicting individual tasks as they would occur chronologically in the procedure, into three separate columns on the form for the three OR subteams (i.e., anaesthesia, surgical, and nursing) primarily responsible for them. Participants were instructed to sort all the given cards into the predefined layout in paper form. Figure 6 below shows the resulting paper version of the layout. Following this exercise, it became apparent that forcing participants to sort tasks in a single chronological sequence and then assigning each task to one of the three subteams would provide additional information as to the overall sequence of tasks for each scenario. This information would otherwise be lost if the tasks were simply divided among subteams, into separate columns.

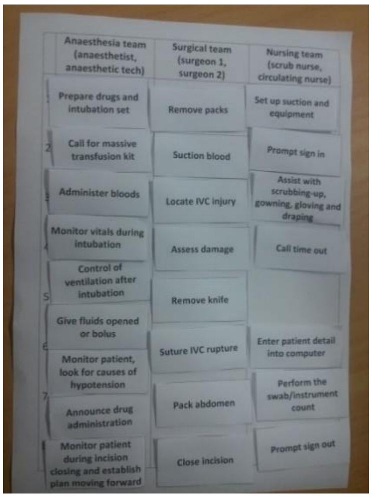


Figure 6. An early paper prototype of the task sort with an example of the completed concept map.

After subsequent iterations and feedback from SMEs, and following exploratory observations, it also became evident that some of the tasks in the task sort could be performed in parallel, rather than having to be forced into a sequence. The subsequent software version of the task sort, built to specification by a software specialist using Visual Studio 2012, was thus adjusted to allow participants to sort tasks in parallel, as well as in a sequence. This was made possible by changing the layout of the form to include a “main action sequence” column, and four additional “parallel actions” columns (Figure 7). In this version, the tasks (i.e., actions) could be manipulated by dragging and dropping from the “unsorted” column on the screen, to the positions in the “main action sequence” and “parallel actions” columns.

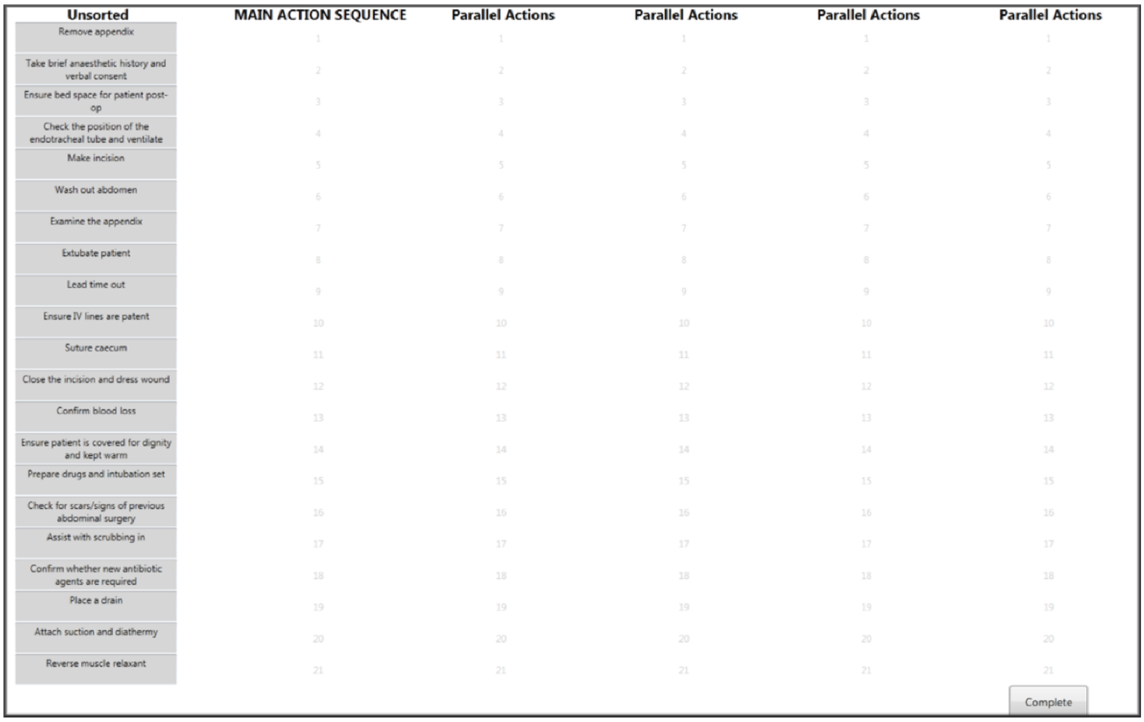


Figure 7. A snapshot of the task sorting software interface allowing tasks to be sorted in parallel as well as in a sequence.

Each time a task was positioned in the "main action sequence" or one of the "parallel actions" columns, a menu automatically popped up offering a choice of one of the three subteams primarily responsible for that task (Figure 8).

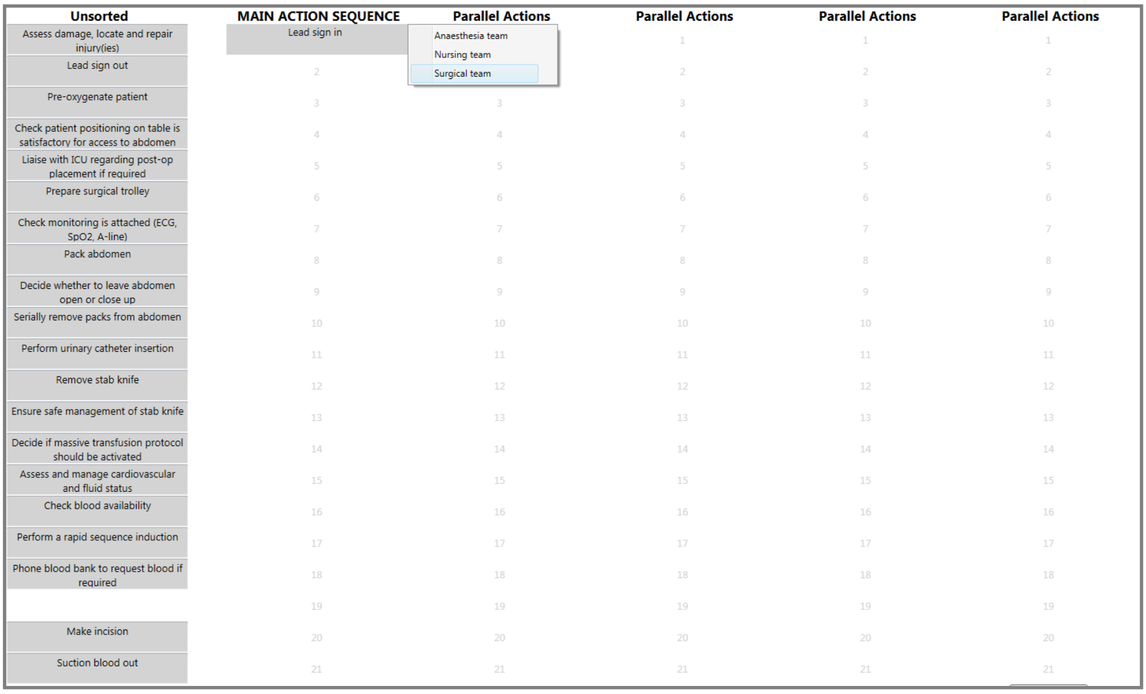


Figure 8. A snapshot of the task sorting tool interface showing a task sort in progress for the laparotomy for an abdominal stab wound scenario. Whenever a task (i.e., action) is dropped on to a position in a column, a menu automatically appears offering a choice of one of the three subteams primarily responsible for that task. Clicking on one of the subteams selects that subteam as primarily responsible for that task.

In the software version of the tool, the data output for each individual task sort is automatically stored in a Microsoft Excel spreadsheet for easy data transfer and analysis. For the mental model of task sequence, the numerical rank assigned to each task is recorded, including tied ranks for those tasks sorted in parallel. For the mental model of responsibility for task, a descriptive code is assigned to each task indicative of the subteam an individual performing the task sort selected as primarily responsible for the task (i.e., A – for anaesthesia subteam; S – for surgical subteam or N – for nursing subteam).

This version of the layout was then piloted on a group of 10 health professionals from the MORSim study research team (including 3 anaesthetists, 3 surgeons, 2 nurses, 1 anaesthetic technician, and 1 simulation technician), as well as on a set of OR clinicians who participated in the MORSim practice run day. The final version of the task sorting tool was based on their feedback and was then used in the main validation study.

2.6. Pilot study

2.6.1. Purpose

The purpose of the pilot study was to test the functionality of the new computer-based task sorting application for capturing mental models, to be used in the subsequent validation study. Of main interest were the participants' verbal feedback on usability of the computer interface, the clarity of the instruction for task completion, and the selection and wording of tasks for the two scenarios. The latter was the final stage in the process of content validation.

2.6.1.1. *Setting, participants and procedure*

The new task sorting software application was piloted as part of the MORSim practice run day conducted at the Simulation Centre for Patient Safety at the University of Auckland. The purpose of this day was to run through one of the study's scenarios and obtain feedback from OR clinicians (other than the ones in the MORSim research group) on the surgical and anaesthesia models and set up, as well as the task sorting tool. A team of six OR clinicians (2 surgeons, 2 nurses, an anaesthetist, and an anaesthetic technician) participated in the pilot. They were assembled ad hoc from various public and private hospitals in the Auckland region.

After the participants completed consent forms, they were familiarised with the simulated OR environment. They were then given a 20 minute talk on teamwork by a member of the MORSim team, followed by participation in an introductory familiarisation scenario[167]. Participants then took part in a realistic simulation of an emergency laparotomy for an abdominal stab wound. This immersive scenario required the whole team to work together effectively in order to manage the case.

Participants performed the computer-based task sort, to capture their mental models related to the upcoming case, immediately prior to the case scenario. They first received a demonstration on how the computer software works by observing the researcher working through a simple example of a task sort using the identical interface, but with a smaller number of tasks other than those in the actual task sort that was to follow. The example task sort was for a simple hand-washing procedure, where a small child is being helped by an adult to wash hands. Six tasks represented salient steps in the procedure to be sorted in chronological order or in parallel: turn tap on; apply soap to hands; rub hands together in circular motion; clean in between fingers; rinse hands in the running water; and check no soap left on hands. Each participant was then presented with a briefing sheet outlining the details of the upcoming simulation scenario (see Appendix III for details of the briefs). They were then instructed to start the task sort which required them to first read on-screen instructions on how to manipulate tasks on the screen (see Appendix IV) before proceeding to sort the tasks. Participants were instructed to use all the tasks available on the screen. If they did not include all the tasks in their sort, a message popped up upon trying to exit the application reminding them to use all the available tasks.

2.6.1.2. ***Results and discussion***

Participants took an average of 9.33 minutes (range = 7-11 minutes) to complete the task sort. Figure 9 shows the variations in ranking sequences of individual team members for the 21 tasks in the procedure. Individual ranks were re-ranked so as to account for tied ranks in those cases where a participant assigned the same rank to multiple tasks when they thought those tasks should occur in parallel during the procedure. This was done by assigning the lowest possible rank to all tasks that were assigned the same rank. Thus, if for example, task A ranked ahead of tasks B and C (which share the same rank) which were both ranked ahead of D, then A was given ranking number 1, B and C were given a ranking number 2 and D was given ranking number 4. As shown in Figure 9, there were variations in how members of the OR team ordered tasks in the procedure. In the extreme cases, some team members believed some of the tasks should be undertaken at the beginning of the procedure, while others thought they should be done towards the end (for example, phoning blood bank to request blood if required, and liaising with ICU regarding post-op placement if required).

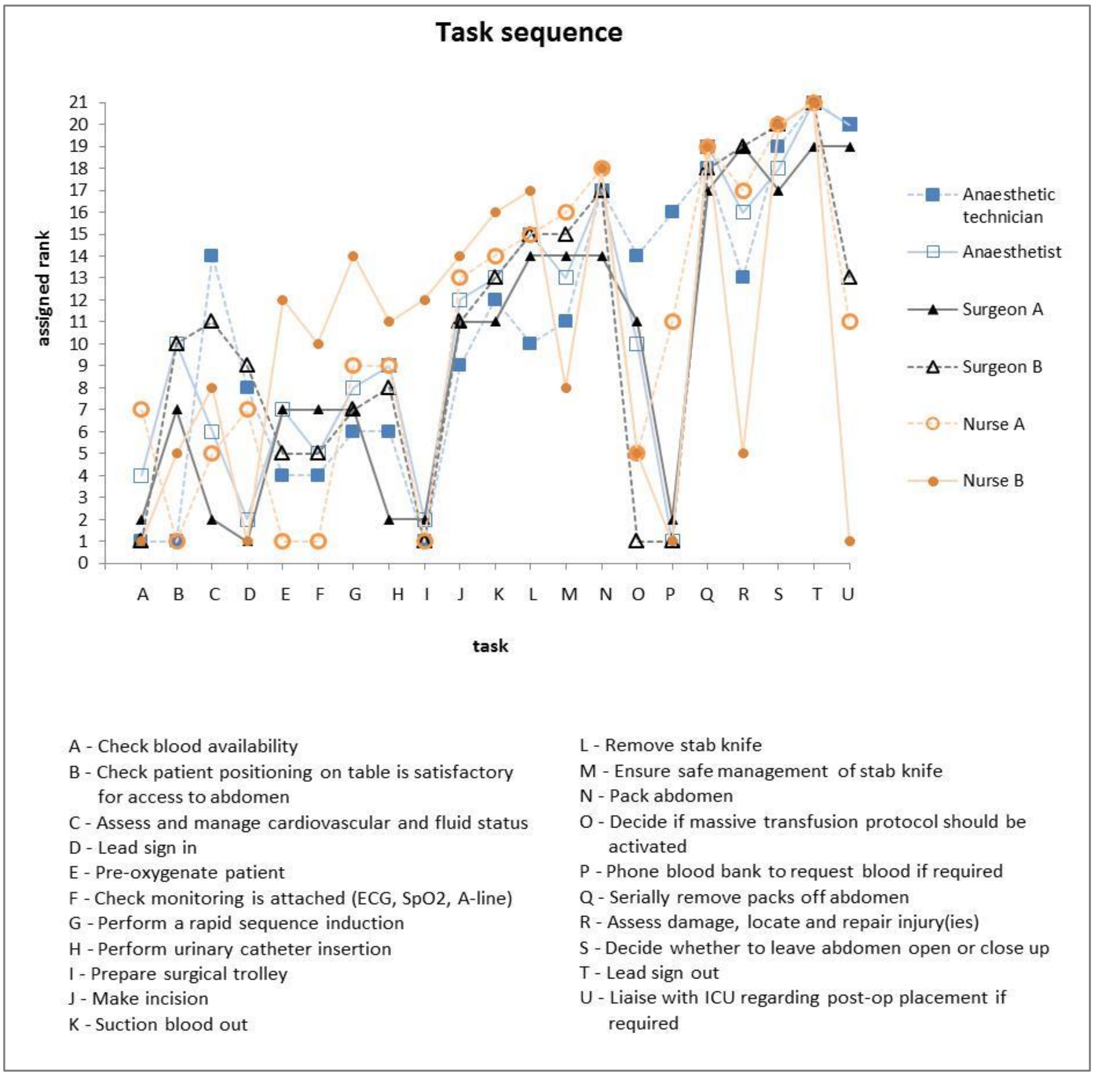


Figure 9. A scatterplot of ranks assigned by individual team members (n=6) to individual tasks representing the order in which the tasks should occur in the procedure.

Figure 10 shows the percentage of team members who selected one of the three subteams as primarily responsible for individual tasks. There was a 100% agreement between team members on which subteam they thought should be primarily responsible for nine out of 21 tasks. The anaesthetic technician and the anaesthetist had the most similar responses to which team they thought was primarily responsible for the tasks in the procedure (95%), while the senior surgeon and nurse B had the least similar responses (48%).

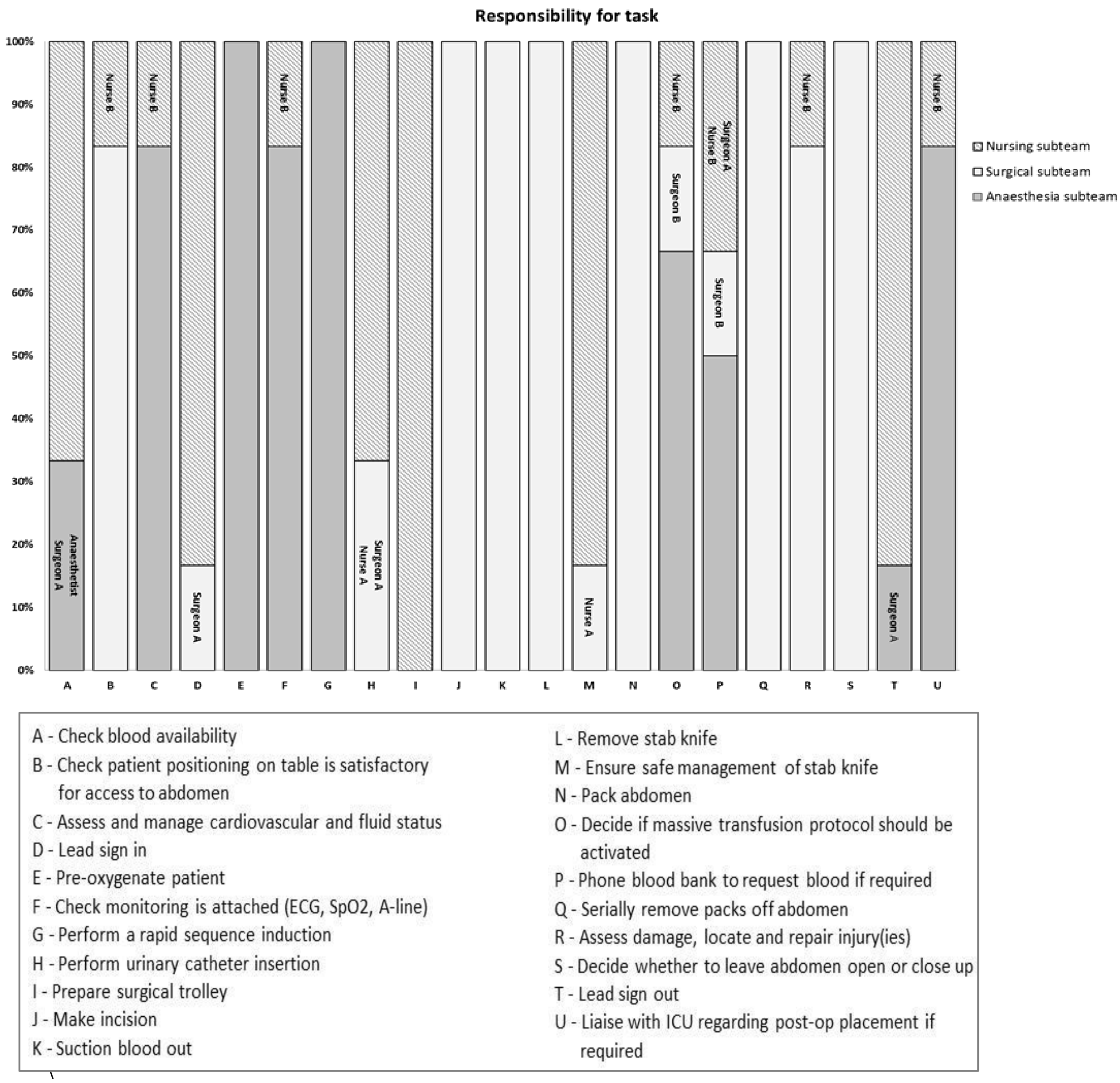


Figure 10. A bar chart showing the percentage of the team (n=6) who selected one of the subteams (nursing - stripes; surgery - light grey background; or anaesthesia - dark grey background) as primarily responsible for each task. The labels inside the bars specify individual team members who selected a subteam different from the majority of their teammates as primarily responsible for the corresponding task.

Based on the feedback from both the research team and the participants of the pilot study, it became apparent that the task sorting tool had the potential to be further refined into a generalised instrument that could be used in different settings. The tool was envisaged as capturing general tasks that occur in all laparotomies, where the answers to the same set of questions would vary given different situations. It was also agreed that such a task list should consist of generic but important "anchor" points or tasks for timing into phases that define the sequence of the procedure (for example, intubation, making the incision, closing the incision). The tool was to be modified to also include two to three general tasks that precede and follow each anchor. In addition, the tool was to allow for inclusion of two to three tasks that could be scenario specific. Such a template would allow for a somewhat cruder, but more meaningful analysis of data, where the exact ordering of tasks into a sequence is not of an essence, as long as the tasks occur between the correct anchors and anchors occur in the correct sequence. In addition, the expert clinicians agreed that an additional "not required" category should be included as an option in the task sorting exercise, where a task may be positively dismissed as unnecessary by a participant and therefore not included in the sort for a given procedure. For the unsorted tasks (i.e., those regarded as "not required" by the individual sorting the tasks), a numerical rank of 0 is to be automatically assigned in the accompanying spreadsheet in which the data on the completed task sort are stored.

The clinicians participating in the piloting of the tool also commented on the ambiguity of some tasks used in the task sort. For example, the task "Assess damage, locate and repair injury(ies)" and "Assess and manage cardiovascular and fluid status" both consist of multiple actions or tasks instead of one. This was rectified in the final version to allow for each label to only represent a single task.

Following the final iteration by the SMEs, the final version of the task sort tool consisted of 18 general tasks (including three "anchors") and two scenario-specific ones, making a total of 20 tasks (Table 4).

Table 4. Final list of tasks in the task sort. For each laparotomy scenario, there were 18 general tasks and two scenario-specific tasks.

General tasks
1. Check blood availability
2. Check for optimal patient positioning on table
3. Initiate sign in
4. Administer anaesthesia induction drugs
5. Perform a rapid sequence induction
6. Ensure appropriate antibiotic prophylaxis
7. Insert urinary catheter
8. Initiate time out
9. Make surgical incision
10. Ensure patient warming devices in place
11. Ensure TED stockings and calf compressors on
12. Monitor ongoing blood loss
13. Organise bed space in PACU (Post-anaesthesia care unit)
14. Close incision
15. Check drains are turned on
16. Confirm estimated blood loss
17. Initiate sign out
18. Provide handover on intraoperative events to PACU staff
Scenario-specific tasks
Laparotomy for an abdominal stab injury (knife in situ) (scenario 1)
19. Inform intensive care unit
20. Remove knife from abdomen
Laparotomy for a perforated viscus (scenario 2)
19. Locate site of perforation
20. Fashion stoma

Figure 11 shows a schematic representation of a snapshot of the final version of the task sort interface at the moment of assigning a team to an action, i.e., task.

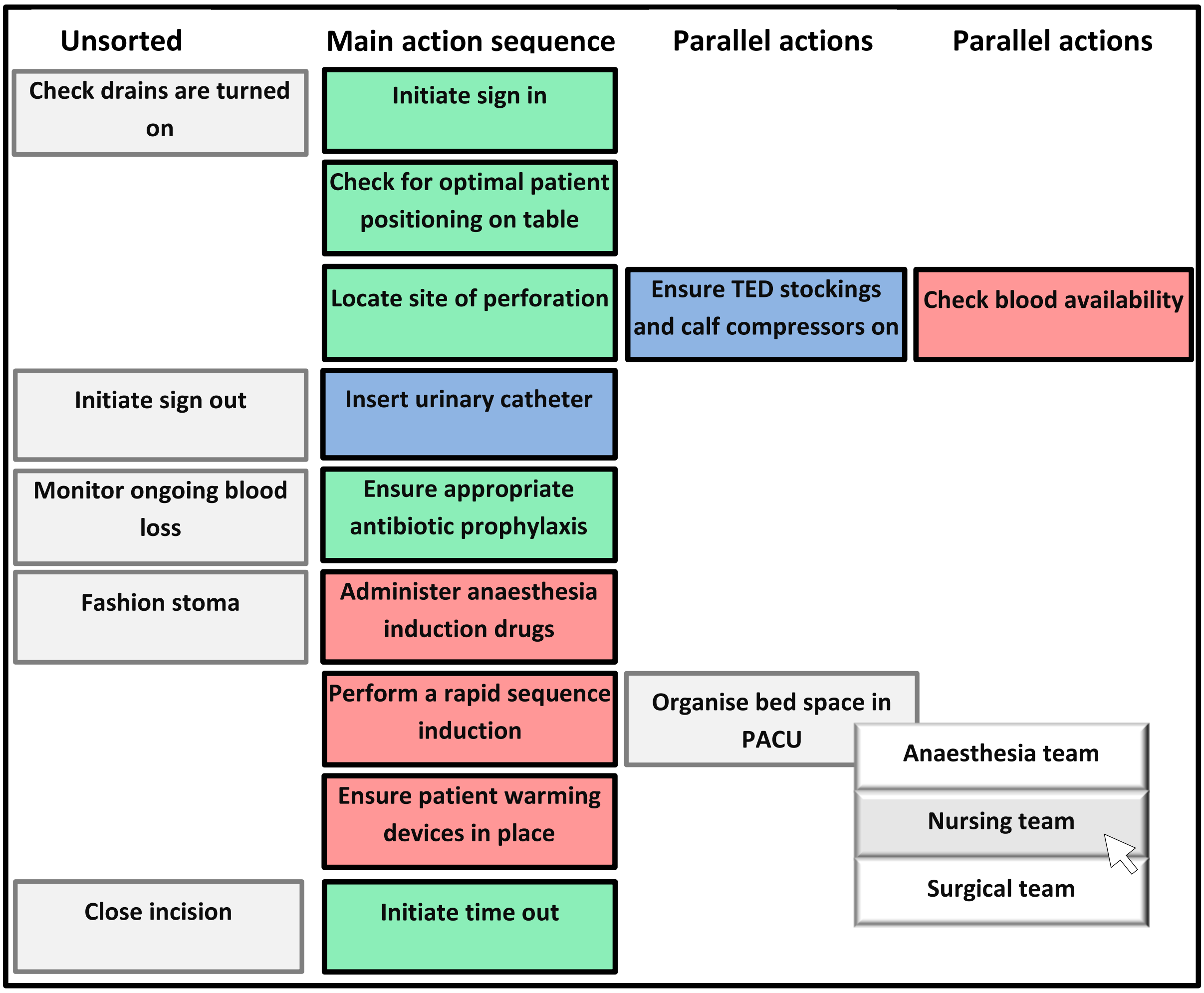


Figure 11. A schematic representation of the top left portion of a computer screen showing the task sorting exercise in progress for the laparotomy for a perforated viscus scenario. An individual can sort tasks chronologically, down the "Main action sequence" column, and in parallel, by dragging and dropping tasks from the "Unsorted" column on to the positions in four "Parallel actions" columns. Whenever a task is dropped on to a position in a column, a menu automatically appears offering a choice of one of the three subteams primarily responsible for that task (as for task "Organise bed space in PACU"). Teams are colour coded (i.e., red=anaesthesia; green=surgical; and blue=nursing), and the "card" depicting the task changes colour accordingly once a team has been assigned to a task.

Participant feedback was that the verbal instructions and the hand-washing example were sufficient to prepare the participants for the task sorting exercise, so the on-screen instructions on how to complete the task sort were excluded from the final version of the tool. The new tool was named Momento, a play on words intended to imply its proposed use as a (to)ol for capturing (ment)tal (mo)dels in the given context.

2.7. Calculating similarity of mental models in the Momento approach

This section describes how calculating the degree of similarity of mental models of OR team members elicited by the Momento tool was approached in the subsequent validation study.

One of the biggest challenges to research on mental models has been the question of how to measure and represent cognition at the team level[145, 170, 171]. So far, there has been no consensus on the best way to assess the degree of similarity of mental models. Mental models in teams are extremely complex, and we should not make definitive conclusions on the basis of a single method used to capture them[145]. Some approaches seem to be more useful for some types of mental models than others[151]. A meta-analysis of literature on mental models in teams[39] revealed that different ways to calculate the degree of similarity have been found to yield different results. In most studies, data were aggregated in order to represent constructs at a group level[145], as described in Chapter 1, section 1.3 of this thesis.

In this project, I aimed to capture a representation of the degree to which different professionals comprising an OR team agreed on two particular aspects of the task to be done, from the moment the patient is brought into the OR to the moment the surgery is completed and post-operative care organised. The resulting data on the order of and responsibility for tasks would therefore provide some insight into the degree to which mental models of task sequence and responsibility for task of these professionals overlap, or are shared. Also, the aim was to provide a standard method for the assessment of mental models in laparotomy. A standard method of assessment that would be applicable in wider research settings has not been reported in literature to date. The biggest challenge by far to this task was finding an optimal way to calculate the scores for the degree of similarity of individual mental models within a team captured using the task sorting approach.

Mohammed et al.[32] argued that there can be no advancement in the field of measurement of mental models at team level if we do not compare and contrast multiple assessment approaches within the same study. Comparing different statistical approaches to aggregation to calculate the degree of similarity of mental models should not be difficult within a single study as it would not require extra data collection efforts[143]. In view of there being no universal method of aggregation of individual scores on mental models, and following the logic that a suite of metrics capturing the same effect in different ways is much more powerful than a single new metric[172], I considered several approaches to assessing similarity of mental models for use in the validation study. For calculating the scores for the degree of similarity of mental model of task sequence (based on comparing task sequences of individual team members), these approaches included calculating the number of common pairs of ranks in two strings, Euclidean distance, Edit distance (Levenshtein distance) , and Krippendorff's alpha for ordinal data. For calculating the scores for the degree of similarity of mental model of responsibility for task (based on comparisons of categories of subteam selected as responsible for tasks) I considered percentage agreement, Cohen and Fleiss' kappa, and Krippendorff's alpha for nominal data. A brief overview of these methods is offered in this chapter; a more detailed description of how the selected approaches were used to analyse the data from the validation study is provided in Chapter 3 – Tool validation.

2.7.1. Calculating similarity scores for mental model of task sequence

The biggest challenge to calculating the similarity scores for mental model of task sequence from the Momento data was how to overcome some of the problems encountered in arbitrary scoring approaches used in previous studies (see Chapter 1, section 1.3.1.1). Another important issue was that the new scoring approach had to accommodate those instances in which one or more team members did not include one or more tasks in the task sort if they thought that those tasks were not required for the management of a patient in a given scenario. Previous scoring approaches did not allow for this option.

The basic similarity score from the data generated in the Momento task sort is calculated by comparing task sequences produced by two members of an OR subteam. These task sequences may contain tied ranks in those cases where team members rank two or more tasks as being done in parallel during the procedure. They may also not include one or more tasks one or both team members decided were not required for the procedure.

The first approach to calculating pairwise similarity scores I considered was a variation on the previous scoring approaches. It involves finding the number of common pairs of ranks in two strings (in this case, two task sequences produced by two team members). Each pair of ranks common to the two sequences is assigned a score of 1, and the similarity score for two team members is the total number of pairs of ranks common to the two task sequences. This approach makes the scoring for those tasks not included in the task sort as deemed unnecessary for the procedure easy, as the “not required” category can simply be assigned an arbitrary rank, or a letter where individual tasks are denoted by letters. The approach also takes into account to some extent the information on the order of tasks, as only those tasks that appear in the same order in the two sequences are considered a pair (for example, A-B is not the same as B-A). However, the problem with this approach is that tied ranks make counting the number of common pairs problematic. As with previous scoring approaches, if a sequence is off by one or more ranks, the score changes, as depicted in Figure 12. This problem becomes more apparent the more tied ranks there are in a sequence. As a result, we may underestimate how similar two task sequences are, leading to overall poor discriminability between similarity scores.

	Team member 1	Team member 2
Task	Rank assigned	Rank assigned
A	1	1
B	2	2
C	3	3
D	4	4

Example 1

	Team member 3	Team member 4
Task	Rank assigned	Rank assigned
A	1	1
B	2	1
C	3	3
D	4	4

Example 2

Figure 12. Two examples of how two members of a team ranked four tasks (represented by letters) in a sequence from 1 to 4. Using the number of common pairs in a string as the scoring approach, the similarity score in Example 1 is 3 (i.e., for three identical pairs of ranks: 1-2, 2-3, and 3-4), but only 1 in Example 2 (i.e., 3-4). In Example 2, tasks A and B are ranked as occurring in parallel; consequently, the sequences for the two team members are off by one rank.

In the search for a scoring approach that would be more successful in detecting differences between pairs of whole sequences of ranks (rather than simply between pairs of individual tasks in a sequence), I considered Edit distance next. Edit distance is a way of measuring how dissimilar two sequences of characters are to one another by counting the minimum number of operations required to transform one sequence into the other. Thus, two strings are similar if only a few operations are needed to transform one string into the other[173]. The most common Edit distance measure is the Levenshtein distance, expressed as the minimum number of characters to be deleted, inserted or substituted in order to convert one string into another[174, 175]. The greater the Levenshtein distance, the more different the strings are[176].

The simple Levenshtein distance LD(a, b) needed to convert string a into string b (through insertion, deletion or substitution) is calculated as:

$$LD(a, b) = \min(i + d + s)$$

For example, the Levenshtein distance between the strings “world” and “swirl” is 3 (one insertion: s; one substitution: o with i; and one deletion: d):

world → **s**world (insertion of “s”)

sw**o**rld → sw**i**rld (substitution of “o” with “i”)

swirl**d** → swirl (deletion of “d” at the end)

Although Edit distance is mostly used in computational biology and computer science to develop software applications such as spelling checkers, one version of this approach used substitution only for comparing a large number of open-ended card sorts to reveal the underlying concepts that occur commonly across different respondents[177].

The problem with using Edit distance on the Momento data, however, is also poor discriminability. Since this approach only counts the number of changes to convert one string into another, and does not deal with the magnitude of these changes (for example, how far apart the ranks are for the one sequence to be transformed into the other), the similarity score between two sequences of task can be overestimated as in the examples shown in Figure 13.

Sequence	Team member 1	Team member 2
1	A	A
2	B	C
3	C	B
4	D	D

Example 1

Sequence	Team member 3	Team member 4
1	A	D
2	B	B
3	C	C
4	D	A

Example 2

Figure 13. Two examples of how two members of a team ranked four tasks (represented by letters) in a sequence from 1 to 4. Levenshtein distance in both examples is 3, although the sequences in Example 2 are quite different, and are only off by one in Example 1. In Example 1, to convert the task sequence of team member 2 into that of team member 1 requires one substitution (C for B), one deletion (B), and one insertion (B before C). In Example 2, one substitution is required in the task sequence for team member 4 (D for A), one deletion (A), and one insertion (A before B).

The optimal approach to calculating the similarity scores for the Momento data would be to count the number of moves required to transform one task sequence into another. This would allow for more discriminability the further away tasks are ranked from each other in the two sequences. Kendal tau distance, also known as the bubble-sort distance (as it can be calculated using a sorting algorithm), between two sets of ranks is calculated by counting the number of pairs that are in opposite order in the two rank sequences[178, 179]. Applied to the examples in Figure 13, Kendal distance in the case of Example 1 would be 1 (B-C and C-B), and in Example 2, it would be 5 (A-B and B-A; A-C and C-A; A-D and D-A; B-D and D-B; and C-D and DC). The bubble-sort algorithm achieves the same by comparing each pair of adjacent tasks and swaps them if they are in the wrong order compared to the task sequence of the other sorter. Thus, in the Example 1 above, C would swap with B, while in Example 2 five swaps are needed (D with B; D with C; A with D; A with C, and A with B). The problem with this approach however, is that it cannot be applied if the “not required” category is to be included in the comparisons of task sequences. Simply coding the “not required” tasks as additional letters (or numbers) makes it impossible to count the number of swaps necessary to transform one sequence into another in those instances where one team member includes all the tasks in his or her task sort and the other member does not. Similarly, this approach cannot be used if more than one task is excluded as "not required" from one of the task sorts in the pair of team members.

In order to attempt to overcome some of the shortcomings of the above methods, I considered Euclidean distance next. Euclidean distance is the shortest distance between two points (i.e., "as-the-crow-flies" distance)[180], or the absolute value of their numerical difference. For two sets of items, it is calculated as the square root of the sum of the squared differences between values for the items:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Thus, the smaller the Euclidean distance, the greater the similarity between two sets of items. Euclidean distance is an easy to understand, intuitive measure. It has been suggested for assessing within-group diversity in organisations[181], and has been used empirically to estimate the levels of shared work values in a team[182]. It has also been used to calculate the degree of similarity of mental models in teams of air traffic controllers[125], where pairwise comparisons of all possible combinations of team members' ratings of task-related statements on a Likert-type scale where made, and then averaged to produce a similarity score for the mental model of the overall team.

In the case of the data from the Momento approach, and unlike some of the other approaches to aggregation, Euclidean distance can be calculated easily for longer task sequences. The resulting similarity scores reflect the order of individual tasks relative to all other tasks in the sequences (i.e., the absolute distance between ranks assigned to a task by two raters increases the further apart the ranks are in their sequence, regardless of tied ranks).

The problem of the “not required” task category can be overcome by assigning the maximum possible distance for the given number of sorted tasks to the pair of ranks where one of the tasks is classified as “not required” while the other one is assigned a rank.

Because of these advantages, I chose to use Euclidean distance to calculate the similarity scores for mental model of task sequence in OR teams in the subsequent validation study. In addition, due to its flexibility, I chose to use Krippendorff’s alpha[183] as an alternative approach to calculating similarity scores in the validation study. Also known as KALPHA, this measure of agreement is suitable for any number of raters, categories, scale values, or measures. It can be used to measure agreement for any type of data (nominal, ordinal, interval, ratio), enabling comparisons across different metrics. This statistic is suitable for any sample size and does not require a minimum. It can also be used for incomplete or missing data. The following general formula for Krippendorff’s alpha applies:

$$\alpha = 1 - \frac{Do}{De}$$

Where Do is the observed disagreement among values assigned to units of analysis:

$$Do = \frac{1}{n} \sum_c \sum_k o_{ck\ metric} \delta_{ck}^2$$

and De is the disagreement due to chance:

$$De = \frac{1}{n(n-1)} \sum_c \sum_k n_c * \ n_{k\ metric} \delta_{ck}^2$$

Although Krippendorff’s formulas appear somewhat complex, Krippendorff himself provides step-by-step guidance for the calculation of alpha for different types of measures in a working example[184]. The first step involves constructing a data matrix where m (1, i, j,…m) raters, value u (1, 2,…u…N) units (i.e., tasks in the procedure, in the case of the Momento) expressed as m_1, m_2,…m_u, m_N. The next step involves tabulating coincidences within units, or constructing a coincidence matrix, which accounts for all pairs of values found in *u:*

$$o_{ck} = \sum_u \frac{number\ of\ c-k\ pairs\ in\ unit\ u}{m_u - 1}$$

KALPHA accounts for different metrics or levels of measurement by weighing the observed and expected coincidences by the squared difference between the coinciding values. For ordinal metric differences, the values have the meaning of ranks and differences between ranks depend on how many ranks they are apart from each other:

$$\text{ordinal}\ \delta^2_{ck=} \left[\sum_{g=c}^{g=k} n_g - \frac{n_c+n_k}{2}\right]^2$$

As noted above, those tasks that have not been assigned ranks because of being deemed not required are not included in the calculation of KALPHA (i.e., they are by default treated as missing data). Treating the “not required” category from the results of the Momento task sort as missing data would inflate KALPHA values and would make individual mental models seem more similar than they would be if one of the members excluded one or more tasks as not necessary for the procedure and the other did not. In the case of the data on mental models generated via the Momento approach, tasks not ranked due to being considered “not required” for a given procedure are a part of the story. They are a part of an individual’s mental representation of which tasks need to (or not) be done for a successful management of a patient. Knowing that some team members deemed a task as unnecessary while others thought them necessary can provide richer information on the degree of overlap of their mental models. Instead of treating the information on the “not required” tasks as missing data, thus automatically excluding it from the calculation of KALPHA, those tasks rated as “not required” were assigned a rank of “0” in the validation study. Such an approach is not entirely free of flaws, especially in the case of short sequences of tasks where the value of KALPHA drops further down in the sequence the “not required” task is placed. For example, in a task sequence ranging from 1 to 5 when one team member assigns a rank of 5 to a task, while the other member ranks the same task as “not required” the value of KALPHA will be substantially lower compared to when one team member assigns a rank of “1” and the other team member a rank of “0” (not required) to the same task. One way to reduce this anomaly is to assign exponential or geometric weights to the “not required” category as its rank, relative to the rank of the other team member for the same task. However, this causes the “smoothing” of the task sequence such that the difference between the rank assigned by one team member and the “not required” “rank” assigned by the other team member for the same task is underestimated. For example, in a sequence of 5 tasks, we can achieve geometric progression for each successive task by multiplying the previous task by one half (for example, 0, 0.5, 1, 1.5, and 2 for a sequence from 1 to 5). These values are then assigned as weights to the rank to which the “not required” task is compared as the new value for the “not required” one. For example, comparing a sequence of 1-2-3-4-5 to a sequence of 1-2-3-4-“not required”, the “not required” value compared to rank “5” in the first sequence now becomes 5+2=7 when a weight of 2 is assigned to the rank of “5” and substituted for the “not required” value in the second sequence. Such smoothing of the sequence, where the new value for the “not required” category becomes closer in value to the comparing rank, does not appear optimal if we wish to distinguish the “not required” option as a substantial deviation from all other possible ranking options. The resulting similarity score would be inflated where having a disagreement over whether or not one or more of the key tasks are required at all should bear more weight than simply a disagreement over when the task should be done. Moreover, the re-ranking of the “not required” task becomes more complex when we wish to compare task sequences from multiple team members, as in the case of the entire OR team. In this project, assigning a rank of “0” to the “not required” tasks was thus a compromise between the two extremes: a) of excluding the category altogether; and b) undermining its importance where KALPHA was calculated to arrive at a similarity score for mental model of task sequence.

2.7.2. Calculating similarity scores for mental model of responsibility for task

The problem of how to compare ranks in those instances where one or more team members thought a task was not required for a given procedure was not relevant for calculating the similarity scores for mental model or responsibility for task. Here, the "not required" responses were simply coded as an extra nominal category (together with anaesthesia, nursing, and surgical subteam categories).

I considered percentage agreement as the most intuitive first step to calculating similarity scores in the case of mental model of responsibility for task. Percentage agreement is a popular method for calculating inter-rater consensus due to its common sense value and ease of calculation and interpretation[185, 186]. In this thesis, percentage agreement is defined as the proportion of occasions on which individual team members select the same subteam as each other as primarily responsible for a given task, multiplied by 100.

A study that reports complex statistics, but omits reporting raw percentage agreement may fail to inform readers at a practical level[187]. A study that reports multiple statistics, including raw percentage agreement, should however add to the validity of the analysis if it tells a consistent story. One disadvantage of calculating percentage agreement only, however, is that it does not take into account agreement that would be expected by chance (i.e., raters randomly assigning categories rather than actual agreement)[188]. Thus, for example, two raters would agree 33% of the time purely by chance when three rating categories are used, and 25% of the time when four categories are used.

Consequently, I considered Cohen's κappa[189] as another metric for calculating the similarity scores for mental model of responsibility for task. Cohen's kappa has been considered by far the most popular statistic of inter-rater agreement for nominal categories[190, 191] and it takes into account random agreement. It is the amount by which the observed agreement (P_o) exceeds that expected by chance alone (P_e), divided by the maximum which this difference could be:

$$k = \frac{P_o - P_e}{1 - P_e}$$

P_o is the observed percentage of occasions on which pairs of raters agree on assigning a case to the same category, and P_e is determined by using the observed data to calculate the probabilities of each rater randomly selecting each category. Kappa coefficient can range from -1 to 1, where a kappa of 1 would equate to perfect agreement, a value of 0 to chance agreement, and negative values would indicate agreement less than chance, or potential systematic disagreement between the observers[192].

There are claims that kappa can provide more information than a simple calculation of the raw proportion of agreement[192]. However, despite being the method of choice for calculating inter-rater agreement, the kappa coefficient is not free of limitations. For example, the method used in the calculation of kappa for trying to correct levels of observed agreement for an amount attributable to chance has been labelled arbitrary[193]. In addition, the same rating procedure may result in different and potentially very disparate values of kappa when the proportions of cases belonging to various categories vary from population to population[194]. This may make comparisons of results across studies and generalisations from the results of a single study difficult. Maclure[195] argued that kappa for two sets of observations is free of the abovementioned arbitrariness in the case where the multiple categories used by the two raters are natural or fixed by convention and have no inherent order; in that case, kappa may be the best measure of overall agreement. However, since Cohen’s kappa deals with only two raters, I decided to use Fleiss' extension of kappa, called the generalised kappa that accommodates for ratings from multiple raters[196], to calculate the similarity scores for mental model of responsibility for task in the Momento validation study. Fleiss' kappa, is defined as:

$$k = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e}$$

where the factor $1 - \bar{P}_e$ is the degree of agreement that is attainable above chance, and $\bar{P} - \bar{P}_e$ is the degree of agreement actually achieved above chance.

Fleiss' kappa involves first calculating the proportion P_j of all assignments of subject (i.e., in this case tasks in the procedure) i=1, … N to category (i.e., subteams selected as primarily responsible for a task) j=1, … k

$$P_j = \frac{1}{Nn} = \sum_{i=1}^{N} n_{ij}, \qquad 1 = \frac{1}{n}\sum_{j=1}^{k} n_{ij}$$

where n_{ij} is the number of raters who assigned category i to subject j. The next step involves calculating P_j - the extent of agreement between raters for subject i , or how many rater-rater pairs are in agreement, relative to the number of all possible rater-rater pairs:

$$P_i = \frac{1}{n(n-1)} \sum_{j=1}^{k} n_{ij}(n_{ij} - 1) = \frac{1}{n(n-1)} \sum_{j=1}^{k} (n_{ij}^2 - n_{ij}) = \frac{1}{n(n-1)} \left[\left(\sum_{j=1}^{k} n_{ij}^2 \right) - (n) \right]$$

The mean $\bar{P}$ of the P_i, and $\bar{P}_e$ in the formula for k are then computed as:

$$\bar{P} = \frac{1}{N} \sum_{i=1}^{N} P_i = \frac{1}{Nn(n-1)} (\sum_{i=1}^{N} \sum_{j=1}^{k} n_{ij}^2 - Nn)$$

and

$$\bar{P}_e = \sum_{j=1}^{k} p_j^2$$

In addition to percentage agreement and Fleiss' kappa, and due to its flexibility in dealing with different types of data, number of raters and categories, I decided to also use Krippendorff's alpha for nominal data to calculate the similarity scores for mental model of responsibility for task. The formulas described in the previous section of this chapter are used for calculating KALPHA for nominal data, minus the calculation for the ordinal metric differences. A working step-by-step example of how to calculate KALPHA for nominal data is provided by Krippendorff himself[184]. The exact way in which the similarity scores for mental models was calculated using this metric in the main validation study is described in more detail in Chapter 3 – Tool validation.

In the next chapter I describe the central study of my project, conducted in order to address the second objective of my thesis, which was to test specific assumptions that would provide initial support for the validity of the Momento approach.

3. TOOL VALIDATION

3.1. Purpose

As discussed in chapter 1, section 1.4, I chose to evaluate the new approach to assessment of shared mental models in OR teams using the unified approach to validity while focusing on content validity and concurrent validity. Content validity was addressed in the tool development stage (chapter 2, section 2.4). In this chapter, I describe a study designed to contribute to the process of validation of the Momento approach by testing specific assumptions that would support validity (see section 1.5, the second objective of the thesis). This was achieved through interpretation of a series of statistical tests conducted to examine the postulated relationships between relevant variables. In this study, data were analysed at different levels within an OR team, and using different metrics to calculate the scores for the degree of similarity of mental models. These analyses were conducted in order to examine the concurrent validity of the Momento approach.

The validation study took place on MORSim simulation study days. Approval for the study was obtained from the Central Regional Ethics Committee (CEN/12/03/002).

3.2. Method

The method for the validation study is described in the context of the MORSim simulation study days (see Chapter 2, section 2.1).

3.2.1. Participants, sample size and power

A total of 120 participants attended the MORSim simulation days (see section 3.1). They comprised 20 complete six-member teams each including: an anaesthetist and an anaesthetic technician (anaesthesia subteam); a surgeon and a surgical trainee (surgical subteam); and two nurses (nursing subteam). Teams were recruited from the Adult and

Trauma department at Auckland City Hospital and Middlemore Hospital general surgery. Participants were recruited using a first-come first-enrolled approach. On any one study day, the majority of participants were from the same hospital, and would have previously worked together.

There are no obvious data on which to base a sample size estimate for this work, as there is no previous literature on the relationship between the variables in the OR explored in the current study. The only empirical study exploring the similarity of mental models in the OR is the one described earlier by Burtscher et al.[56]. This study included 32 two-person anaesthesia teams participating in a single simulated scenario, but offered no rationale for the selected sample size. The sample size in the validation study followed the requirements of the MORSim study which was guided by previous simulation studies[163, 165].

Data on mental models were collected prior to two high-fidelity simulations representing realistic surgical cases run over the course of the day for each team, generating 240 individual responses to each measure over the course of the 40 simulated scenarios.

3.2.2. Setting, procedure and measures

Written informed consent was obtained from all participants prior to data collection. At the beginning of each simulation day, participants filled out a simple demographic questionnaire including questions on gender, clinical speciality, clinical experience, experience in the OR, and experience working with the other members of the team (i.e., "team familiarity"). The latter involved members of each multidisciplinary OR team reporting how often they worked with the other five members of their team in the past, on a descriptive scale ranging from 1("never before") to 4 ("we have worked together often").

Each team of six participants attended for one full MORSim day and took part in the three scenarios. Each scenario lasted approximately 40 minutes. The first and the last scenarios provided the vehicle for collecting the mental model data for the current study - as mentioned earlier, the items used in the task sorting tool were developed from the clinical requirements of these two simulated cases. Scenarios were presented in random order to control for time-of-day and order effects.

Before starting each scenario, participants received a demonstration on how the Momento computer application works by observing the researcher working through a simple example of a task sort, the hand–washing procedure used in the pilot study (see section 2.6). They were then given written case briefs on the upcoming case (see Appendix III) and the time to read them. Participants were then asked to apply the Momento task sorting tool to that case. They were verbally instructed to drag the cards depicting actions or tasks in the sequence in which they believed the tasks should be performed in the given case from the "Unsorted" column on the screen, and drop and sort them down the "MAIN ACTION SEQUENCE" column. They were also instructed to assign the subteam (i.e., anaesthesia, surgery, or nursing) they thought should be primarily responsible for each relevant task, by selecting the subteam from the pop up menu appearing next to each selected task. Finally, the participants were told they could sort tasks in parallel, using the "Parallel actions" columns on the screen, for each step in the procedure for those tasks they believed should be done simultaneously, regardless of the subteam responsible for the task. If they believed one or more tasks were not required for the given case, the participants were instructed to leave them in the "Unsorted" column on the screen. The task sort was completed when the participant clicked on the "Complete" button in the lower right hand corner of the screen (see Figure 7, section 2.5 in Chapter 2). If there were any tasks left unsorted, a pop up message automatically appeared asking the participant to confirm that he or she truly wanted to exclude the tasks from the sort. This was to ensure that the unsorted tasks were not simply overlooked by the participants, and therefore guard against missing data.

3.3. Data analysis

3.3.1. **Calculating scores for similarity of mental models**

For each participant, the Momento produced a set of numerical ranks for sequence of the 20 tasks (including tied ranks for those tasks sorted in parallel) and a subteam category for responsibility for each task. At its basic level of analysis, the similarity score was calculated for a pair of team members, where their task sequences and the subteams assigned to tasks were compared. A similarity score calculated using the scoring methods described in sections 2.7.1 and 2.7.2 was assigned to the pair for each of the two types of mental models – task sequence and responsibility for task.

3.3.1.1. ***Calculating similarity scores for task sequence***

The score for the degree of similarity of mental model of task sequence for a pair of team members was calculated in two steps:

(1) Individual ranks in the task sequence of each team member within an OR team were re-ranked so as to account for tied ranks, as described in section 2.6.1.2, Chapter 2. This was implemented in those cases where a participant assigned the same rank to multiple tasks if they believed those tasks should occur in parallel during the procedure, rather than sequentially. The re-ranking of tasks did not include those tasks participants considered as "not required" for the given procedure. These were at this stage in the analyses not assigned a rank.

(2) Comparisons of ranks for 20 tasks were conducted for pairs of participants using Euclidean distance and Krippendorff's alpha (KALPHA) for ordinal data (described in section 2.7.1.).

Euclidean distance was calculated using Microsoft Excel 2010. For each of the 20 ranked tasks in the sequence, this involved calculating the distance between the ranks assigned to a task by two team members. The resulting distances for each task were then squared and added up, and the square root of this sum represented the similarity score for the task sequence for the two team members. If one of the two team members classified a task as “not required” for the given procedure instead of assigning it a rank, then the Euclidean distance for the pair was the maximum possible distance between two ranks in the list of tasks. For a list of 20 tasks, this is calculated as the difference between the last and the first rank: 20-1=19. The distance scores for each task could vary between 0 (total agreement, where both participants assigned the same rank to the same task, including those instances where two “not required” options were selected for the same task), and 19 (total disagreement) for a pair of ranked tasks.

Euclidean distance is a measure of how far apart the scores are from each other. Thus, in order to convert distance scores to similarity scores, we first need to work out the maximum possible distance (or dissimilarity) score for the entire sequence of tasks. Such maximum disagreement is theoretically possible if one of the two team members whose ranks are being compared sorts all tasks as “not required” in a procedure. For a list of 20 tasks, this is $\sqrt{(20x19^2)} = 84.97$. The given distance score for two sets of ranks is then divided by their maximum possible distance and the result subtracted from 1:

$$\text{similarity score} = 1\text{-}\left(\frac{\text{distance (x, y)}}{\text{Max distance(x,y)}}\right)$$

This results in rescaled Euclidean distance scores which now range from 0 (total disagreement) to 1 (total agreement or similarity). Such rescaling can be applied to any number of tasks in a sequence.

The similarity score for the whole OR team for task sequence was the mean rescaled Euclidean distance of all possible pairwise combinations of rank sequences within the OR team (i.e., 15, in the case of 6-member teams).

To calculate KALPHA, a macro designed by Hayes and Krippendorff[197] for computing KALPHA for subjective judgments made at any level of measurement, any number of judges, with or without missing data, was imported and used in IBM SPSS version 21. The calculation of KALPHA to arrive at the similarity score for the entire OR team involved simultaneously comparing mental models of all six members of the team by tabulating and comparing the differences between all pairs of rankings assigned to each task (as described in section 2.7.1). As discussed in detail in section 2.7.1, those tasks that an individual omitted as "not required" in the given procedure were assigned a rank of 0 when calculating similarity scores using this method.

3.3.1.2. *Calculating similarity scores for responsibility for task*

For the assessment of the similarity of mental models of responsibility for task, I took the following approach to scoring. For each of the 20 tasks, pairwise comparisons involved assigning a score of 1 if the two individuals assigned the task to the same subteam (e.g., anaesthesia). Alternatively, a score of 0 was assigned if the two individuals did not assign the task to the same subteam, or if one of the individuals did not assign a subteam to a task due to believing the task to be "not required" in the given procedure.

Where percentage agreement was used to calculate similarity scores, the total similarity score for a pair of team members was the mean of individual similarity scores for 20 tasks. Thus, the total similarity score could range from 0 (total disagreement) to 1 (total agreement) when expressed as proportions, or 1 to 100 when expressed as percentages.

Fleiss' kappa was calculated using a Microsoft Excel spreadsheet created for this purpose[198] (available for download at http://www.ccit.bcm.tmc.edu/jking/homepage). This macro takes into account multiple raters of multiple tasks, where the degree of observed agreement is calculated as described in section 2.7.2.

To calculate the similarity scores using KALPHA, I used the same SPSS macro as the one used to calculate KALPHA for the similarity of task sequence (see section 3.3.1.1), but with "nominal" as the data type option.

In the case of both Fleiss' kappa and KALPHA, a '"not required" response was treated as a separate, fourth nominal category.

3.3.2. Calculating team familiarity scores

As described in section 3.2.2, within each OR team, members reported on how often they worked with the other five members of their team in the past, on an ordinal scale ranging from 1("never before") to 4 ("we have worked together often"). This resulted in five individual ratings by each participant of their fellow team members. These ratings were averaged over participants belonging to the same subteam so that, for each pair of participants in a subteam (e.g., anaesthetist and anaesthetic technician for the anaesthesia subteam), the mean of their ratings of each other was taken as that subteam's average familiarity score. To arrive at the mean OR team-level familiarity score, the average of all possible familiarity scores for pairs of team members was calculated.

3.3.3. Assessing validity – levels of analysis

In order to test the assumptions that would provide initial support for the validity of the Momento approach (see sections 1.5 and 3.1), similarity scores for mental models (for both task sequence and responsibility for task) were calculated within each OR team at three main levels of analysis:

1) Overall OR team:

a) When Euclidean distance was used to calculate the similarity score for mental model of task sequence, and percentage agreement was used to calculate the similarity score for responsibility for task, the similarity score for each OR team was calculated by averaging all possible pairwise comparisons within the OR team.

b) When KALPHA for ordinal data was used to score similarity for task sequence, and Fleiss' kappa and KALPHA for nominal data used to score similarity for responsibility for task, the similarity scores of all members of the OR team were simultaneously compared.

2) Intra-team:

a) **for each subteam** – where pairwise comparisons of ranks and categories for similarity scores were conducted separately for:

i) Anaesthesia subteam (A) - consisting of an anaesthetist (AN) and an anaesthetic technician (AT)

ii) Nursing subteam (N) - consisting of Nurse A (Na) and Nurse B (Nb), and

iii) Surgical subteam (S) - consisting of a consultant surgeon (Sa), and a junior surgeon (Sb).

This resulted in a separate similarity score for each individual subteam within each OR team, for each type of mental model and scenario.

b) **overall intra-team** – where the score for similarity of mental model was the mean of intra-team scores for the three subteams within that OR team, for each type of model and scenario.

3) Inter-team:

a) for comparisons with members from other subteams within the OR team

When similarity score was calculated using Euclidean distance and percentage agreement, the following pairwise comparisons were made:

i) Anaesthesia subteam with nursing subteam (AN) – this was the mean similarity score for pairings A-Na, A-Nb, AT-Na, and AT-Nb

ii) Anaesthesia subteam with surgery subteam (AS) – i.e., mean similarity score for A-Sa, A-Sb, AT-Sa, and AT-Sb pairings; and

iii) Nursing subteam and surgery subteam (NS) – i.e., mean similarity score for Na-Sa, Na-Sb, Nb-Sa, and Nb-Sb pairings

This resulted in a separate similarity score for each of the three inter-team groupings (i.e., AN, AS, and NS) within each OR team.

When similarity was calculated using Fleiss' kappa and KALPHA, the inter-team comparisons involved simultaneously comparing the ranks and categories assigned by all the members within each of the three intra-team groupings (i.e., AN, AS, NS), rather than calculating the mean of the corresponding pairwise comparisons.

b) **overall inter-team** – where the similarity score was the mean of inter-team scores for the three inter-team groupings in an OR team.

Table 5 provides a summary of all the levels of analysis used in the validation study.

Table 5. A summary of the levels at which data were analysed within each OR team.

Groupings	Pairwise comparisons (Euclidean distance, percentage agreement)*	Total individual similarity scores for comparison (for 20 teams)
OR team	mean of A-AT, Na-Nb, Sa-Sb, AN, AS, and NS	300
Intra-team	Individual subteam	
	anaesthesia **A** (A-AT)	20
	surgery **S** (Sa-Sb)	20
	nursing **N** (Na-Nb)	20
	Overall intra-team	
	mean of A-AT, Na-Nb, and Sa-Sb	60
Inter-team	Individual inter-team	
	AN (mean of A-Na, A-Nb, AT-Na, and AT-Nb)	80
		80
	AS (mean of A-Sa, A-Sb, AT-Sa, and AT-Sb	80
	NS (mean of Na-Sa, Na-Sb, Nb-Sa, and Nb-Sb)	240
	Overall inter-team	
	mean of AN, AS and NS	

*When KALPHA and Fleiss' kappa were used to score similarity of mental model, the similarity scores of all members of the OR team were simultaneously compared, as well as those for overall intra-team and overall inter-team groupings.

3.3.3.1. *Assessing the relationship between similarity scores for mental models and team familiarity scores*

In order to address the second objective of this thesis – beginning the process of validation of the new Momento approach - I examined the link between the similarity scores for mental models and team familiarity scores. The assumption that would support validity here is that OR teams whose members have worked with each other more frequently in the past are also likely to have more similar scores on mental models of the sequence of tasks and responsibility for task, as assessed by the new Momento tool. A Spearman's rank-order correlation was calculated to test this assumption. Mean familiarity scores for the OR teams were compared to mean similarity scores for the OR team, for each type of mental model and for each scenario.

3.3.3.2. ***Assessing the relationship between intra-team and inter-team similarity scores for mental models***

The relationship between the similarity scores for mental model and subteam groupings was tested in order to provide further support for the validity of the Momento approach. Here, the assumption was that OR subteams would have more similar scores for mental models of task sequence and responsibility for task within their own subteam, than when compared inter-team.

First, the intra-team versus inter-team relationship was examined at the level of the OR team. This involved comparing the overall intra-team similarity score to the overall inter-team similarity score within each OR team. Second, intra-team similarity scores for each individual subteam were compared to inter-team similarity scores of that subteam with other subteams. Thus, the intra-team scores for the subteam A were compared to the inter-team scores for the combined AS and AN groupings, respectively; subteam S scores were compared to AS and to NS scores; and subteam N scores were compared to NS and to AN scores (see Table 5). It could be expected that members of a professional subteam, such as those working closely together in anaesthesia, might have greater similarity scores for their mental models than might be expected between one subteam and another, for example comparing anaesthesia with nursing.

To test the above relationships, a paired samples t-test was calculated in the cases where the similarity scores were normally distributed, and a Wilcoxon signed rank test where the data did not satisfy the normality criteria. Normality of the distribution of differences of intra-team versus inter-team means was assessed used the Shapiro-Wilk test[199], which was found to be superior to nine other statistical procedures for evaluating the normality of a sample[200]. The null hypothesis for this test is that the data are normally distributed. Therefore, if the *p*-value < 0.05, then the null hypothesis is rejected and there is evidence that the data tested are not from a normally distributed population.

3.3.3.3. ***Comparing multiple methods for calculating similarity scores for mental models***

As discussed in Chapter 1, section 1.4, one of the different ways to demonstrate that a given assessment approach actually captures the same cognitive structures of interest is to test the assumption that the same relationships between variables could be found using different scoring methods. Therefore, for each method used to calculate the score for the degree of similarity of mental model (described in sections 2.7.1 and 2.7.2), I examined the relationship between similarity scores for mental model and team familiarity score, and intra-team versus inter-team similarity scores for mental model. Spearman's rank-order correlation was calculated to explore the degree to which similarity scores calculated using different scoring methods were correlated for each type of mental model and each scenario.

3.4. Results

3.4.1. **Demographics**

One hundred-and-twenty participants (20 consultant and 20 junior surgeons; 20 consultant anaesthetists or senior anaesthetic fellows and 20 anaesthetic technicians; and 40 nurses) completed the task sort related to scenario 1 (laparotomy for an abdominal stab wound). The data from one participant (Nurse A from team 11) was subsequently excluded from the analysis of data from scenario 1 because she misunderstood the instructions for completing the task sorting exercise. The task sort for scenario 2 was completed by 119 participants because a junior surgeon had to leave the course early for personal reasons.

Within each OR team, there were more females (62.5%) than males. Figure 14 shows the participants' mean self-reported clinical experience and experience in the OR for each of the six members of an OR team.

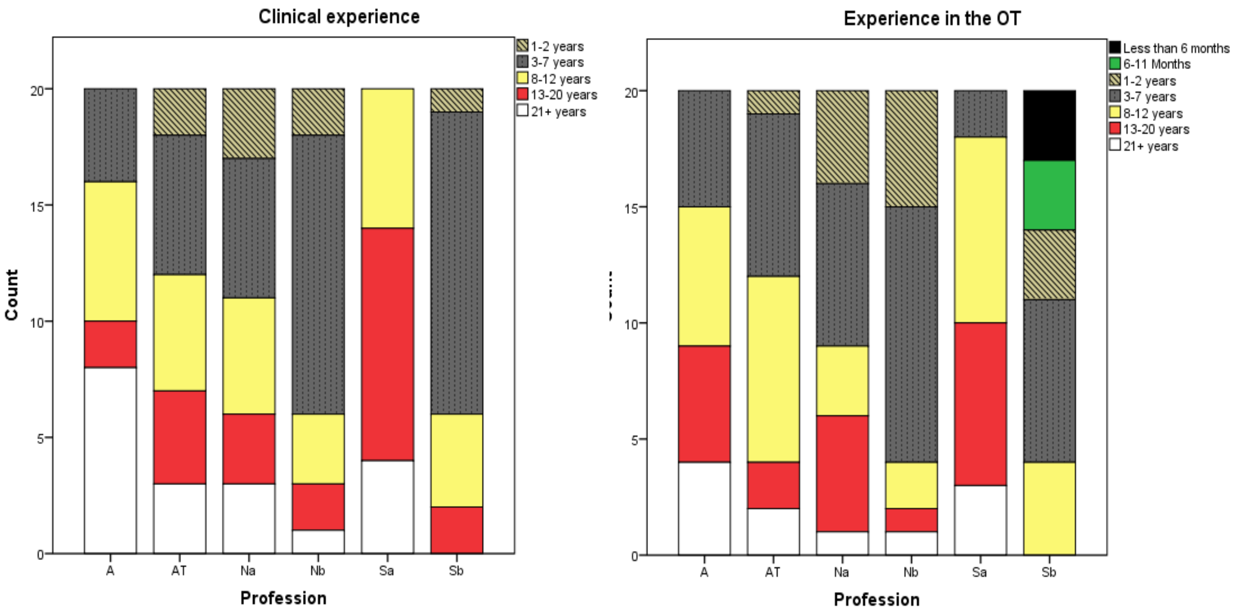


Figure 14. Participants' mean self-reported years of clinical experience and experience in the OR

3.4.2. **Similarity scores for mental models and team familiarity scores**

The descriptive statistics for the similarity scores for task sequence calculated using rescaled Euclidean distance and KALPHA for ordinal data for scenario 1 and 2 are summarised in Table 6. The descriptive statistics for the similarity scores for responsibility for task for Scenario 1 and 2 calculated using percentage agreement, Fleiss kappa, and KALPHA for nominal data are summarised in Table 7.

Table 6. Summary statistics for the similarity scores for mental model of task sequence for scenario 1 and 2 for the OR team, intra-team, and inter-team. Mean similarity scores are expressed as rescaled Euclidean distance and KALPHA for ordinal data.

Task sequence	Scenario 1 Laparotomy for an abdominal stab wound							Scenario 2 Laparotomy for a suspected perforated viscus						
				95% CI							95% CI			
	Mean	SD	Std error of the mean	Lower bound	Upper bound	Min	Max	Mean	SD	Std error of the mean	Lower bound	Upper bound	Min	Max
OR team														
Rescaled Euclidean distance	0.77	0.04	0.01	0.75	0.79	0.68	0.86	0.83	0.04	0.01	0.81	0.85	0.74	0.86
KALPHA	0.72	0.11	0.02	0.67	0.77	0.45	0.88	0.83	0.07	0.02	0.80	0.87	0.68	0.93
Overall intra-team														
Rescaled Euclidean distance	0.78	0.05	0.01	0.76	0.80	0.67	0.86	0.83	0.04	0.01	0.81	0.85	0.74	0.88
KALPHA	0.75	0.10	0.02	0.71	0.80	0.50	0.89	0.84	0.07	0.02	0.81	0.88	0.68	0.93
A subteam														
Rescaled Euclidean distance	0.76	0.07	0.02	0.73	0.79	0.60	0.93	0.83	0.08	0.02	0.79	0.87	0.65	0.95
KALPHA	0.72	0.17	0.04	0.64	0.80	0.15	0.97	0.83	0.15	0.03	0.76	0.90	0.40	0.98
S subteam														
Rescaled Euclidean distance	0.80	0.08	0.02	0.77	0.84	0.61	0.92	0.83	0.07	0.02	0.80	0.86	0.67	0.90
KALPHA	0.79	0.14	0.03	0.72	0.85	0.55	0.97	0.84	0.10	0.02	0.80	0.89	0.61	0.95
N subteam														
Rescaled Euclidean distance	0.79	0.08	0.02	0.75	0.83	0.65	0.90	0.83	0.09	0.02	0.79	0.87	0.56	0.92
KALPHA	0.76	0.18	0.04	0.67	0.85	0.36	0.96	0.85	0.14	0.03	0.78	0.92	0.34	0.97
Overall inter-team														
Rescaled Euclidean distance	0.77	0.04	0.01	0.75	0.79	0.67	0.86	0.83	0.04	0.01	0.81	0.85	0.74	0.86
KALPHA	0.72	0.10	0.02	0.68	0.77	0.46	0.88	0.83	0.07	0.02	0.80	0.87	0.68	0.93
AN														
Rescaled Euclidean distance	0.78	0.05	0.01	0.76	0.81	0.68	0.87	0.83	0.05	0.01	0.81	0.85	0.69	0.89
KALPHA	0.74	0.13	0.03	0.68	0.80	0.42	0.90	0.83	0.10	0.02	0.79	0.88	0.57	0.93
AS														
Rescaled Euclidean distance	0.77	0.06	0.01	0.74	0.79	0.66	0.85	0.83	0.06	0.01	0.81	0.86	0.68	0.91
KALPHA	0.73	0.12	0.03	0.67	0.78	0.51	0.90	0.83	0.10	0.02	0.78	0.88	0.64	0.95
NS														
Rescaled Euclidean distance	0.76	0.06	0.01	0.73	0.78	0.62	0.86	0.82	0.05	0.01	0.80	0.85	0.69	0.88
KALPHA	0.71	0.12	0.03	0.65	0.77	0.37	0.89	0.84	0.08	0.02	0.80	0.87	0.60	0.93

Note: A=anaesthesia; S=surgical; N=nursing; AN=anaesthesia and nursing subteam combined; AS=anaesthesia and surgical subteam; NS=nursing and surgical subteams combined

Table 7. Summary statistics for the similarity scores for mental model of responsibility for task for scenario 1 and 2 for the OR team, intra-team, and inter-team. Mean similarity scores are expressed as percentage agreement, Fleiss kappa, and KALPHA for nominal data.

Task sequence	Scenario 1 Laparotomy for an abdominal stab wound							Scenario 2 Laparotomy for a suspected perforated viscus						
				95% CI for mean							95% CI for mean			
	Mean	SD	Std error of the mean	Lower bound	Upper bound	Min	Max	Mean	SD	Std error of the mean	Lower bound	Upper bound	Min	Max
OR team														
Percentage agreement	0.69	0.05	0.01	0.67	0.71	0.62	0.78	0.72	0.06	0.01	0.69	0.74	0.61	0.82
Fleiss' kappa	0.53	0.07	0.02	0.50	0.57	0.43	0.67	0.57	0.08	0.02	0.54	0.61	0.40	0.73
KALPHA	0.54	0.07	0.02	0.51	0.57	0.44	0.67	0.58	0.08	0.02	0.54	0.62	0.44	0.74
Overall intra-team														
Percentage agreement	0.74	0.06	0.01	0.71	0.76	0.60	0.83	0.75	0.05	0.01	0.73	0.77	0.67	0.83
Fleiss' kappa	0.60	0.08	0.02	0.56	0.63	0.38	0.71	0.62	0.07	0.02	0.58	0.65	0.49	0.73
KALPHA	0.60	0.08	0.02	0.57	0.64	0.40	0.72	0.63	0.07	0.02	0.59	0.66	0.50	0.73
A														
Percentage agreement	0.76	0.11	0.02	0.71	0.81	0.55	0.90	0.78	0.07	0.02	0.75	0.81	0.65	0.90
Fleiss' kappa	0.63	0.16	0.04	0.56	0.70	0.33	0.84	0.66	0.11	0.02	0.61	0.71	0.49	0.85
KALPHA	0.64	0.15	0.03	0.57	0.71	0.35	0.85	0.67	0.10	0.02	0.62	0.72	0.50	0.85
S														
Percentage agreement	0.73	0.10	0.02	0.68	0.76	0.50	0.90	0.73	0.08	0.02	0.69	0.76	0.55	0.85
Fleiss' kappa	0.58	0.16	0.04	0.50	0.65	0.20	0.85	0.57	0.12	0.03	0.52	0.63	0.31	0.77
KALPHA	0.59	0.16	0.03	0.51	0.66	0.22	0.85	0.58	0.11	0.03	0.53	0.64	0.33	0.78
N														
Percentage agreement	0.72	0.09	0.02	0.68	0.77	0.50	0.85	0.75	0.08	0.02	0.71	0.78	0.60	0.85
Fleiss' kappa	0.58	0.14	0.03	0.51	0.65	0.22	0.76	0.61	0.12	0.03	0.55	0.67	0.37	0.77
KALPHA	0.58	0.14	0.03	0.52	0.65	0.24	0.77	0.62	0.12	0.03	0.56	0.68	0.39	0.78

Table 7 continued

Task sequence	Scenario 1 Laparotomy for an abdominal stab wound							Scenario 2 Laparotomy for a suspected perforated viscus						
				95% CI for mean							95% CI for mean			
	Mean	SD	Std error of the mean	Lower bound	Upper bound	Min	Max	Mean	SD	Std error of the mean	Lower bound	Upper bound	Min	Max
Overall inter-team														
Percentage agreement	0.68	0.05	0.01	0.65	0.70	0.59	0.78	0.71	0.06	0.01	0.68	0.74	0.58	0.83
Fleiss' kappa	0.54	0.07	0.02	0.51	0.57	0.43	0.67	0.58	0.08	0.02	0.54	0.62	0.46	0.73
KALPHA	0.55	0.07	0.02	0.52	0.58	0.43	0.67	0.59	0.08	0.02	0.55	0.62	0.47	0.73
AN														
Percentage agreement	0.72	0.08	0.02	0.68	0.75	0.56	0.84	0.74	0.09	0.02	0.70	0.78	0.46	0.85
Fleiss' kappa	0.58	0.11	0.02	0.53	0.63	0.35	0.75	0.62	0.11	0.02	0.57	0.67	0.34	0.77
KALPHA	0.59	0.11	0.02	0.54	0.64	0.36	0.76	0.62	0.11	0.02	0.57	0.67	0.35	0.77
AS														
Percentage agreement	0.67	0.07	0.01	0.64	0.70	0.51	0.76	0.71	0.07	0.02	0.68	0.75	0.61	0.85
Fleiss' kappa	0.54	0.08	0.02	0.51	0.58	0.37	0.66	0.59	0.09	0.02	0.54	0.63	0.47	0.79
KALPHA	0.55	0.07	0.02	0.51	0.58	0.38	0.66	0.59	0.10	0.02	0.55	0.64	0.47	0.85
NS														
Percentage agreement	0.64	0.06	0.01	0.61	0.67	0.50	0.75	0.68	0.07	0.02	0.64	0.71	0.51	0.88
Fleiss' kappa	0.50	0.08	0.02	0.47	0.54	0.36	0.62	0.54	0.09	0.02	0.50	0.58	0.37	0.80
KALPHA	0.51	0.08	0.02	0.47	0.54	0.37	0.62	0.54	0.09	0.02	0.50	0.59	0.38	0.80

Note: A = anaesthesia; S = surgical; N = nursing; AN = anaesthesia and nursing subteam combined; AS = anaesthesia and surgical subteam; NS = nursing and surgical subteams combined

Table 8 summarises the descriptive statistics for the variable “team familiarity” for the OR team. No significant correlation was found between team familiarity scores and similarity scores for mental model of task sequence in scenario 1 (Spearman’s rho = -0.304, p = 0.176) or scenario 2 (Spearman’s rho = - 0.070, p = 0.769) when similarity scores were calculated using rescaled Euclidean distance. The relationship was also found to not be statistically significant when similarity scores were calculated using KALPHA for ordinal data in scenario 1 (Spearman’s rho = -0.228, p = 0.333) and scenario 2 (Spearman’s rho = 0.069, p = 0.774).

There was also no significant correlation between team familiarity scores for the OR team and similarity scores for mental model of responsibility for task in scenario 1 (Spearman’s rho = 0.170, p = 0.474) or scenario 2 (Spearman’s rho = -0.040, p = 0.867) when percentage agreement was used to calculate similarity. Similarly, no significant correlation was found between team familiarity scores when either Fleiss’ kappa was used to calculate similarity scores in scenario 1 (Spearman’s rho = 0.198, p = 0.402) and scenario 2 (Spearman’s rho = -0.069, p = 0.774), or KALPHA for nominal data in scenario 1 (Spearman’s rho = 0.198, p = 0.402) and scenario 2 (Spearman’s rho = 0.069, p = 0.774).

No statistically significant relationship was found between the similarity scores for mental model and team familiarity scores at intra- team level, inter-team level, or for individual professional groups for either type of mental model or scenario (p<0.05).

Table 8. Summary of statistics for the variable "team familiarity" at OR team level (N=20).

"Team familiarity"			95% Confidence Interval for Mean			
	Mean	SD	Lower Bound	Upper Bound	Min	Max
Mean OR team	**2.60**	**0.36**	**2.56**	**2.63**	**1.93**	**3.33**
Mean intra-team	**2.99**	**0.59**	**2.72**	**3.27**	**2.00**	**3.83**
A	3.05	0.93	2.61	3.49	1.00	4.00
S	2.50	1.19	1.94	3.06	1.00	4.00
N	3.43	0.69	3.10	3.75	2.00	4.00
Mean inter-team	**2.51**	**0.37**	**2.34**	**2.68**	**1.83**	**3.24**
AN	3.07	0.59	2.80	3.35	1.50	4.00
AS	2.16	0.36	1.99	2.33	1.63	2.88
NS	2.30	0.57	2.03	2.57	1.50	3.88

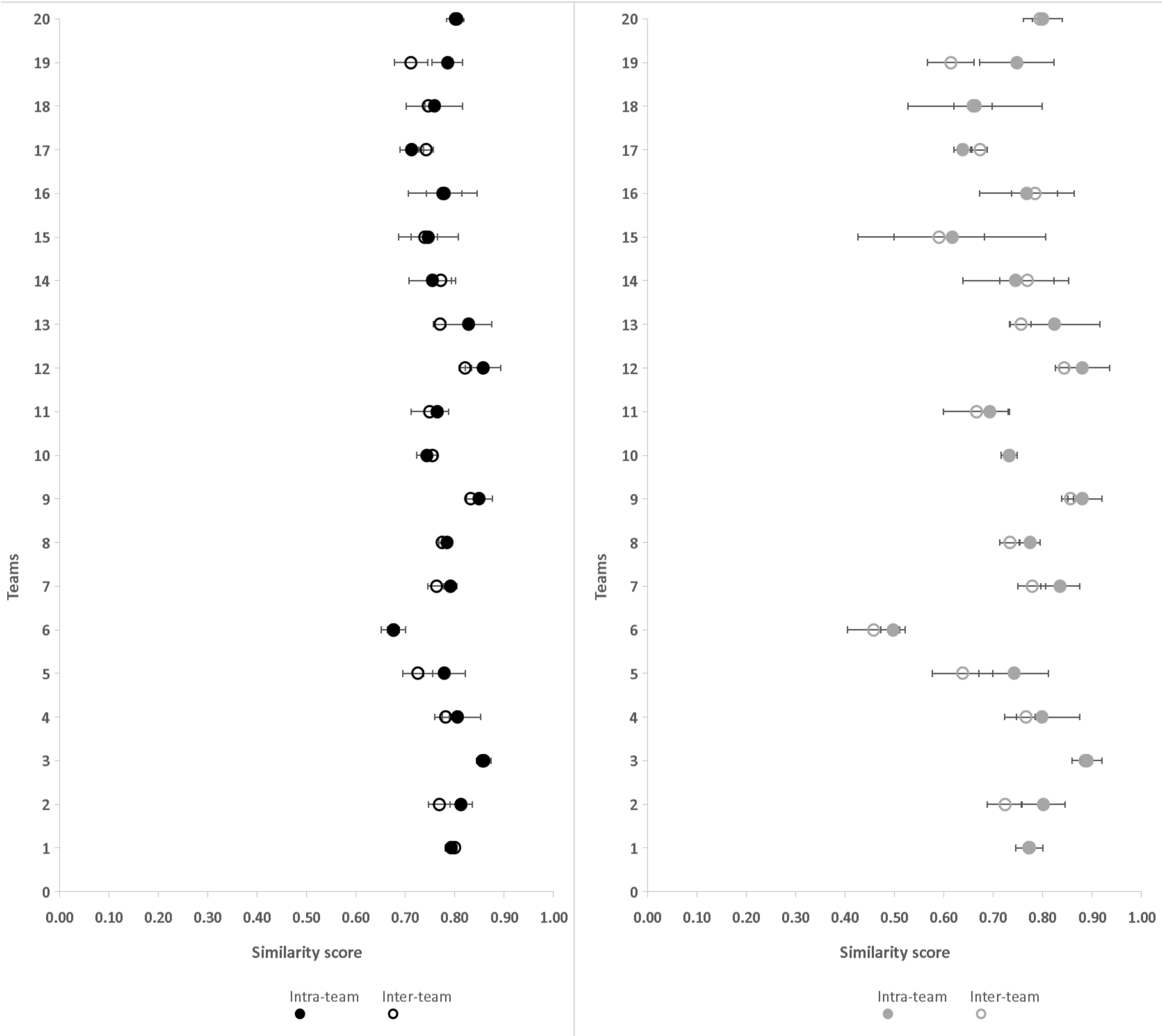
Note: Team familiarity score was an average response to the question "How long have you worked with the other participants?" and was coded as 1 = "Never before", 2 = "Maybe once or twice before"; 3="A number of times before"; and 4="We have worked together often".

3.4.3. Similarity scores for mental model and subteams

3.4.3.1. *Similarity scores for mental model of task sequence intra- versus inter-team*

The difference scores for the overall intra-team and inter-team similarity using both rescaled Euclidean distance and KALPHA for ordinal data in Scenario 1 were normally distributed, as assessed by the Shapiro-Wilk test (p=0.598, and p=0.384 , respectively). In the case of scenario 1, similarity scores for mental model of task sequence were significantly greater intra-team (0.78±0.01) than inter-team (0.77±0.01), when similarity was calculated using rescaled Euclidean distance $t(19)$=2.513, p=0.021, *Cohen's d*=0.56, mean difference

0.02, 95% confidence interval 0.003 to 0.03. Similarly, a significant relationship was found using KALPHA for ordinal data to calculate similarity scores intra-team (0.75±0.02) and inter-team (0.72±0.02), $t(19)=3.222$, $p=0.04$, *Cohen's d*=0.72, mean difference 0.03 , 95% interval 0.01 to 0.05). Figure 15 shows overall intra-team and inter-team scores and 95% confidence intervals calculated using rescaled Euclidean distance (a) and KALPHA for ordinal data (b) for individual teams (n=20) in scenario 1.



a) Rescaled Euclidean distance **b) KALPHA**

Figure 15. Mean overall intra-team versus inter-team scores using a) rescaled Euclidean distance and b) KALPHA for ordinal data with 95% confidence intervals for similarity of mental model of task sequence, for scenario 1 (laparotomy for an abdominal stab wound); p=0.021. The further apart the circles, the larger the mean difference in similarity scores intra-team versus inter-team.

In the case of scenario 2, the difference scores for the overall intra-team and inter-team similarity were not normally distributed, as assessed by the Shapiro-Wilk test (p=0.002 for rescaled Euclidean distance, and p=0.006 for KALPHA for ordinal data), and by the inspection of skewness and kurtosis values (1.94 and 5.33 respectively for rescaled Euclidean distance, and 1.62 and 3.19 for KALPHA for ordinal data). As a result, the nonparametric Wilcoxon Signed-Rank test was conducted to determine if there were differences between similarity scores for task sequence intra-team versus inter-team. No statistically significant difference in the overall similarity scores intra-team (*Mdn*=0.84) versus inter-team (*Mdn*=0.85) was found when rescaled Euclidean distance was used to calculate the similarity scores (z=-0.373, p=0.709), or when KALPHA was used to compare overall intra-team (*Mdn*=0.85) with overall inter-team (*Mdn*=0.86) scores (z=-1.381, p=0.167). Table 9 shows the statistics for the intra-versus inter-team comparisons for mental model of task sequence for both scenarios.

3.4.3.1.1. *Subteam versus inter-team comparisons of scores for mental model of task sequence for individual subteams*

In the case of the **anaesthesia subteam (A)**, their subteam similarity scores were compared to the inter-team similarity scores for anaesthesia and nursing subteams (**AN**), and to the inter-team scores for anaesthesia and surgical subteams (**AS**), respectively. As shown in Table 9, no significant differences were found between the similarity scores of anaesthesia subteams compared to the AN inter-team scores, when either rescaled Euclidean distance or KALPHA was used, in either scenario. Similarly, no significant differences were found between the anaesthesia subteam scores and AS inter-team scores, for either rescaled Euclidean distance or KALPHA were used to score similarity, in either scenario.

In the case of the **surgical subteam (S)**, their individual subteam similarity scores were compared to the inter-team scores for surgical and anaesthesia subteams (**AS**), and to the inter-team similarity scores of surgical and nursing subteams (**NS**), respectively. In scenario 1, surgical subteams had significantly higher similarity scores for mental models of task sequence within their own subteam than the surgical and anaesthesia subteams combined, when both rescaled Euclidean distance and KALPHA were used to calculate similarity. Similarity scores for mental models of surgical subteams in scenario1 were also significantly higher for the subteam than for surgical and nursing subteams together. This was the case regardless of the metric used to calculate similarity.

In the case of scenario 2, there were no significant differences between the similarity scores for surgical subteams and inter-team scores for surgical and anaesthesia subteams, regardless of the metric used to score similarity. Similarly, no statistically significant difference was found between the similarity scores for surgical subteams and inter-team scores of the surgical and nursing subteams. Again, this was the case when both rescaled Euclidean distance and KALPHA were used to arrive at similarity scores.

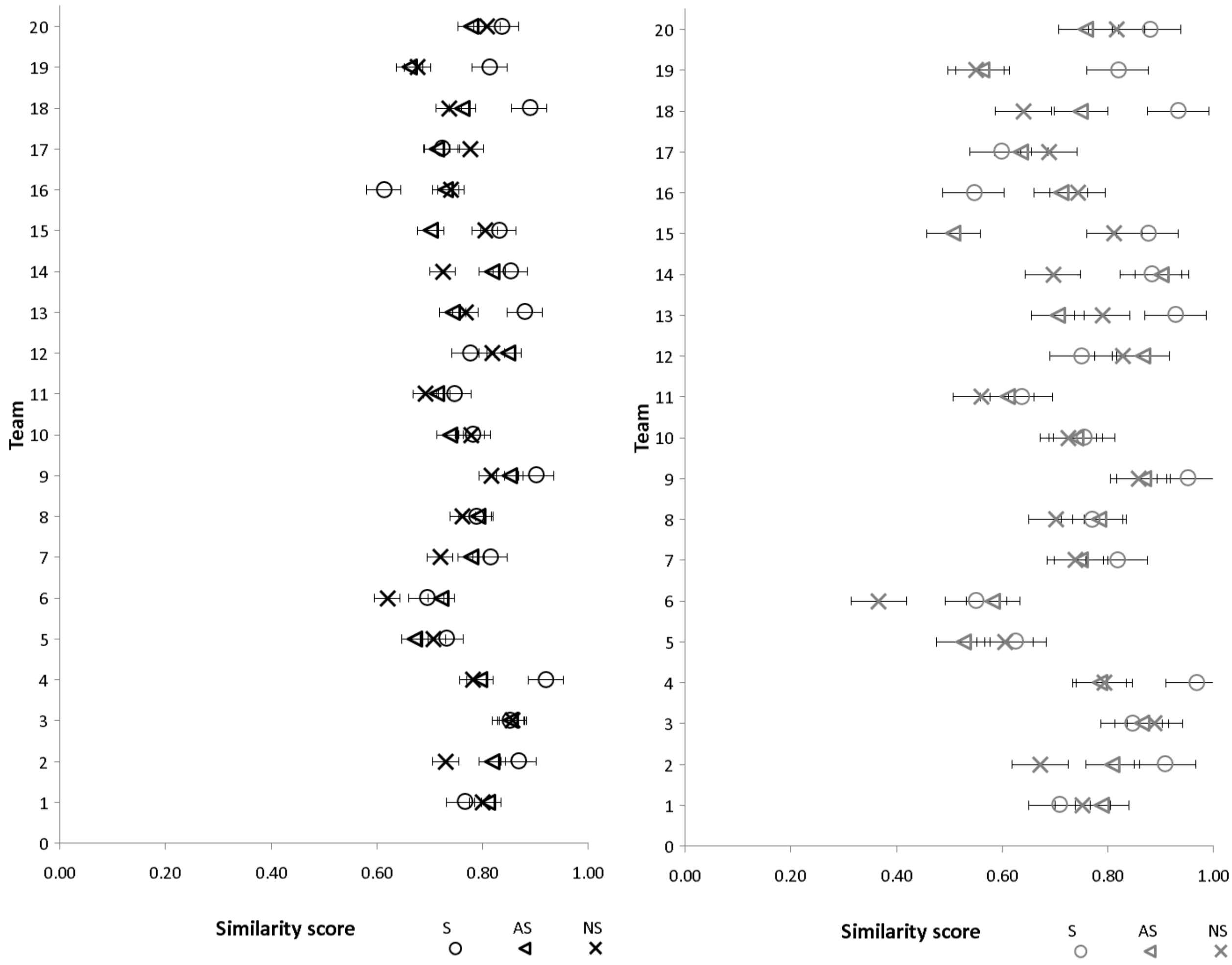
In the case of the **nursing subteam (N)**, their individual subteam similarity scores for task sequence were compared to the inter-team scores for nursing and anaesthesia subteams (**AN**), and to the inter-team scores for nursing and surgical subteams (**NS**), respectively. For both scenario 1 and scenario 2, no significant differences were found between the similarity scores of nursing subteams and the inter-team similarity scores of nursing and anaesthesia subteams, regardless of the method used to score similarity. Similarly, no significant differences were found between the similarity scores of nursing subteams and inter-team similarity scores of nursing and surgical subteams in either scenario, and regardless of whether rescaled Euclidean distance or KALPHA were used to score similarity.

The statistically significant intra- versus inter-team differences in the similarity scores for mental model of task sequence between the surgical and anaesthesia, and surgical and nursing subteams in scenario 1 are depicted graphically in Figure 16.

Table 9. Statistics for the intra-versus inter-team comparisons of similarity scores for mental model of task sequence for scenario 1 and 2. All statistics are shown only where significant relationships were found.

	Scenario 1 (laparotomy for an abdominal stab wound)							Scenario 2 (laparotomy for a perforated viscus)						
Similarity for task sequence	Mean difference	Std error of the mean	t-test	p value	95% CI		Effect size (Cohen's d)	Mean difference	Std error of the mean	t-test	p value	95% CI		Effect size (Cohen's d)
					Lower	Upper						Lower	Upper	
Overall intra- vs inter-team														
Rescaled Euclidean distance	0.02	0.01	2.513	0.021*	0.003	0.03	0.56	0.004	0.004		0.709	-0.01	0.01	
KALPHA	0.03	0.01	3.222	0.004*	0.01	0.05	0.72	0.009	0.005		0.167	-0.002	0.02	
Individual intra- vs inter-team														
Anaesthesia subteam:														
A-AS														
Rescaled Euclidean distance	-0.01	0.01		0.502	-0.04	0.03		-0.0001	0.01		0.841	-0.03	0.02	
KALPHA	-0.004	0.03		0.897	-0.07	0.06		0.02	0.02		0.921	-0.05	0.05	
A-AN														
Rescaled Euclidean distance	-0.02	0.01		0.083	-0.05	0.003		0.007	0.02		0.996	-0.04	0.04	
KALPHA	-0.01	0.03		0.585	-0.07	0.04		0.008	0.03		0.370	-0.06	0.05	
Surgical subteam:														
S-AS														
Rescaled Euclidean distance	0.04	0.02	2.362	0.029*	0.004	0.07	0.52	0.004	0.01		0.992	-0.02	0.02	
KALPHA	0.06	0.03	2.070	0.052*	-0.001	0.12	0.46	0.02	0.02		0.387	-0.02	0.06	
S-NS														
Rescaled Euclidean distance	0.05	0.02	2.758	0.013*	0.01	0.08	0.60	0.02	0.01		0.778	-0.02	0.04	
KALPHA	0.08	0.03	2.660	0.015*	0.02	0.14	0.60	0.01	0.02		0.809	-0.04	0.06	
Nursing subteam:														
N-AN														
Rescaled Euclidean distance	0.01	0.01		0.562	-0.02	0.04		0.004	0.01		0.794	-0.02	0.03	
KALPHA	0.03	0.03		0.391	-0.04	0.09		0.02	0.02		0.179	-0.03	0.06	
N-NS														
Rescaled Euclidean distance	0.03	0.02		0.079	-0.004	0.06		0.02	0.02		0.445	-0.02	0.04	
KALPHA	0.04	0.03		0.216	-0.03	0.10		0.01	0.02		0.563	-0.03	0.06	

* $p<0.05$



a) Rescaled Euclidean distance **b) KALPHA**

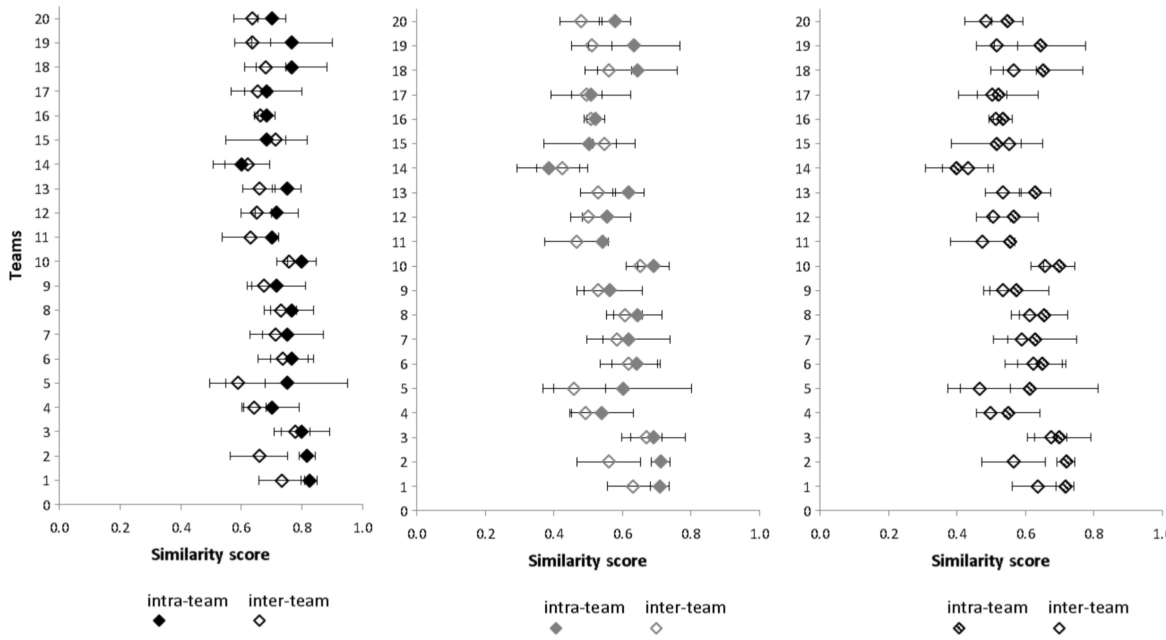
Figure 16. Individual subteam scores with mean 95% group confidence intervals for similarity of mental model of task sequence for the surgical subteam (S) and the inter-team scores for surgical and anaesthesia subteams (AS), and surgical and nursing subteams (NS), for scenario 1 (laparotomy for an abdominal stab wound), for the 20 teams. The further apart the markers are, the greater the intra- versus inter-team difference in similarity scores.

3.4.3.2. *Similarity scores for mental model of responsibility for task intra- versus inter-team*

Table 10 shows the statistics for the intra-versus inter-team comparisons for mental model of responsibility for task for both scenarios, when similarity was calculated using percentage agreement, Fleiss' kappa, and KALPHA for nominal data. The difference scores for the intra-team and inter-team similarity in Scenario 1 were normally distributed, as assessed by the Shapiro-Wilk test (p = 0.343 for percentage agreement; p=0.637 for Fleiss' kappa; p=0.504 for KALPHA for nominal data). In the case of scenario 1, the similarity scores for mental model of responsibility for task were significantly higher intra-team than inter-team when all three methods for calculating similarity were used. In the case of scenario 2, the difference scores for the intra-team and inter-team similarity were normally distributed when similarity was calculated using percentage agreement (p=0.167), Fleiss' kappa (p=303), and KALPHA for nominal data (p=0.203), as assessed by Shapiro-Wilk test. The paired samples t-test showed that the similarity scores for mental model of responsibility for task were significantly higher intra-team than inter-team, regardless of the method used to calculate similarity.

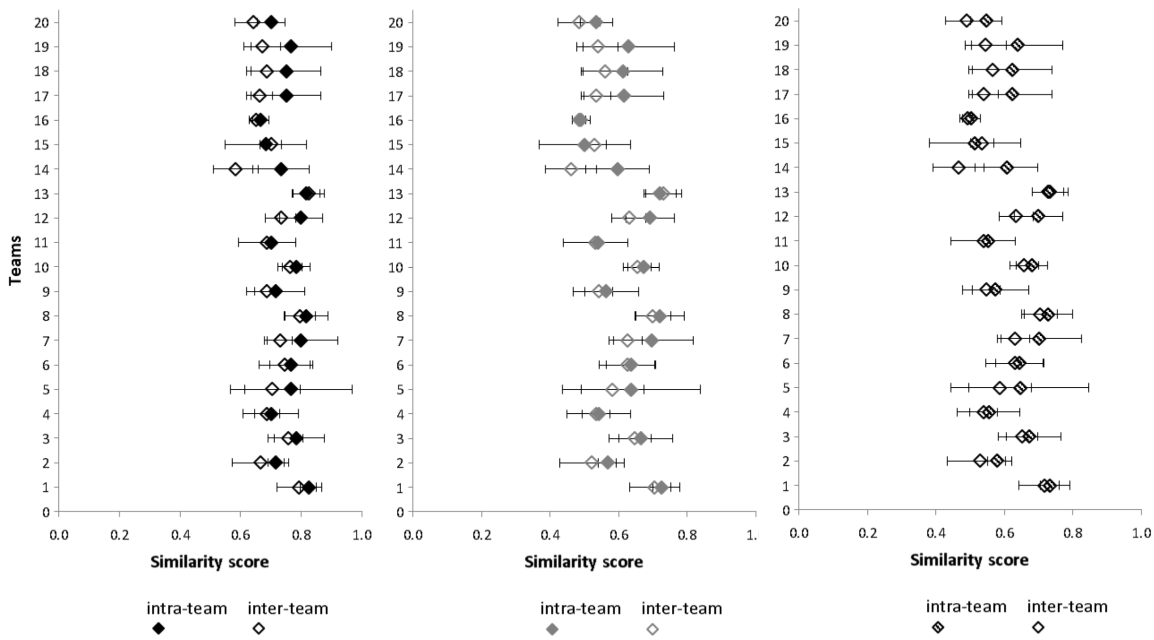
The forest plots in Figure 17 show the mean similarity scores for mental model of responsibility for task for intra-team and inter-team groupings for the two scenarios.

Scenario 1 (laparotomy for an abdominal stab wound)



a) Percentage agreement b) Fleiss' kappa c) KALPHA

Scenario 2 (laparotomy for a perforated viscus)



b) Percentage agreement b) Fleiss' kappa c) KALPHA

Figure 17. Mean overall intra-team versus inter-team scores with 95% confidence intervals for similarity of mental model of responsibility for task calculated using three methods for scenario 1, and scenario 2. The further apart the diamonds, the larger the mean difference in similarity scores intra-team versus inter-team when each method was used to calculate the degree of similarity.

3.4.3.2.1. *Subteam versus inter-team comparisons of scores for mental model of responsibility for task for individual subteams*

The similarity scores for the **anaesthesia subteam** (**A**), for mental model of responsibility for task were compared to the inter-team similarity scores for anaesthesia and the nursing subteams (**AN**) and the anaesthesia surgical subteams (**AS**) respectively. As shown in Table 10, for scenario 1, anaesthesia subteams had a significantly higher similarity score for mental model of responsibility for task for their subteam than inter-team with nursing subteams, and with surgical subteams. This was true for all three metrics used to calculate similarity. In scenario 2, the similarity scores for mental model of anaesthesia subteams were significantly higher than the inter-team scores of anaesthesia and surgical subteams, regardless of the methods used to calculate similarity. The difference between similarity scores for mental model of responsibility of task for anaesthesia subteam and the inter-team scores for anaesthesia and nursing subteams in scenario 2 was not significant for either metric used to calculate similarity.

The similarity scores for the **surgical subteam** (**S**) were compared to the similarity scores for the inter-team model of the surgical and anaesthesia subteams (**AS**) and the similarity scores for the inter-team model of surgical and nursing subteams (**NS**) respectively. As shown in Table 10, for scenario 1, on average, similarity scores of surgical subteams were significantly higher than the inter-team scores of surgical and anaesthesia subteams when percentage agreement was used to calculate similarity. However, no significant difference was found when Fleiss' kappa and KALPHA were used to calculate similarity. In the same scenario, the similarity scores for the surgical subteams were on average significantly higher than inter-team similarity scores of the surgical and nursing subteams. In scenario 2, no statistically significant difference was found between the similarity scores for surgical subteams and inter-team similarity scores of surgical and anaesthesia subteams, for either method used to calculate similarity. In the same scenario, the surgical subteam had a significantly higher similarity score for their subteam than inter-team with the nursing subteam when similarity of mental model was calculated using percentage agreement. However, the difference between the similarity scores of surgical subteams and inter-team scores of surgical and nursing subteams was not significant when similarity was calculated using Fleiss' kappa and KALPHA for nominal data.

The similarity scores for the **nursing subteam** (**N**) for responsibility for task were compared to the inter-team similarity scores of nursing and anaesthesia subteams combined (**AN**) and nursing and surgical subteams combined (**NS**) respectively. For both scenarios, the similarity scores for the nursing subteam were significantly higher within their subteam than inter-team scores with the surgical subteam, when similarity was calculated using all three methods (see Table 10).

The similarity scores for responsibility for task of the nursing subteam were not significantly different from the scores of the combined nursing and anaesthesia subteams in either scenario and when similarity was calculated using different methods.

The significant differences in individual subteam versus inter-team scores are graphically depicted in Figures 18-21.

Table 10. Statistics for the intra-versus inter-team comparisons for mental model of responsibility for task for scenario 1 and 2. All statistics are shown only where significant relationships were found.

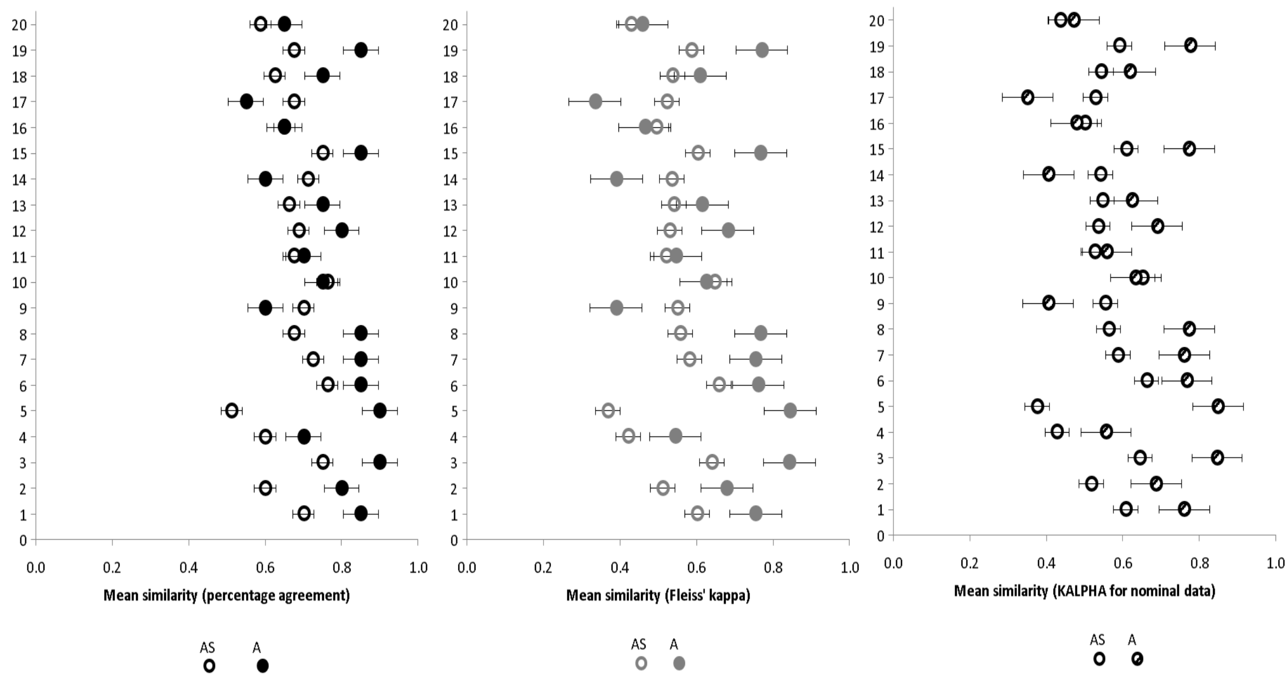
	Scenario 1 (laparotomy for an abdominal stab wound)							Scenario 2 (laparotomy for a perforated viscus)						
Similarity for task sequence	**Mean difference**	**Std error of the mean**	**t-test**	**p value**	**95% CI**		**Effect size (Cohen's d)**	**Mean difference**	**Std error of the mean**	**t-test**	**p value**	**95% CI**		**Effect size (Cohen's d)**
					Lower	**Upper**						**Lower**	**Upper**	
Overall intra- vs inter-team														
Percentage agreement	0.06	0.01	5.221	<0.0005*	0.04	0.08	1.17	0.04	0.01	4.932	<0.0005*	0.03	0.06	1.10
Fleiss' kappa	0.05	0.01	4.539	<0.0005*	0.03	0.08	1.01	0.04	0.01	4.277	<0.0005*	0.02	0.05	0.96
KALPHA	0.06	0.01	4.951	<0.0005*	0.03	0.08	1.11	0.04	0.01	4.786	<0.0005*	0.02	0.06	1.07
Individual intra- vs inter-team														
Anaesthesia subteam:														
A-AS														
Percentage agreement	0.09	0.03	3.201	0.005*	0.03	0.14	0.72	0.07	0.02	4.285	<0.0005*	0.04	0.10	0.96
Fleiss' kappa	0.09	0.03	2.588	0.018*	0.02	0.16	0.58	0.08	0.02	4.042	0.001*	0.04	0.12	0.90
KALPHA	0.09	0.03	2.758	0.013*	0.02	0.16	0.62	0.08	0.02	4.083	0.001*	0.04	0.12	0.91
A-AN														
Percentage agreement	0.04	0.02	2.652	0.016*	0.01	0.08	0.59	0.04	0.02		0.094	-0.01	0.09	
Fleiss' kappa	0.05	0.02	2.228	0.038*	0.003	0.09	0.50	0.05	0.03		0.135	-0.01	0.10	
KALPHA	0.05	0.02	2.476	0.023*	0.008	0.10	0.55	0.05	0.03		0.083	-0.005	0.11	
Surgical subteam:														
S-AS														
Percentage agreement	0.05	0.03	2.110	0.048*	0.0004	0.11	0.47	0.02	0.02		0.257	-0.02	0.06	
Fleiss' kappa	0.04	0.03		0.299	-0.03	0.11		-0.03	0.02		0.877	-0.05	0.04	
KALPHA	0.04	0.03		0.299	-0.03	0.11		0.002	0.02		0.928	-0.04	0.05	
S-NS														
Percentage agreement	0.09	0.03	3.344	0.003*	0.03	0.14	0.75	0.06	0.02	2.110	0.01*	0.01	0.10	0.47
Fleiss' kappa	0.08	0.03	2.258	0.036*	0.01	0.15	0.51	0.04	0.02		0.141	-0.01	0.09	
KALPHA	0.08	0.03	2.443	0.025*	0.01	0.15	0.55	0.04	0.02		0.092	-0.008	0.09	

Table 10 continued

Similarity for task sequence	Scenario 1 (laparotomy for an abdominal stab wound) Mean difference	Std error of the mean	t-test	p value	95% CI Lower	95% CI Upper	Effect size (Cohen's d)	Scenario 2 (laparotomy for a perforated viscus) Mean difference	Std error of the mean	t-test	p value	95% CI Lower	95% CI Upper	Effect size (Cohen's d)
Nursing subteam:														
N-AN														
Percentage agreement	0.01	0.02		0.750	-0.03	0.04		0.01	0.01		0.629	-0.02	0.04	
Fleiss' kappa	-0.005	0.03		0.853	-0.06	0.05		-0.007	0.02		0.691	-0.04	0.03	
KALPHA	-0.006	0.02		0.791	-0.10	0.04		-0.002	0.02		0.917	-0.04	0.03	
N-NS														
Percentage agreement	0.08	0.02	3.431	0.003*	0.03	0.12	0.79	0.07	0.02	4.117	0.001*	0.03	0.10	0.93
Fleiss' kappa	0.07	0.03	2.505	0.022*	0.01	0.13	0.57	0.07	0.02	3.302	0.004*	0.03	0.12	0.74
KALPHA	0.07	0.03	2.464	0.024*	0.01	0.13	0.57	0.08	0.02	3.698	0.002*	0.03	0.12	0.83

*p<0.05

Scenario 1 (laparotomy for an abdominal stab wound)



Scenario 2 (laparotomy for a perforated viscus)

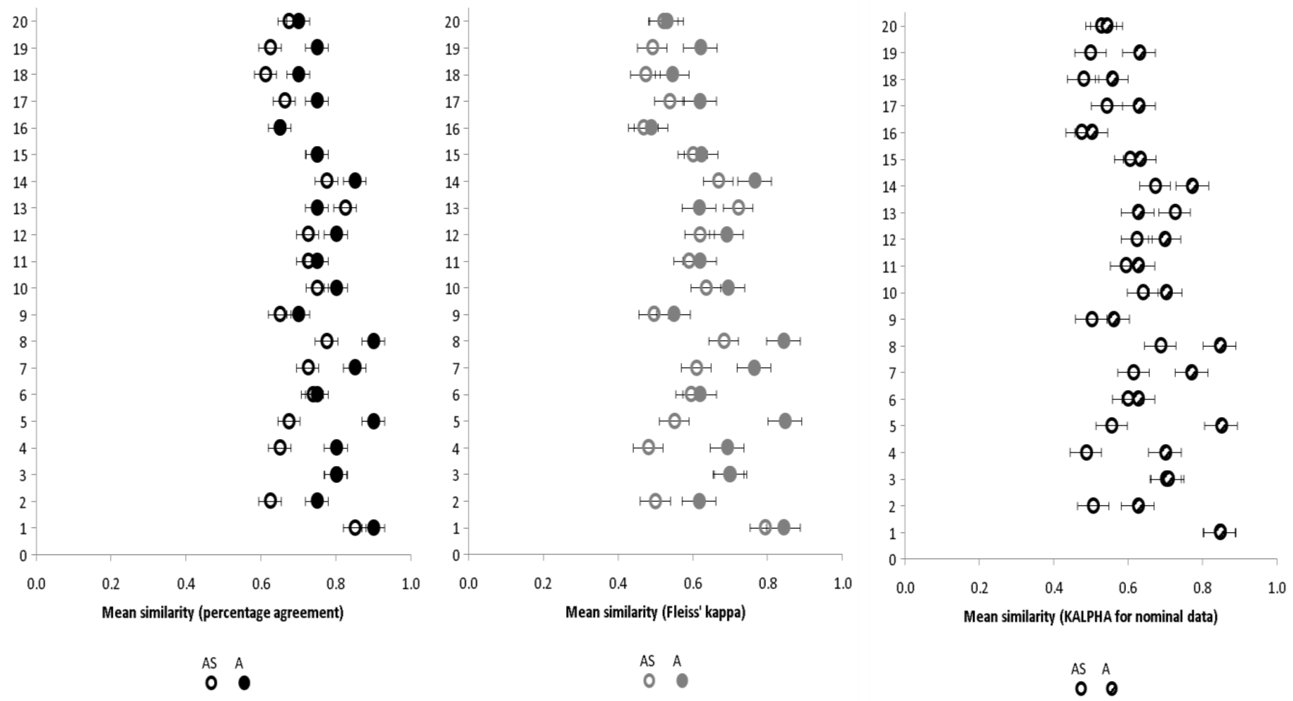


Figure 18. Mean similarity scores with mean 95% group confidence intervals for mental model of responsibility for task for anaesthesia subteams (A) and inter-team similarity score for anaesthesia and surgical subteams (AS), for scenario 1 and 2. The further apart the markers are, the greater the subteam versus inter-team difference in similarity scores.

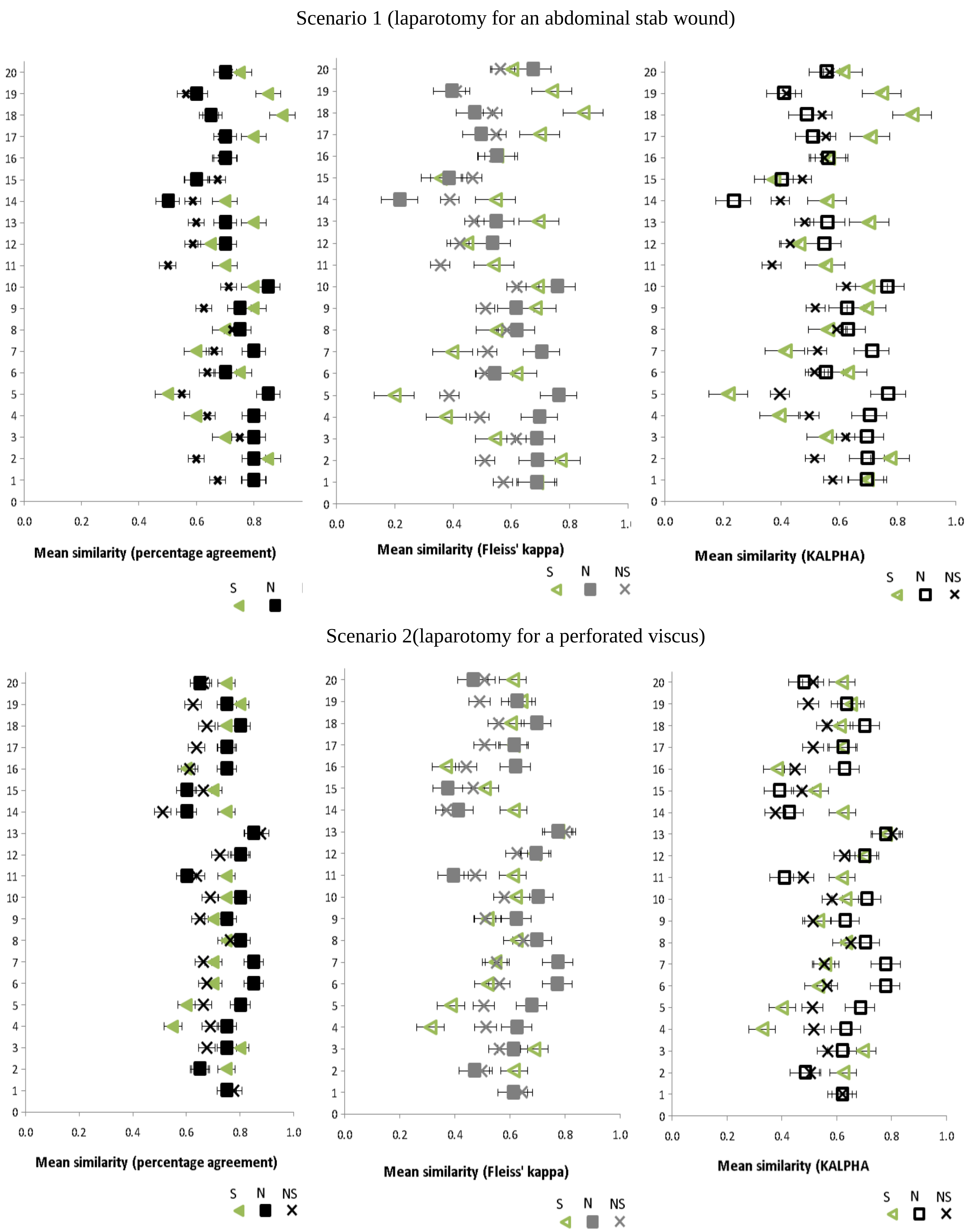


Figure 19. Mean individual subteam similarity scores with mean 95% group confidence intervals (using three methods to calculate similarity) for mental model of responsibility for task for the nursing (N) and surgical (S) subteams, and mean inter-team similarity scores for surgical and nursing subteams (NS) for scenario 1 and 2, respectively. The further apart the markers are, the greater the intra- versus inter-team difference in similarity scores.

Scenario 1 (abdominal stab wound)

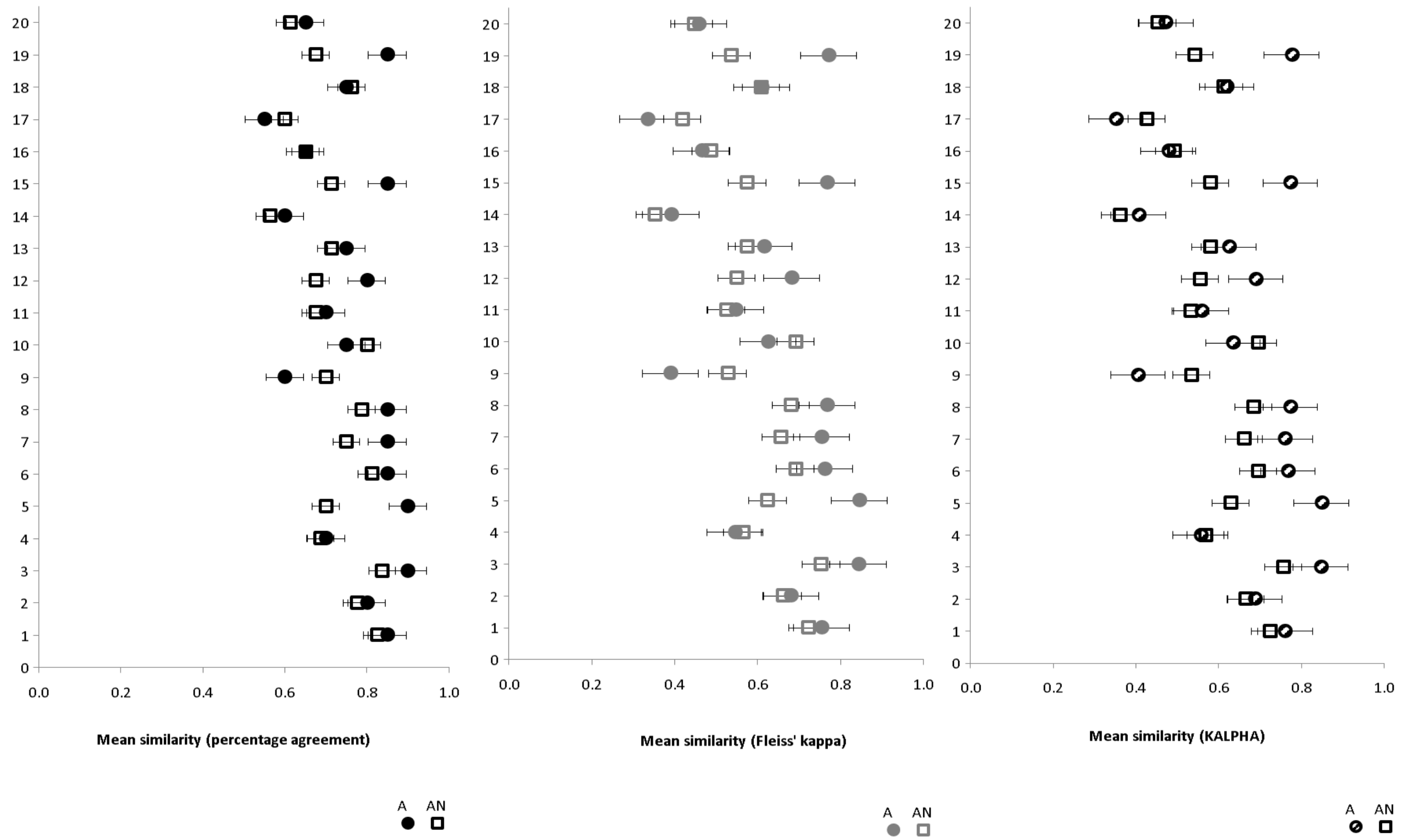


Figure 20. Mean individual subteam similarity scores with mean 95% group confidence intervals (using three methods of calculating similarity) for mental model of responsibility for task for the anaesthesia subteam (A) and mean inter-team similarity scores for anaesthesia and nursing subteams (AN) for scenario 1. The further apart the markers are, the greater the intra- versus inter-team difference in similarity scores.

Scenario 1 (abdominal stab wound)

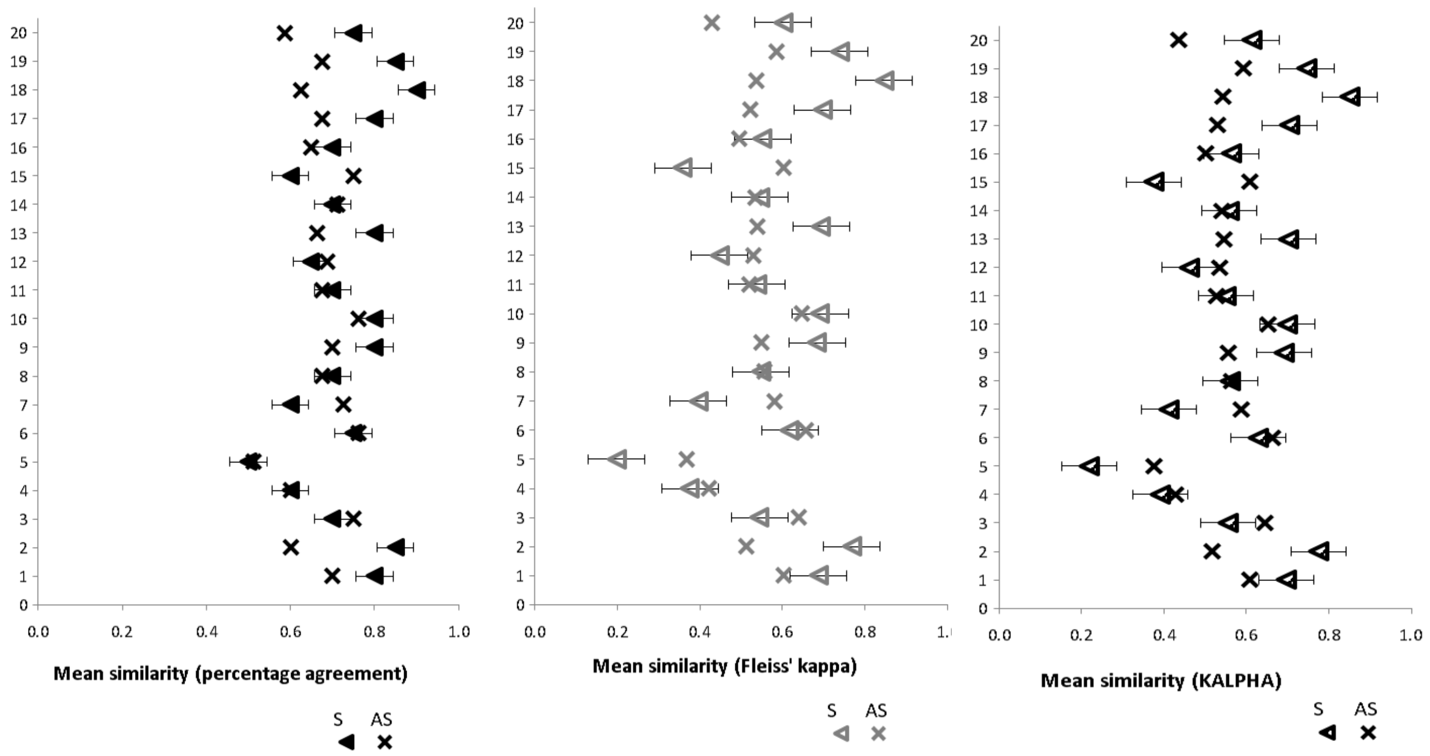


Figure 21. Mean individual subteam similarity scores with mean 95% group confidence intervals (using three methods to calculate similarity) for mental model of responsibility for task for the surgical subteam (S) and mean inter-team similarity scores for anaesthesia and surgical subteams (AS) for scenario 1. The further apart the markers are, the greater the intra- versus inter-team difference in similarity scores. The intra- versus inter-team difference was significant (p<0.05) when similarity was scored using percentage agreement, but not when Fleiss' kappa and KALPHA were used.

3.4.3.3. ***Using multiple methods to calculate the similarity scores for mental models***

The final assumption that would support validity of the Momento approach was that the identified relationships between the similarity scores for mental models and team familiarity scores, and intra-team versus inter-team differences in similarity scores for mental models should be demonstrated when similarity scores were calculated using different metrics. Table 11 shows bivariate correlations of the variables examined in the study. Spearman's rank-order correlation showed that different metrics used to calculate the similarity scores for the two types of mental models were highly correlated. Rescaled Euclidean distance and KALPHA for ordinal data, used to calculate the similarity scores for mental model of task sequence, were significantly positively correlated in the case of scenario 1 ($r_s(18) = 0.908$, $p < 0.0005$) and scenario 2 ($r_s(18) = 0.932$, $p < 0.0005$). Similarly, there was significant positive correlation between all the scores for mental model of responsibility for task calculated using the three methods in scenario 1 ($r_s(18) = 0.992$, $p < 0.0005$ between percentage agreement and both Fleiss' kappa and KALPHA for nominal data; and $r_s(18) = 1.00$, $p < 0.0005$ between Fleiss' kappa and KALPHA for nominal data), and scenario 2 ($r_s(18) = 0.996$, $p < 0.0005$ between percentage agreement and both Fleiss' kappa and KALPHA for nominal data; and $r_s(18) = 1.00$, $p < 0.0005$ between Fleiss' kappa and KALPHA for nominal data).

No statistically significant relationship was found between the similarity scores for mental model and team familiarity scores, for either type of mental model and regardless of the metric used to calculate the degree of similarity in both scenarios. Where the relationship between similarity scores for mental model of task sequence and subteams was investigated, similar patterns of results emerged when both Rescaled Euclidean distance and KALPHA for ordinal data were used to calculate similarity. Here, there was more agreement intra-team than inter-team that was due to surgical subteams having more similar scores for mental models within their subteam than inter-team with anaesthesia, and with nursing subteams, regardless of the method used to calculate the similarity scores (see Table 9). In the case of mental model of responsibility for task, similar patterns of results emerged regardless of which of the three metrics (i.e., percentage agreement, Fleiss' kappa, and KALPHA for nominal data) were used to calculate the similarity scores. The exception was the intra-team versus inter-team relationship between the similarity scores for mental model of the surgical subteam and inter-team similarity scores of surgical and anaesthesia subteams in scenario 1, and the inter-team model of surgical and nursing subteams in scenario 2, where a significant difference was found only when percentage agreement was used to calculate similarity scores. Therefore, the last assumption regarding validity was largely supported.

Table 11. Descriptive statistics and correlations at the dyad-Level for scenario 1 and 2.

	Mean	SD	1	2	3	4	5	6	7	8	9	10
1. Team familiarity	2.60	0.36										
SMM of task sequence												
Scenario 1 (STAB WOUND)												
2. Rescaled Euclidean distance	0.77	0.04	-0.32									
3. KALPHA	0.72	0.11	-0.23	0.91*								
Scenario 2 (PERFORATED VISCUS)												
4. Rescaled Euclidean distance	0.83	0.04	-0.07									
5. KALPHA	0.83	0.07	-0.07			0.93*						
SMM of responsibility for task												
Scenario 1 (STAB)												
6. Percentage agreement	0.69	0.05	0.17									
7. Fleiss' kappa	0.53	0.07	0.20					0.99*				
8.KALPHA	0.54	0.07	0.20					0.99*	1.00*			
Scenario 2 (PERFORATED VISCUS)												
9. Percentage agreement	0.72	0.06	-0.04									
10. Fleiss' kappa	0.57	0.08	-0.07								1.00*	
11.KALPHA	0.58	0.08	-0.07								1.00*	1.00*

* Spearman's rank-order correlation , $p<.0005$

3.5. Summary of results and discussion

The results of the validation study do not provide support for the assumption that being familiar with teammates is related to higher similarity scores for mental models of task sequence and responsibility for task within OR teams. The assumption that similarity scores for mental models within subteams would be significantly higher than inter-team similarity scores was largely confirmed by the findings for responsibility for task, and partially confirmed for task sequence. Further, results provide partial support for the differences between OR subteams in understanding of the key tasks and their sequence, and responsibility for those tasks. The new Momento approach to assessing mental models in OR teams distinguished between the surgical subteam as a group having greater similarity scores for mental models of task sequence within their subteam than inter-team with both anaesthesia and nursing subteams. It also distinguished between the surgical subteam as a professional grouping having higher similarity scores for responsibility for task within their subteam than inter-team with the nursing subteam in the case of the laparotomy for an abdominal stab injury. It also distinguished the anaesthesia, and the nursing subteams from the surgical subteam on the degree of similarity of mental model of responsibility for task in both scenarios. Finally, the findings of the validation study provide support for the assumption that different measures of similarity of mental models should demonstrate similar results if they are in fact measuring the same real phenomena. These findings are discussed in more detail in the section below.

3.5.1. Similarity scores for mental models and team familiarity scores

I found no significant relationship between team familiarity scores and similarity scores for either type of mental model investigated in this study. There are several possible explanations for this finding.

A likely reason for this finding is that team member agreement on the sequence of key tasks in two laparotomy procedures is not related to how often members of OR teams have worked with each other in the past. Likewise, for mental model of responsibility for those tasks. It may be that there are other variables, not captured in the present study, that are more likely to be related to similarity of these types of mental models than how familiarity teammates are with each other. For example, as proposed in the IPO model of teamwork depicted in Figure 2, personality traits, being trained at different times in different institutions or countries could have influenced mental models studied here more than how frequently team members worked with each other in the past.

Another possible reason for this finding is that teams in the current study were all established OR teams who have worked with each other before – a characteristic that was one of the requirements of the wider MORSim project. Therefore, there may not have been sufficient variability in the sample of participants on how often they had worked with each other in the past to distinguish between "non-familiar" and extremely "familiar" teams.

Another possible explanation is that simply having worked with team members in the past may not be enough to develop similar mental models. The present findings could alternatively be explained in the context of the nature of the task at hand. Familiarity with teammates, in terms of knowing how they function, what skills and abilities they possess, and what they are likely to do next, becomes more important the more dynamic the team's task is[30]. If a task is highly proceduralised and it does not matter which team member does what, knowing your teammates is not as important as in the case of a more complex and dynamic task which requires higher adaptability and greater anticipation of the needs and actions of teammates. This type of knowledge is thought to help individuals in a team form viable expectations for performance[30]. Thus, because emergency surgery is highly dynamic, it could be that knowledge of teammates' personal characteristics, rather than simply having worked with teammates in the past, may be more relevant for similar mental models to develop in a team. This broader definition of how familiar team member are with each other was not explored in this study. Related to this is the previous discussion on more versus less specialised teams. If we follow the Venn diagram analogy presented earlier to visually describe the concept of a shared mental model in a team, it is unlikely that there will ever be a complete overlap of mental models of individual members of OR teams as they consist of professionals with largely specialised roles. Highly dynamic and unpredictable situations such as those in emergency surgery require that the team relies to a high degree on specialised knowledge and expertise. Consequently, the overlap between models of individual team members related to tasks will be smaller in teams of specialists than in teams with less specialised knowledge. A distinction between teams based on the level of specialisation of team members has not previously been made in the context of mental models and their relationships with other variables. Therefore, it is possible that although the extent to which team members are familiar with each other may be related to mental models in teams that are less reliant on specialised knowledge, it may not make a difference in teams of specialists. In such teams, there is a limit to how similar mental models of individuals should be if the team is to also rely on specialised knowledge.

Similarly, it may be that other broader factors related to how familiar teammates are with each other, but not investigated in the current study, influenced the findings. For example, having prior positive experiences in a team, such as good coordination of activities with fellow team members, or prior success with the team achieving positive outcomes, is more likely to positively affect mental models[201]. Alternatively, if some team members have had negative experiences in terms of coordination with other team members, or poor individual or team performance in the past, they may not develop mental models that complement those of other team members[201]. Further, as the effect of how familiar teammates are with each other seems to be most beneficial during the team formation stage and decreases over time[138], it could be that, for the already established teams of participants in the current study, the association between familiarity with teammates and mental models had diminished.

In addition, it could be that the 4-point Likert-type scale used to capture the extent to which team members have previously worked with each other was not sensitive enough to distinguish the possible different levels of "team familiarity". Relying on subjective reports and recollections of participants introduces random error that is usually balanced out by obtaining multiple responses to the same questions[202]. However, there were examples in some of the teams in the current study where one team member reported having worked "a number of times before" with one of the other team members, while that team member reported to have "never before" worked with the same team member. Averaging out such contradictory responses could have potentially skewed the team's familiarity score and potentially influenced the findings on the relationship with similarity scores for mental models.

Finally, it could be that the extent to which team members are familiar with each other is not related to similarity of mental models in all situations. The level of uncertainty as to the potential surgical complications prior to the actual surgery in the scenarios the Momento task sort was based on in this research was low. It could be that how familiar teammates are with each other becomes relevant only in unpredictable emergent situations, when crisis occurs during surgery, rather than where complications are not anticipated.

3.5.2. **Similarity scores for mental models and subteams**

I found that, overall, members of the same OR subteam had more similar scores for mental model of task sequence within their own subteam (i.e. intra-team), compared to members of different subteams (i.e. inter-team) in scenario 1. The intra-team versus inter-team difference in similarity scores was not significant in scenario 2. However, this relationship was confirmed in both scenarios for mental model of responsibility for task. These findings thus largely support the validity of the Momento approach at the overall intra- versus inter-team level.

3.5.2.1. ***Similarity scores for task sequence and subteams***

Similarity scores for mental model of task sequence were significantly greater intra-team than inter-team in scenario 1, but not in scenario 2. One possible explanation for this difference is that participants were more likely to have encountered a laparotomy for a perforated viscus (scenario 2) than a stab wound requiring laparotomy. This in turn could have contributed to the greater overall agreement among OR teams as to the required steps and their order in the case of the more familiar procedure. The ability to discriminate between a more common and a less common surgical scenario provides further support for the validity of the Momento approach.

Individual subteam versus inter-team analysis showed that the overall intra- versus inter-team difference was due to the members of the surgical subteam having greater similarity scores on the sequence of tasks in scenario 1 within their own subteam than inter-team with anaesthesia, and nursing subteams. The finding of no differences in similarity scores for task sequence for the more familiar procedure (laparotomy for a perforated viscus) was confirmed at individual subteam versus inter-team level for all comparisons. The greater alignment of mental models within OR teams as to the order of the key steps is also implied by higher mean similarity scores in this scenario. Thus, the assumption that would support the validity of the Momento approach that there would be greater similarity of models on task sequence within individual subteams than inter-team was only confirmed for the surgical subteam and the less common scenario.

3.5.2.2. ***Similarity score for responsibility for task and subteams***

Similarity scores for mental model of responsibility for task were significantly higher for members belonging to the same subteam than between members of different subteams (inter-team). Thus, the second assumption that would support validity that mental models of subteams are more similar than inter-team mental models is confirmed for this type of model. Looking more closely at individual subteam versus inter-team comparisons, the postulated relationships were task-dependent. In scenario 1, the assumption was supported for:

- Anaesthesia subteam – where similarity score was higher within their subteam than between members of anaesthesia and surgical, and anaesthesia and nursing subteams;
- Surgical subteam – who had higher similarity score than with nursing, but not the anaesthesia subteam (except when similarity was calculated using percentage agreement); and
- Nursing subteam – who had a higher similarity score than inter-team, with members of the surgical, but not anaesthesia subteam.

In scenario 2, the second assumption was supported for:

- Anaesthesia subteam – whose similarity score was higher than the inter-team score with the surgical, but not the nursing subteam; and
- Nursing subteam – whose similarity score was higher than their inter-team one with the surgical, but not anaesthesia subteam.

The findings distinguish the anaesthetist and an anaesthetic technician as a subteam having a higher similarity score between each other than with surgeons or nurses on who should be primarily responsible for the key tasks in scenario 1, despite the fact that anaesthetists and surgeons are medical specialists and anaesthetic technicians are not. Similarly, both the subteam of surgeons and the subteam of nurses had a higher similarity score on who should be responsible for tasks in scenario 1 in their subteams than inter-team with each other. This may point to individual subteams relying more on their background-specific knowledge and experience when faced with a more uncommon task (scenario 1). In scenario 2, both anaesthesia and nursing subteams had higher similarity scores within their respective subteams than inter-team with the surgical subteam. In both scenarios, the nursing subteam had higher similarity scores within their subteam than inter-team with the surgical subteam. This finding has practical implications for the smooth collaboration between the scrubbed members of the OR team – the surgical subteam and the scrub nurse. The high level of coordination required of these professionals could be compromised if the scrub nurse has different expectations from the surgeons of who should be responsible for what in both common and uncommon laparotomy.

Regardless of the type of mental model, professional differences in the degree of agreement were less pronounced in the case of laparotomy for a perforated viscus than the less common laparotomy for an abdominal stab wound scenario. This can have practical implications. Having a similar mental model becomes more important during dynamic tasks requiring more flexibility and team member adaptability, where team members' knowledge of each other's responsibilities and needs may enhance their ability to develop viable expectations for performance[30].

3.5.3. Using multiple methods to calculate the similarity scores for mental models

The third assumption that would support validity stated that different measures of similarity of mental models should demonstrate similar results if they are in fact measuring the same real phenomena. The findings largely support this assumption.

All methods used to calculate similarity produced scores for each type of model that were highly, significantly, and positively correlated. Similar patterns of results emerged when different methods were used to represent similarity for both types of mental models. No statistically significant relationship was found in this study between team familiarity scores and similarity scores for mental models, although various metrics were used to calculate similarity. Similar patterns were also found when comparing similarity scores of intra- versus inter-team groupings when similarity was calculated using different metrics. The only exceptions were a significant difference between similarity scores of the surgical subteam and their inter-team scores with the anaesthesia subteam in scenario 1, and with the nursing subteam in scenario 2, when percentage agreement was used to calculate the scores for responsibility for task. As discussed in section 2.7.2 of this thesis, percentage agreement is the most intuitive, but lax measure of agreement, as it does not take into account chance agreement. This issue is reflected in the findings of the present study, where mean agreement for all professional groupings was consistently higher when similarity scores for mental model of responsibility for task were calculated using percentage agreement, compared to the other two metrics. The above divergence of findings when this metric was used to calculate similarity scores could be explained by not accounting for agreement due to chance in the calculation of similarity using percentage agreement. Despite this, I argue that percentage agreement can still be useful, informative, and possibly the most intuitive way to assess agreement on who should be responsible for what in practical situations. When time is limited, which is often the case in emergency surgery, data on individual mental models of who should be responsible for what in an upcoming procedure could be relatively effortlessly aggregated using percentage agreement. This information can then be used to initiate discussion in the OR team in order to align individual team members' understandings of responsibilities prior to the procedure. This is discussed in more detail in the next chapter of this thesis.

In summary, the data from the validation study largely support the validity of the Momento approach. The first assumption on the relationship between similarity scores for mental models and team familiarity scores was not supported by the data. However, the second assumption that the Momento approach would distinguish between similarity scores within subteams from the similarity scores for inter-team mental models was largely supported in the case of the mental model of responsibility for task in both laparotomy scenarios. This assumption was also partially supported for the mental model of task sequence in less common scenario 1. The data supported the third assumption that multiple ways to calculate similarity scores would yield similar results. The Momento approach was also able to distinguish between a more common and a less common surgical scenario. This finding provides further support for the validity of the Momento approach.

3.5.4. **Clinical significance of the findings**

What are the implications of these findings for practice? Despite the relatively high p values, seemingly large effect sizes and relatively narrow 95% CIs for the intra- versus inter-team comparisons, it is difficult to know whether these statistically significant differences may be of clinical relevance. Looking at figures 15 and 17, depicting overall intra- versus inter-team comparisons at team level, variation within each group of scores is relatively small, and except for only one or two of the 20 intra- versus inter-team pairs, the intra-team scores are higher than the inter-team ones. However, the difference between the overall intra- and inter-team scores is relatively small. In the case of mental model of task sequence in scenario 1, the effect sizes of the overall and surgical subteam intra- versus inter-team difference in similarity were moderate as per established criteria[203], with relatively wide 95% confidence intervals for the mean differences. This indicates that intra-team versus inter-team differences in the degree of similarity of mental models found in the present study are likely to be too small to be of practical relevance.

Further, the absence of similar studies, especially in the healthcare context, makes it difficult to compare the effect sizes found in this study to those reported in prior studies of a similar nature. Comparisons could potentially be made to radiology-specific studies of the rates of agreement on interpretations of imaging studies and disease detection by radiologists. These studies have found a 2 to 6 percent clinically significant differences in the rates of agreement in radiology practice[204-206], with higher disagreement rates (4.2%) for those more serious cases where the discrepancy could potentially adversely affect outcome[205]. In a study comparing emergency physicians and radiologists in the interpretation of head CTs, the level of agreement between the two types of specialist was 83.8%[207]. The authors regarded the level of disagreement of 16.2% found in this study as significant and warranting intervention, such as continuous medical education workshops, to improve the level of agreement between specialists. In other studies, discrepancy between emergency physicians and radiologists in their readings of patients' X-rays has been reported to be between 0.95% and 16.8%[208, 209].

Related to the problem of data interpretation in the clinical context is the statistical issue of whether or not to correct for multiple comparisons, as when examining intra- versus inter-team differences at several different levels of analysis.

When we test many hypotheses, each test has a specified probability of Type I error, which involves the incorrect rejection of the null hypothesis or the finding of a "false positive" relationship between variables. The probability of committing Type I errors can increase sharply the more tests we conduct. This can have serious implications if the results of multiple tests are to be interpreted as a whole and if important decisions are to be based on these results[210, 211].

There are contrasting views on how to deal with the issue of multiple comparisons. These depend on how much importance is assigned to the problem of controlling for Type I error while maintaining the likelihood of detecting an effect if it exists[211, 212]. No single solution will be acceptable for all situations[211]. For example, some of the common adjustment methods, such as Bonferroni adjustment, have been criticised for being too conservative and of limited practical value[213-215]. Authors generally agree that rigorously controlling for an error rate in the case of multiple comparisons is a must in confirmatory studies, where the goal is to provide a definitive proof of a predefined key hypothesis that is to guide final decision making[210, 216], such as trials of new drugs. However, when the research is pursuing novel or exploratory hypotheses, multiple test adjustments are not strictly required[210, 216, 217]. Some authors go even further to argue that any form of correction for multiple comparisons is a "penalty for peeking" (p.29)[218], where labelling some observed relationships as chance findings prevents a more intensive scrutiny and future research that may provide us with a better understanding of the potentially interrelated phenomena[213, 219].

A moderate view proposes that, regardless of whether corrections for multiple comparisons are performed or not, in exploratory studies multiple significance tests should only be used with caution and for descriptive purposes, and not for decision-making[216, 217]. The findings of such studies should be tested further in confirmatory studies[216], where the cut off for significance level should be higher[214].

This thesis represents novel and exploratory research, and is a first step in validation of a new tool for assessing phenomena not previously quantified in the multidisciplinary OR setting. Therefore, I argue that to formally correct for multiple comparisons in the beginning stages of this research would hinder the potential to explore what may be real, clinically relevant differences between groups of OR team members and how they perceive each other's tasks and responsibilities, in subsequent studies. In view of the findings and recommendations from literature, and given the high risk nature of the healthcare environment where potential for error is greater and consequences may be more damaging than in most industries, a small significant difference in similarity of mental models of responsibility for task found in the present study warrants further inspection.

In the chapter that follows, I use the data gathered in the validation study to demonstrate how the Momento approach could be extended to inform clinical practice. Here, the degree of similarity of mental models is explored at the level of individual tasks in the surgical procedure for the entire OR team and for individual subteams.

4. TOOL APPLICATION

This chapter addresses the third objective of my thesis, which was to demonstrate how the Momento approach to assessment of mental models in OR teams could be extended for use in clinical practice. Ultimately, the goal of this analysis was to provide a quick and easy way to establish the degree of agreement within an OR team on the sequence in which tasks should be performed in the upcoming procedure and which subteam should be primarily responsible for each task. In one application, this information could be obtained before starting an operation and then shared among the members of the team so that they could discuss the potential discrepancies and subsequently align their mental models before the upcoming surgery.

The specific aim here was to show how the Momento approach could be used to answer the following questions:

1) To what degree do OR team members agree prior to the procedure on when each key task in the procedure should be done. This information would be indicative of their shared mental model of task sequence; and

2) To what degree do OR team members agree on who is primarily responsible for each key task, prior to the procedure. This information would be indicative of their shared mental model of responsibility for task.

4.1. Analysis of the data at the level of the task

The data gathered during the validation study described in chapter 3 were analysed at the level of individual tasks in each of the two study scenarios.

As in the validation study, for each participant, the data from the Momento task sort were presented as a set of numerical ranks, where each of the 20 tasks was represented by a number denoting its rank in the sequence, and a subteam category assigned to each task representing a subteam chosen as primarily responsible for that task. All possible pairwise comparisons of individual ranks and categories within each OR team were conducted to calculate the degree of similarity of the mental models. To help understand the potential applicability of the Momento approach in the clinical context, the similarity scores are here referred to as "agreement scores".

4.1.1. **Calculating agreement scores on task sequence within a subteam**

I calculated the agreement scores for the sequence of each task between pairs of subteam members as the absolute difference between the ranks assigned to the task divided by the maximum possible score for any given task (19 is the maximum possible agreement score for any given task, for two individuals, given a list of twenty tasks; a maximum score at this stage of the calculation implies the least possible agreement). The result was then subtracted from 1, in order to reverse the scale so that ascending scores represented ascending agreement. Thus, perfect agreement on when a particular task should be undertaken in the sequence would produce a score of 1-(0/19) = 1, and the worst possible agreement would produce a score of 1-(19/19) = 0. These scores were then expressed as percentages.

4.1.2. **Calculating agreement scores on responsibility for task within a subteam**

I calculated agreement scores for responsibility for each task between pairs of subteam members as "1" if the two participants agreed and "0" if they disagreed and then calculated the mean agreement score for each subteam, expressed as a percentage.

4.1.3. **Calculating agreement scores within a multidisciplinary team**

For both task sequence and responsibility for task, the agreement score for the multidisciplinary OR team for each task was calculated as the mean of all possible pairwise combinations within the OR team. For each task and type of mental model in each scenario, I then calculated the mean similarity score for twenty participating OR teams.

4.2. Results

Tables 12 and 13 show the mean agreement scores for task sequence and responsibility for task, respectively.

For the whole team, the overall mean agreement score for both scenarios[1] was 87% (range 57-97%, median 86%) for task sequence and the mean agreement score on task sequence exceeded 80% for all but two tasks in one of the scenarios (Table 12). The overall mean agreement score for both scenarios was 70% (range 38-100%, median 70.5%) for task responsibility, and more than half of the items (26 out of 40) scored less than 80%.

As seen in Tables 12 and 13, for both scenarios, “making a surgical incision” was the task with the highest mean agreement score both for task sequence and for responsibility (for which the score was 100%). As shown in Table 12, the lowest mean agreement score for sequence was given for when to inform the intensive care unit during a laparotomy for an abdominal stab wound (scenario 1) (57%), followed by when to confirm estimated blood loss in both scenarios (72% and 80%). As shown in Table 13, all participants indicated that the anaesthesia subteam is primarily responsible for performing a rapid sequence induction in scenario 1, and administering anaesthesia induction drugs in scenario 2.The lowest score for responsibility was given for checking for optimal patient positioning (38% and 39%, respectively) closely followed by estimating blood loss (39% in both scenarios), as shown in Table 13.

Mean agreement scores on task sequence were largely consistent across the three subteams (see Table 12). Within some subteams, mean agreement scores on responsibility for

[1] The mean agreement for the two scenarios was the mean of the OR team scores for the 20 tasks (as shown in the last row of Tables 12 and 13) for scenario 1 and scenario 2

some of the tasks in the procedure were lower than in other subteams (see Table 13). For example, mean agreement scores were lower among surgical subteams than the anaesthesia and nursing subteams for who should be primarily responsible for ensuring patient warming devices are in place. Similarly, the mean agreement score was higher in anaesthesia subteams than the surgical subteams for who they thought should be primarily responsible for ensuring appropriate antibiotic prophylaxis.

Table 12. Mean sequence rank assigned to each task and mean agreement scores on task sequence (expressed as percentages) for OR team and subteams (anaesthesia, surgical, nursing) for both scenarios.

Agreement on task sequence	Scenario 1: Laparotomy for an abdominal stab wound					Scenario 2: Laparotomy for a perforated viscus				
	Mean task rank	OR Team	Anaesthesia subteam	Surgical subteam	Nursing subteam	Mean task rank	OR Team	Anaesthesia subteam	Surgical subteam	Nursing subteam
1. Check blood availability	2.3	90	88	92	88	2.0	92	93	92	90
2. Check for optimal patient positioning on table	5.6	83	80	89	88	4.9	84	82	85	85
3. Initiate sign in	2.2	90	86	93	93	1.9	91	91	95	91
4. Administer anaesthesia induction drugs	5.4	85	82	85	91	5.1	88	86	88	90
5. Perform a rapid sequence induction	6.0	86	85	88	91	5.5	86	87	86	89
6. Ensure appropriate antibiotic prophylaxis	7.7	86	86	88	85	6.7	84	81	84	86
7. Insert urinary catheter	7.2	86	84	87	85	6.8	85	88	89	82
8. Initiate time out	9.3	85	85	83	89	8.9	86	89	78	88
9. Make surgical incision	12.0	94	93	95	96	11.3	97	96	98	97
10. Ensure patient warming devices in place	6.1	84	83	82	87	5.6	86	86	88	88
11. Ensure TED stockings and calf compressors on	6.3	83	81	88	87	5.4	86	84	89	87
12. Monitor ongoing blood loss	12.3	85	84	83	89	12.5	88	91	84	86
13. Organise bed space in PACU	15.5	80	85	86	79	15.2	81	85	86	76
14. Close incision	15.8	92	92	94	88	15.7	92	93	93	87
15. Check drains are turned on	17.0	89	89	89	92	17.0	90	91	89	92
16. Confirm estimated blood loss	13.1	72	69	74	75	14.5	80	72	82	84
17. Initiate sign out	18.0	87	84	91	89	18.3	91	93	95	88
18. Provide handover on intraoperative events to PACU staff	19.3	92	98	90	86	19.5	95	98	89	98
Inform intensive care unit (scenario1)	8.6	57	66	74	64	n/a	n/a	n/a	n/a	n/a
Remove knife from abdomen (scenario 1)	13.3	92	88	93	92	n/a	n/a	n/a	n/a	n/a
Locate site of perforation (scenario 2)	n/a	n/a	n/a	n/a	n/a	12.3	95	93	97	96
Fashion stoma (scenario 2)	n/a	n/a	n/a	n/a	n/a	14.6	91	93	93	89
Mean agreement		**85**	**84**	**87**	**87**		**88**	**89**	**88**	**88**

Table 13. Mean agreement scores for OR team and subteams (anaesthesia, surgical, and nursing) on responsibility for each task for both scenarios.

Agreement on responsibility for task	**Scenario 1: Laparotomy for an abdominal stab wound**				**Scenario 2: Laparotomy for a perforated viscus**			
	OR Team	Anaesthesia subteam	Surgical subteam	Nursing subteam	OR Team	Anaesthesia subteam	Surgical subteam	Nursing subteam
1. Check blood availability	51	55	50	74	55	65	32	65
2. Check for optimal patient positioning on table	39	50	85	37	38	45	79	55
3. Initiate sign in	78	80	85	63	76	80	84	65
4. Administer anaesthesia induction drugs	96	100	95	95	100	100	100	100
5. Perform a rapid sequence induction	100	100	100	100	98	100	95	100
6. Ensure appropriate antibiotic prophylaxis	61	90	55	63	64	90	37	70
7. Insert urinary catheter	55	75	70	68	53	60	84	75
8. Initiate time out	73	70	75	84	73	75	74	75
9. Make surgical incision	100	100	100	100	100	100	100	100
10. Ensure patient warming devices in place	54	80	45	74	55	80	37	80
11. Ensure TED stockings and calf compressors on	83	85	75	84	87	90	68	100
12. Monitor ongoing blood loss	61	90	55	53	58	85	37	45
13. Organise bed space in PACU	69	80	65	74	72	85	79	65
14. Close incision	96	95	100	95	98	100	100	95
15. Check drains are turned on	46	35	55	68	48	45	63	45
16. Confirm estimated blood loss	39	40	50	37	39	55	37	35
17. Initiate sign out	81	85	85	84	77	65	89	75
18. Provide handover on intraoperative events to PACU staff	44	40	45	37	49	45	58	55
Inform intensive care unit (scenario 1)	50	70	75	58	n/a	n/a	n/a	n/a
Remove knife from abdomen (scenario 1)	98	100	90	100	n/a	n/a	n/a	n/a
Locate site of perforation (scenario 2)	n/a	n/a	n/a	n/a	98	95	100	100
Fashion stoma (scenario 2)	n/a	n/a	n/a	n/a	97	100	100	90
Mean agreement	**69**	**76**	**73**	**72**	**72**	**78**	**73**	**75**

Frequency scatterplots in Figure 22(a) and 22(b) show the spread of ranks in the task sequence assigned by individual team members (20 within each of the six-member categories) to individual tasks on which there was lowest (57%) and highest average agreement (97%) within OR teams. Figure 22(c), provided here for illustrative purposes, shows a plot showing ranks assigned to a task with a mid-range agreement (80%). The plots demonstrate that the lower the OR team mean agreement score for the sequencing of a task in the procedure, the greater the spread of ranks assigned by team members to that task, with the lowest mean score (in the case of informing the intensive care unit in scenario 1) generating the greatest spread of ranks. By contrast, the highest OR team mean score (in the case of making a surgical incision in scenario 2) had the most unified rankings, regardless of team member. In the case of informing the intensive care unit (Figure 22(a)), the majority of surgeons thought this task should be performed at the start of the procedure, while the majority of nurses and anaesthesia subteams believed it should be done sometime in the second half of the procedure.

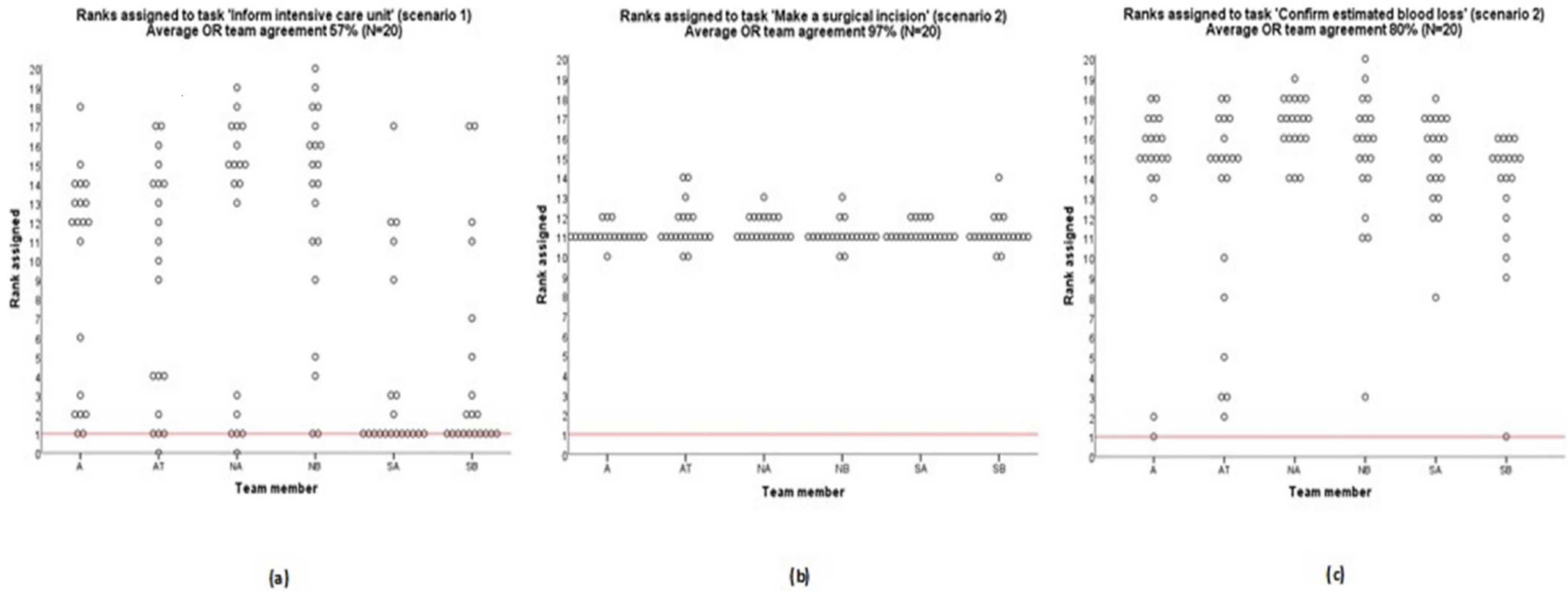


Figure 22. Frequency scatterplots of ranks assigned by team members to the tasks with the lowest (a), highest (b), and mid-range (c) OR team agreement on position in the sequence of tasks in the procedure. The horizontal line denotes the rank of 1 (i.e., first in the sequence); a rank of 0 means that a participant omitted the task from the task sort as "not required" in the given scenario.

Table 14 shows the mean percentage of times each subteam believed which of the three possible OR subteams was primarily responsible for a task. On average, the anaesthesia subteam believed they were responsible for 10 out of 20 tasks in scenario 1, and 9 out of 20 tasks in scenario 2. The surgical subteam believed their subteam was responsible for 9 out of 20 tasks in both scenarios, while the nursing subteam believed they were responsible for half the tasks in each scenario. Subteams produced variable responses for the tasks with lower mean agreement score on responsibilities. For example, all three subteams chose their own subteam as being primarily responsible for checking for optimal patient positioning on the OR table. While the majority of anaesthesia and surgical subteam members agreed that providing handover on intraoperative events to PACU (post-anaesthesia care unit) staff was primarily the responsibility of the anaesthesia subteam, most nurses thought they should be primarily responsible for this task. For confirming estimated blood loss, one of the tasks with lowest agreement scores for both task sequence and responsibility for task, there was also a split within the subteams, with the nurses splitting primary responsibility between all three subteams.

Table 14. Mean agreement scores within individual subteams on which one of the subteams (A=Anaesthesia subteam; N=Nursing subteam; or S = Surgical subteam) they thought was primarily responsible for each task*. Tasks for which there were lower mean agreement scores are shown in bold.

	Who is primarily responsible?**								
	Rated by A subteam			Rated by S subteam			Rated by N subteam		
Task	**A**	**N**	**S**	**A**	**N**	**S**	**A**	**N**	**S**
Check blood availability	83%	19%	1%	56%	21%	23%	63%	37%	0%
Check for optimal patient positioning on table	53%	11%	36%	3%	8%	90%	3%	58%	39%
Initiate sign in	10%	85%	3%	6%	92%	1%	15%	82%	3%
Administer anaesthesia induction drugs	100%	0%	0%	99%	0%	0%	99%	1%	0%
Perform a rapid sequence induction	100%	0%	0%	99%	0%	0%	100%	0%	0%
Ensure appropriate antibiotic prophylaxis	95%	0%	5%	52%	1%	47%	84%	6%	10%
Insert urinary catheter	0%	44%	56%	0%	14%	86%	0%	73%	25%
Initiate time out	1%	53%	46%	0%	43%	57%	1%	61%	38%
Make surgical incision	0%	0%	100%	0%	0%	100%	0%	0%	100%
Ensure patient warming devices in place	90%	10%	0%	48%	48%	4%	44%	56%	0%
Ensure TED stockings and calf compressors on	1%	94%	4%	1%	87%	11%	0%	96%	3%
Monitor ongoing blood loss	94%	4%	3%	63%	6%	30%	71%	14%	15%
Organise bed space in PACU	45%	54%	1%	25%	73%	1%	37%	61%	1%
Close incision	0%	1%	99%	0%	0%	100%	0%	0%	97%
Check drains are turned on	0%	65%	34%	0%	20%	77%	0%	73%	27%
Confirm estimated blood loss	63%	8%	30%	39%	16%	44%	48%	29%	23%
Initiate sign out	1%	83%	14%	0%	92%	6%	0%	89%	10%
Provide handover on intraoperative events to PACU staff	71%	29%	0%	72%	18%	10%	33%	66%	0%
Inform intensive care unit (scenario1)	73%	5%	20%	40%	0%	60%	77%	13%	8%
Remove knife from abdomen(scenario1)	0%	0%	100%	0%	3%	98%	0%	0%	100%
Locate site of perforation (scenario 2)	0%	3%	98%	0%	0%	100%	0%	0%	100%
Fashion stoma (scenario 2)	0%	0%	100%	0%	0%	100%	3%	0%	95%
	40%	**26%**	**34%**	**27%**	**25%**	**48%**	**31%**	**37%**	**32%**

*For example, on average, 83% of the anaesthesia subteam thought their own subteam was primarily responsible for checking blood availability in the two scenarios, while 19% of the same subteam thought the nurses were primarily responsible; similarly each subteam believed checking for optimal patient positioning on the table was primarily the responsibility of their own subteam.

**For those tasks that some participants believed not to be required in the procedure, the total agreement score for the three OR subteams as rated by a subteam may be less than 100%.

4.3. Discussion

In this study, I used the Momento approach to assess the level of agreement in an OR team on the sequence of individual tasks and which subteam should be primarily responsible for each task. Using the data on the 20 teams from the MORSim study, I found that there was poor agreement between OR team members on responsibility for task for half the tasks in each procedure. This has potentially important and concerning implications for safe and efficient team work. OR team members had largely similar understandings of when tasks should be done in an upcoming procedure for all but two tasks, which was more reassuring. Members of the three OR subteams believed their own subteam was primarily responsible for around half the tasks in each procedure.

It has been suggested that agreement over who is responsible for what may be more important for team performance than agreement over the sequence in which tasks should be done[47]. Redundancy in perceived responsibility for a task may be seen as making that task less likely to be forgotten, but may also result in the (possibly unjustified) assumption that it can be left to others. Furthermore, it would seem to be less efficient to have more than one person taking responsibility for a task, especially in a crisis where time and resources are precious.

This study served as a proof of concept demonstration, showing that it is possible to capture the information on certain types of mental models in multidisciplinary OR teams. Moreover, the differences between members of OR teams found in this study – using a relatively lax measure of agreement in the case of agreement on responsibility for task that would have somewhat inflated the level of agreement due to not accounting for chance agreement - indicate that this line of work is worth further expanding on. By completing the Momento task sort, the OR team has access to practical information than can be used to identify the potential differences between team members in their understanding of the key tasks and related responsibilities in an upcoming surgical procedure. Such data analysis can be made into a fully automated process with preset formulas in Microsoft Excel for calculating team agreement at different levels that would make the data on team agreement readily available. Also, the task sort could be easily modified for use in different clinical scenarios by using different lists of tasks generated to be similarly representative of the procedure.

Making team members aware of the extent of the discrepancies in individual mental models prior to embarking on a case gives them an opportunity to regroup and address the gaps in shared understanding, to make sure all team members are "on the same page" as to who should be responsible for which crucial tasks and when. In theory, providing time for team members to agree, perhaps through a briefing session or in relation to the WHO Surgical Safety Checklist ahead of a procedure, on the order of crucial tasks and on who should be responsible for each task, could help clarify potential ambiguities and better align mental models[220, 221]. The level of disagreement seen in this study for some tasks reinforces an increasing body of evidence supporting a pre-procedure briefing to align understandings and circumvent intra- and postoperative complications and reduce wasted time[222-224]. However, due to the logistics and constraints in the clinical setting, I do not envisage that the Momento tool will be used in day-to-day clinical practice. Instead, its strength lies in education and research. This is discussed in more detail in the next chapter.

The final chapter of this thesis is a discussion of all the phases of this project, including its contributions and limitations, and suggestions for future research.

5. DISCUSSION

5.1. Main findings

The findings of the tool validation study of this thesis largely support the concurrent validity of the Momento approach for measuring mental models of task sequence and responsibility for task. The first assumption was that OR teams who have worked more frequently with each other in the past would also have more similar mental models. This was not supported by the findings, as discussed in section 3.5 of chapter 3. The second assumption that mental models would be more similar within professional subteams than with colleagues from other subteams was partially supported. The Momento approach distinguished between subteam and inter-team groupings in both scenarios on responsibility for task, and in scenario 1 for task sequence. On closer inspection, the latter difference was mainly due to members of the surgical subteam agreeing more among themselves on task order for the abdominal stab wound scenario than with members of the nursing or the anaesthesia subteam. In the case of responsibility for task, the Momento approach distinguished between the degree of similarity of the anaesthesia subteam as being significantly higher than that of their inter-team model with the surgical subteam in both scenarios, and with the nursing subteam in scenario 1. It also distinguished between the nursing subteam as having a more similar mental model of responsibility for task within their subteam than inter-team with the surgical subteam in both scenarios. The above relationships were found when different metrics were used to calculate the degree of similarity. This suggests that they are in fact measuring the same phenomenon, confirming the third validity assumption of this thesis. The surgical subteam had more similar mental models of responsibility for task within their own subteam than inter-team with the nursing subteam in scenario 1 when all three methods were used to calculate similarity scores for responsibility for task. These findings also provide support for the third validity assumption that different approaches used to calculate similarity of mental models should demonstrate similar results if they are in fact capturing the same real phenomena. However, the agreement on responsibility for task in the surgical subteam was on average significantly greater than inter-team agreement with the anaesthesia subteam in scenario 1, and with nursing subteam in scenario 2, only when percentage agreement was used to measure agreement. This finding provides support for the need to use multiple scoring methods in the same study[32] when examining the relationships between similarity of mental models and other variables. The finding that the Momento approach was able to distinguish between a more common and a less common surgical scenario provides further support for its validity.

In the tool application study, I demonstrated that the new Momento approach could be used to capture certain types of mental models in multidisciplinary OR teams. The findings of this study showed some differences in mental models of members of OR teams and subteams regarding some of the key tasks and responsibilities in laparotomy. Differences were particularly pronounced in the case of who should be primarily responsible for certain crucial tasks in laparotomy. This was the case regardless of the scenario the task sort was related to in this study. The level of disagreement on some tasks found in this study warrants further investigation and the new Momento approach could next be used to elicit mental models in a wider range of clinical scenarios and on different teams.

5.2. Contributions to the literature on mental models

The aim of this project was to develop a new approach to assessing mental models of the order in which certain key tasks should be undertaken in an operation and of who should be responsible for them within OR teams. Further, the aim was to begin the process of validation of the new approach through examining a series of assumptions for which certain findings could reasonably be expected, and seeing to what extent those findings were obtained with Momento. Finally, the aim was to demonstrate how the new approach could potentially be used in clinical practice. Through addressing these aims, several contributions were made to the existing literature on the measurement of mental models in teams.

A major achievement of this project is its contribution to the assessment of mental models in healthcare teams. The participants were members of real OR teams, consisting of professionals of different backgrounds typically encountered in the OR. Further, OR teams consisting of individuals who typically work together took part. Mental models have not previously been studied in the context of established, multidisciplinary healthcare teams. The study by Burtscher et al.[56] remains so far the only previous study to have attempted to quantify the information on the similarity of mental models in the OR context. These authors, however, only focused on mental models of two-person anaesthesia subteams. Larger teams may have different dynamics than smaller ones. They may engage in different and more complex processes than smaller teams, sometimes with poorer coordination and higher conformity[225]. Larger teams are also likely to have a greater diversity of viewpoints and knowledge because of multiple team members[226]. Thus, the focus on larger multidisciplinary teams in the current project is a step up from the previous studies that have focused on mental models of small ad hoc teams[40, 54], or of only part of the team[56].

Although the Momento approach was inspired by assessment approaches employed in previous studies (see Chapter 1, Table 1), it offers several important advantages over these approaches, namely:

- The selection of tasks for the Momento task sort used an extensive multistage, iterative process guided by literature, informed by preliminary in-theatre observations of teamwork, and led by multiple subject matter experts. Such an exhaustive process of task selection has not been reported in similar studies on the assessment of mental models in teams. This process contributed to the content validity of the new approach.

- The Momento simultaneously captures the information on mental models of task sequence and of responsibility for task, but allows for the degree of similarity to be assessed separately for each model. This may help save time during data collection, while providing more detailed information on individual types of models.
- It provides insight into mental models within the entire OR team, rather than a single professional subteam. The similarity scores can subsequently be calculated at the level of the OR team, as well as individual subteams, allowing for richer data analyses.
- The Momento task sort involves all members of the OR team sorting all the listed key tasks in the surgical procedure in the order in which they occur, regardless of which subteam should be responsible for them and then assigning subteams to tasks. Previous approaches involved team members sorting tasks for individual subteams, making it difficult to discern the overall sequence in which tasks should be performed in the procedure.
- As well as allowing tasks in the surgical procedure to be sorted sequentially into a prespecified layout, the Momento allows for tasks that occur *in parallel at different points in the sequence* to be recorded, *regardless of which team performs them*. The result is a more complex and detailed concept map, allowing insight into more complex and detailed mental models.
- While the Momento task sort was designed in this work for two defined surgical cases, it was envisaged to be customisable to laparotomy for any cause. A standard method of measurement of mental models in teams that would be applicable in wider healthcare research settings (other than just anaesthesia subteams performing induction[56]) has not been reported in literature to date. In the Momento approach, this is made possible through having a set of generic tasks common to all general laparotomies where the answers to the same set of questions would vary given

different situations, and where a task may be positively dismissed as unnecessary by a participant. The tool also includes important 'anchor' points/tasks for timing into phases that define the sequence of the procedure and two to three options that could be scenario specific, but would not reveal the possible non-routine events that could be built into a scenario.

- The approach to calculating the degree of similarity of mental models of individual team members quantified using the Momento approach involves the computation of several well-established statistical measures of agreement, as opposed to a single arbitrary scoring method. Calculating the degree of similarity using multiple methods and comparing these allows for the most suitable and feasible approach to be selected that can then be applied in different contexts.

One strength of this research lies in the fact that all the participating teams performed the task sort for the same cases. This was made possible by the use of two simulated cases that participants were about to undertake, rather than clinical cases. Because participants were preparing to actually manage the simulated cases, it is likely they were more engaged with reading the case briefs and responding to the task sorting exercise, in comparison with just providing participants with written scenarios.

The findings of the main study of this thesis suggest that different types of mental models, different ways of calculating similarity scores for mental models, and different tasks can impact differently on the relationship between mental models and other variables, such as subteams in the OR. The findings thus support the notion that it is important to broaden the conceptualisation and measurement of mental models within teams through evaluating multiple types of the mental model construct, using multiple methods and in different contexts.

5.3. Implications for clinical practice

The findings of the study on the application of the tool of differences in the mental models of OR team members have implications for effective team function and patient safety. This is true both in terms of who is responsible for certain tasks, and the variation regarding the order of tasks in an emergency.

If used in clinical practice, the Momento approach has the potential to help elucidate and better align the mental models of OR team members about surgical procedures and thereby improve teamwork and outcomes for patients.

The Momento task sort is a computer application with a user friendly graphical interface that is easily customisable, allows for a quick task completion and automatic capture of results in an easily transferable format.

Theoretically, in the clinical setting, the Momento tool could be used prior to a procedure, so that members of OR teams individually complete the Momento task sort related to the upcoming laparotomy or other surgery. This could be done on portable electronic devices and with the Momento task sort as an online application. The outputs of task sorts of individual members would then be automatically consolidated and agreement on task sequence and responsibility for task calculated by the application for each task. Ideally, the whole team could then convene and discuss the points of disagreement during a pre-procedure briefing. This would allow team members to clarify potential ambiguities and re-align their mental models prior to starting surgery.

There are however, obvious barriers to this kind of application of the Momento approach. The biggest hurdle at this point is its applicability in a wide range of procedures, as key case-specific tasks would have to be determined ahead of time. For application in the clinical setting to be more feasible, the Momento would first need to be modified and tested for a wider range of surgical procedures. More realistically, therefore, because of the logistics, unpredictability, and time constraints characteristic of the clinical setting, the Momento approach is more likely to first find application in research and training of healthcare teams.

One implication of the findings of the study on the potential applicability of the Momento approach in clinical practice is that the low agreement between team members on the tasks in the surgical procedure could be rectified through team training. However, one issue that has been plaguing the field of mental model theory and research is the question of what constitutes an optimal degree of agreement or overlap of mental models in a team. It is still not clear what constitutes an optimal degree of shared understanding either conceptually[227], or in relation to the agreement scores obtained in this study. The Venn diagram analogy developed in this project to describe the shared mental model construct contributes to this discussion - conceptualising the degree of overlap of mental models of individual members as greater for teams with less specialised knowledge than for teams of specialists reflects the notion that there may never be perfect agreement in multidisciplinary teams working in high-risk domains such as in the OR. The relationship between the extent of similarity of mental models within a team and the team's performance has not been well defined[124]. Important decisions in an organisational setting presumably warrant a very strong degree of agreement among the decision-makers[228], but there is no empirical evidence to guide the quantification of this in multidisciplinary healthcare teams. The problem of optimal degree of shared knowledge in a team has been tackled by Cannon-Bowers, Salas, and Converse[30] who argue that fostering shared mental models in teams as much as possible can only lead to positive outcomes. Thus, having too little shared knowledge within a team can lead to poorly coordinated teams who are in turn likely to fail. This would hold true for both teams with more and teams with less specialised members. In contrast, having "too much" (p. 237)[30] shared knowledge could lead to excessive conformity resulting in incorrect team decisions. The latter would hold true for OR teams. Further, the advantages of teams with members with diverse backgrounds go beyond simply avoiding conformity, it allows a team to take on more complex tasks than expecting everyone in the team to be able to do everything. In this way, tasks or endeavours which require a larger skill domain which exceeds the skills of any one individual can be undertaken. Cannon-Bowers, Salas, and Converse[30] further argue that the extremes can be overcome if teams are equipped with strategies such as being taught assertiveness skills and provided with systems that alert them to alternative situation-specific explanations and solutions.

In training team members to develop similar mental models, it is not enough for individuals to be allowed to simply interact unaided with the system or to simply provide theoretical instruction[30, 100]. Instead, training must be enhanced by guided practice, where team members are provided with explicit conceptual models of the system, including explicit information on which specific major objects and actions are required, and in which order[30, 100]. Training of shared mental models should focus on generic, task-contingent competencies – these are competencies that are applicable to specific tasks regardless of team member configuration[229]. Thus, teams could be trained to develop mutual task-specific mental models that are generalisable to any team configuration that consistently performs that task[230]. It follows on then that the Momento approach could be used as part of the team training process, to improve team communication. Key surgical tasks with low level of agreement could serve to initiate discussion to ensure all members are "on the same page" as to the order of tasks and distribution of responsibilities prior to the team working together on the case.

Another example of the type of training into which the Momento approach could be incorporated is Team-oriented Medical Simulation that provides interdisciplinary team training for all members of the OR team, that can consist of didactic training, followed by the team taking part in a series of simulated cases, followed by debriefing[231-233]. In this context, the Momento could be used as a research tool for evaluating the success of team training interventions, to assess whether the mental models of team members have been re-aligned following intervention. As such, the Momento approach could be incorporated into existing comprehensive medical training programs[231, 234]. Unlike in the research setting, when assessing similarity of mental models in the clinical training setting, it may not be possible for practical reasons to use multiple methods of calculating agreement. Dynamic teams, such as OR teams, are often faced with time pressures. Thus, if the information on the degree of similarity of mental models is to be used to initiate discussion in order to ensure team members have a mutual understanding of what key tasks need to be done, when, and who should be responsible for them in the training context, a simple percentage agreement may be enough to identify the differences between team members.

Ultimately, by informing quality improvement initiatives in healthcare the new Momento approach to the assessment of shared mental models has the potential to reduce harm to patients.

With modification for particular procedures, the Momento task sort could next be used to identify dissimilarities of mental models for various other surgical procedures. Given its ease-of-use and customisability, the tool also has the potential to be used for team training and research in domains reliant on teamwork other than healthcare.

5.4. Limitations and future research directions

This research presents the first step towards a comprehensive and flexible approach to the assessment of mental models in multidisciplinary OR teams, with potential to be extended to other areas dependent on inter-professional teamwork. Although developing an approach for quantifying mental models meaningfully is important, this research has a number of limitations that could be addressed in future studies.

One area of future research is to continue the process of validation of the Momento approach. A single exploratory study may not be sufficient for validation of a new assessment approach, thus, establishing the psychometric soundness of the Momento approach is an ongoing process. For example, other measures assessing constructs theoretically related to mental models or alternative assessment approaches could be used in support for the convergent validity of the Momento approach. The assessment of concurrent validity could further be extended by comparing other professional groupings within OR teams. For example, the degree of similarity of mental models of the consultant surgeon and anaesthetist, or between the scrub nurse and the surgeons, could be examined more closely, as these team members form subteams of their own during various stages of surgery. Such analyses have the potential to further clarify the extent of shared knowledge between team members and identify areas or professionals who would benefit the most from training to align their mental models with the rest of the team.

The validation process could also be continued by testing the Momento approach on other types of laparotomy and comparing the findings to the current ones. This research was limited to cases requiring laparotomy. Future research should also extend this work to other types of surgery and surgical specialties where the same principles would apply but would have to be validated in those settings.

Future research should also focus on further investigation of the lack of significant effects found in this thesis. This could be accomplished in several ways. First, the definition of "team familiarity" could be extended to include the information on the quality of previous interactions with teammates, for example in terms of whether working with a teammate was a positive experience, or had led to positive outcomes. The study of potential antecedents of mental models should also be extended to investigate the relationship between other variables identified in the IPO model of teamwork (see Chapter 1, Figure 2) and mental models as assessed by the Momento approach.

The main limitation of the research presented in this thesis is that the extent to which similarity of mental models influences subsequent team processes (such as coordination), performance and patient outcome was not investigated. Although the positive relationship between the similarity of mental models and team processes and performance has been repeatedly established in previous empirical studies (see Chapter 1, section 1.2.2), there is a need for these relationships to be further investigated in healthcare teams. Consequently, this remains an area for further research.

The question of what constitutes an optimal degree of overlap of mental models has not been addressed in this research and remains unanswered. As with intelligence tests, using the Momento approach in future studies to gain insight into mental models of task sequence and responsibility for tasks on a larger representative sample of OR teams and calibrating scores against each other could potentially lead to a standard to be set for what similarity score is optimal for which type of mental model and in which context. As with the measurement of many other cognitive phenomena, there is the problem of arbitrariness of the scales used to quantify the similarity of mental models in this research. Although the scales used to calculate similarity scores were pegged to range from 0 to 100 percent, it is difficult to compare the different levels of agreement found in this research, and the existing approaches to assessing the quality of agreement[203] do not offer any more clarity. Further, even though the similarity scores for both types of mental models studied in this thesis vary between 0 and 1 (or 0 and 100 percent), because the similarity of mental models of task sequence and of responsibility for task were calculated differently, we cannot make direct comparisons between the scores on the two types of models. In fact, Momento is primarily a qualitative rather than a quantitative tool. Thus, although agreement on task sequence may appear to be higher than agreement on responsibility for task, the selection of task sequence was done on a different scale (1 to 20) and similarity scores calculated using different metrics from those for the selection of subteams responsible for tasks, where participants had to choose one of the four options.

The similarity of mental models in this thesis was assessed at the beginning of the two cases. Mental models of surgical procedures are likely to be dynamic and the degree of similarity of mental models may well change as cases progress and team members communicate with each other. The extent to which a team has a shared mental model may be greater in routine situations than in unpredictable emergent situations. This could be explored in future studies by assessing mental models of task sequence and responsibility for task pre- and post-performance.

Other potential limitations of this research include the possibility that there were inaccuracies in task sorting by participants. Participants did not have the option of choosing more than one OR subteam as responsible for a task, as the exercise used a forced choice design. The split within OR teams on who should be primarily responsible for certain tasks found in the tool application study might therefore reflect that there should be joint responsibility for those tasks. Future versions of the task sorting tool should be upgraded to allow for the selection of multiple subteams for those tasks for which there is likely to be joint responsibility.

This project did not set out to study the accuracy of mental models in OR teams. This can be construed as the degree to which these converge with those of guidelines or subject experts, or more generally the degree to which they are grounded in reality. Although accuracy was to an extent addressed by having various subject matter experts agree through an iterative process on a final set of key tasks for the task sort as in a previous study[41], there was no single accurate response as to the exact task sequence for the two procedures, especially as some tasks could be performed in parallel by different subteams rather than in a sequence. For many tasks in a surgical procedure, however, it may be more important that team members agree on what should be done, by whom and when, than that this agreement necessarily reflects received wisdom. In some instances however, accuracy in this latter sense probably does matter, and the findings of this research may support the idea that agreement tends to be higher in such circumstances. For example, the mean similarity scores in the tool application study were high for who should make the surgical incision and for who should administer the anaesthetic drugs.

6. CONCLUSION

This thesis described the development and the beginning of the process of validation of a new computer-based task sorting tool. The purpose of the tool is to gather information on individual mental models of multidisciplinary members of OR teams of the sequence of key tasks in the upcoming laparotomy and of who should be primarily responsible for those tasks. The new tool, named Momento, is easy to use and the similarity of individual mental models it captures can be calculated using several established metrics. More importantly, Momento as a tool illustrates the potential for a more general approach to exploring and mapping mental models about specific aspects of surgical operations. This is what I have called “the Momento approach.”

The specific outlined assumptions that would support validity of the Momento approach were largely supported in the main study of this thesis. No relationship was found between the degree of similarity of mental models of OR team members and team familiarity. However, the Momento approach distinguished between subteam versus inter-team professional groupings within an OR team in the case of task sequence in a less common laparotomy scenario, and for responsibility for task for both common and uncommon scenarios. Results for individual subteam versus inter-team comparisons also provide support for the validity of the new approach. Finally, validity was supported by the finding of similar relationships between similarity scores for mental models and team familiarity scores, and intra- versus inter-team groupings when different methods were used to calculate similarity scores.

In the study demonstrating how the Momento approach could be applied to inform clinical practice, differences were found in the mental models of some OR team members about responsibility for and order of certain tasks in an emergency laparotomy. Participants also believed their own subteam was primarily responsible for approximately half the tasks in each procedure.

Momento is a tool that can be used in team training and research. Ultimately, it has the potential to help elucidate and better align the mental models of OR team members about surgical procedures and thereby improve teamwork and outcomes for patients. The next logical steps for future research would be to extend the Momento approach to a greater range of surgical procedures, and evaluate its effectiveness as a tool in team training research and intervention.

7. APPENDICES

7.1. Appendix 1. Details of the MORSim scenarios 1 and 2

Scenario 1 -laparotomy for an abdominal stab wound

A stab wound with lacerated inferior vena cava (IVC) complicated by cardiovascular collapse

Background:

0800h on a weekend. Stabbed patient urgently brought into theatre for exploratory laparotomy for knife in situ.

Unknown male (Ian Peterson) is a 48 year old (weight unknown) obese man with a history of paroxysmal atrial fibrillation, hypercholesterolaemia, hypertension, iv drug use, but almost no records in the hospital system (recently moved into this catchment area). Medications include warfarin, statin, cilazapril, methadone. He also has a history of recreational drug use, current smoker and alcohol abuse. Vague when asked about history, medication, allergies, reflux. Airway unremarkable.

Ian was stabbed when a “P”-party went out of control; partner injured also. Brought to ED by ambulance, weapon in situ periumbilical. Ian is high on marijuana and drunk so confused and is admitted as “Unknown Male” so no old computer records are seen initially.

Trauma call, A B D E ok, C moderately shocked, isolated injury, iv lines, CXR unremarkable, initial bloods for Gp and Screen, coag, toxicology EtOH. Arterial line and IDC inserted, FAST scan not indicated, no other imaging, rapid transfer to OR by ED registrar and ED nurse, trauma team currently managing second victim.

Participants receive briefs and asked to go to OR to perform exploratory laparotomy.

In OR, patient on operating table with confederate ED registrar and confederate OR nurse (ED nurse has left) who hand over to anaesthesia and nursing team. Patient identity now confirmed and new bracelet applied, new stickers available but old bracelet and stickers remain in folder.

Nurses should continue preparations, sign in. Anaesthetist should assess, check preparations, and begin induction. Surgeon/s should prep and drape prior to induction. (This can be forced by worsening hypotension). Timing of time-out check will be problematic. Anaesthetic tech will be busy with monitoring, fluids and induction.

Anaesthetist and surgeons will have challenges with massive venous bleeding when knife is removed. Surgeons will have challenges with finding/suturing IVC injury; nurses and tech will have high taskload.

Once bleeding is controlled by clamp/ pack, some stabilisation occurs but is followed on further exploration by signs of venous air embolism with sudden marked hypotension responding to control of possible VAE from open IVC, and supportive measures.

Scenario ends after 40 min.

Start sequence:

- Written briefs to all participants
- Participants to complete the Momento task sort individually
- Participants to be taken into the simulated OR
- Faculty ED registrar to give handover to participant anaesthetist and leave
- Faculty OR nurse will stay in theatre and handover to incoming theatre nurses.

Scenario 2 - laparotomy for a suspected perforated viscus

Appendicitis complicated by sepsis and subsequent allergic reaction

Scenario background:

0800 on a weekend. Patient brought into theatre for urgent laparotomy for septic appendicitis.

Brian Richards is a 70 year old (90kg, BMI 33) man from UK visiting family. History of mild asthma, hypertension, ischaemic heart disease. Not currently wheezy. Medications include fluticasone MDI, salbutamol MDI prn, aspirin, and diltiazem, isosorbide mononitrate, gingko and garlic. NKDA. Airway unremarkable. No history of reflux.

Patient had abdominal pain for 24 hours. Felt more unwell overnight with vomiting and abdominal pain. Arrived from UK two days prior and saw GP with sore leg, was sent home.

Brought in by family to ED at 0630. Required resuscitation with iv fluids. Free air on x-ray. All lines inserted. Low dose noradrenaline started in pre-op.

Written briefs are given to all participants. Scenario starts in OR with patient on operating table with ED bed next to it. Nurses to prep after being handed over by faculty theatre nurse (who stays in OR). Faculty anaesthetist hands over to participant anaesthetist. Anaesthetic tech will be busy with monitoring, fluids and induction.

Patient will be moderately hypotensive after induction.

Once vasopressor and fluids are given post-induction, some stabilisation occurs but is followed on further exploration by further marked hypotension, tachycardia and raised airway pressures due to allergy. Adrenaline and supportive measures will be needed.

Expected discussion about where the patient will be transferred to after surgery.

Anaesthetist will have challenges to deal with acute abdomen and hypotension.

Surgeon will have challenges with finding cause of sepsis – caecal perforation.

Technician will be busy assisting anaesthetist.

Scenario ends after 40 min.

Start sequence:

- Written briefs to all participants
- Participants to complete the Momento task sort individually in the computer room
- Participants to be taken into the simulated OR
- Faculty anaesthetic registrar to give handover to participant anaesthetist and leave
- Faculty theatre nurse will stay in theatre and handover to theatre nurses/tech who come into OR to help setup.

7.2. Appendix 2. Exploratory in-theatre observations

This section provides a more detailed account of the exploratory observations of tasks, roles and team dynamics in the OR carried out during the surgical cases outlined in section 2.3 of this thesis. The observations are summarised into two main categories: a) procedural steps and team members' roles; and b) coordination of information and activities.

a) Procedural steps and team members' roles

Many steps in the surgical procedure proceeded in parallel fashion, as different subteams focused on different tasks at the same point in time. For example, as the scrub nurse prepared the surgical trolley, the anaesthesia team and the remainder of the nursing team positioned the patient on the operating table and attached monitoring; while the surgical team operated on the patient, the scrub and the charge nurse performed an instrument count; while a nurse laid out the correct sterile gloves for the surgeons, anaesthetic technicians (ATs) put on the air warming device on the patient, etc. The preparatory stages of a case were characterised by various independent parallel activities with little explicit interaction, especially inter-team. Numerous examples of what Manser et al. [23, 49] refer to as "coordination via artefacts" (for example, TEE, monitor, infusion pump, patient chart, white board) part of "coordination via the work environment", were observed in the preparatory stages. In one case, the incorrect procedure and estimated surgery time were written down in the operative schedule. This was picked up and commented on by different subteams and finally explicitly shared inter-team at the "time out" phase of the Surgical Safety Checklist (WHO).

Several crucial sequential tasks were observed for all the cases, where one task could only commence after the success of another. Thus, for example, the "sign in" part of the

WHO Safe Surgery Checklist was always completed before the induction of anaesthesia; the surgeon commenced the surgery (i.e., made a skin incision) only after the confirmation with the anaesthetist that the patient had been induced and stable. In one case, no ‘time out’ was performed. The five stages of the surgical process [235] were also observed, namely: procedure set-up (i.e., where non-medical staff bring in supplies, arrange and prepare), surgical preparation 1 (patient brought in; prepared and positioned), surgical preparation 2 (induction/intubation), intraoperative phase (surgical intervention), and handoff phase (preparing patient for transport - transition to the next level of care).

The activities of individual team members appeared highly structured at certain stages of the case. For example, the patient was almost universally greeted in the OR by the anaesthetist, AT and a nurse; where a spinal epidural was required, a nurse always stood in front of the patient ensuring adequate positioning, while the anaesthetist worked on the patient’s back; the AT was most frequently observed to be the first person in the OR, setting up drugs and equipment. The surgical team was in most cases absent during the procedure set-up stage, and had in some cases entered the OR only once the patient had already been sedated. In all the laparotomies, the consultant surgeon always came into the OR to visualise the patient’s abdomen while the patient was still awake, and then left to scrub up.

Apart from the positioning of the surgical team (i.e., between the patient’s legs, as in the urology cases) and there not being an abdominal incision and steps associated in preparing the patient’s abdomen for a laparotomy (as in the vascular surgeries), the only obvious observable differences in procedures and roles that could potentially affect team coordination were noted in urology surgery. Here, the charge nurse seemed much more involved directly with the consultant surgeon, obtaining and handing equipment on the surgeon’s request, than in non-urology surgery. There was much less sterile instrumentation and the urology cases were the only ones observed where the nurses draped the patients with no involvement of the surgical team. It is difficult to ascertain whether the dominance of the charge nurse in the urology cases was due to her dominant personality or the nature of the procedure, although the former may be more likely, as it was conveyed to the observer that the same nursing team, in a similar configuration, also worked on general surgery cases.

Due to the institution being a large teaching hospital, there were trainees in every procedure for different subteams.

The anaesthesia subteam

In all the observed cases, the anaesthetist, and in most cases the AT were present in the OR prior to the patient being brought in, with the AT being usually the first to start the preparations. The anaesthetist often left the OR with the patient and accompanied the patient to PACU. Teaching was observed to be most prolific in the anaesthesia team, often characterised by supportive, positive comments and acknowledgments.

Following the induction/intubation phase, apart from teaching and informal chatting, there was hardly any overt coordinative activity to observe involving the anaesthesia team until the reversal of anaesthesia and the handoff. The intraoperative phase was characterised by the anaesthetist often monitoring the patient's vitals, or by observing the surgical procedure – an implicit behaviour Manser et al. categorise as "coordination via the work environment" through monitoring the surgical team [23, 49]. ATs were often involved in equipment stocktake and sending samples to the lab as initiated by the anaesthetist. Anaesthetists were most frequently in charge of organising and following up on patient transfer to PACU. This was often delegated to the anaesthetic registrar by the consultant anaesthetists. In one laparotomy case, the anaesthetist was the one to explicitly initiate the start of the surgery ("the operation can start"). In some cases, following "time out", the consultant surgeon asked the anaesthesia team if the surgery could start. The AT rarely interacted inter-team, with members from the other two subteams. On one observed occasion, the AT assisted a charge nurse to enter patient details into the computer, and on another, an AT trainee asked a nurse to clarify procedures regarding patient's sample forms. There was only one observed example of an interaction between an AT and a surgeon, where the surgeon asked the AT to relay a message to the anaesthetist.

The surgical subteam

In all the non-urology cases, surgeons carried out the draping of the patient, sometimes assisted by a scrub nurse. Applying an antiseptic solution to the patient's abdomen was usually done by surgeons, although in two cases this was done by a scrub nurse, and only after she asked the anaesthetist if she could do so in one long laparotomy case, and the charge nurse in a urology case. Interestingly, the scrub nurse taking on the abdomen prepping was not explicitly discussed with the surgeons, but rather the scrub nurse seemed to have anticipated the surgeons' workload and stepped in to assist. In four vascular cases, the same surgeon was present throughout the induction/intubation phase. This was the only example of a surgeon being present and observing the actions of the anaesthesia team during their busiest time.

The consultant surgeon often discussed the surgical plan with the surgical registrar just prior to the commencement of the surgery. The intraoperative phase was the busiest time for the surgical team, with little inter-team interaction. The verbal interactions within the surgical team were however difficult to follow, as their conversations were often muffled due to the surgical masks and the surgeons leaning over the patient, away from the observer.

In the urology cases, the surgeon was pivotal in deciding the positioning of the patient, which, in one case, was anticipated by the charge nurse ahead of time:

- Surgeon: "This patient is going to be a bit of a challenge" [positioning-wise]

- Nurse: “Yeah, this is why I said you should come and do it”

The nursing subteam

The nursing subteams differed from the other subteams in the flexibility of their roles, where nurses routinely took turns at being the scrub nurse. In one long case, the three members of the nursing team had a discussion at the beginning of the day on who was going to do what, making the allocation of roles a diplomatic process.

The circulating or the charge nurse was often observed to have a dominant personality, coordinating everyone’s activities, delegating tasks, monitoring the nursing team and anticipating the surgeons’ needs ahead of time. An extreme example of this was observed during one urology case, where the charge nurse volunteered a comment to the observer, regarding her setting up the x-ray machine:

“What I’m doing here is not a nursing job. But I’m the only one that can get away with it.”

In this study, nurses were frequently observed to be fierce at protecting the sterility of the environment in the OR. For example, in one long laparotomy case, a medical trainee on a surgical rotation was about to approach the operating table after gowning and gloving. The charge nurse noticed he had not tied the gown around properly, spoke up (”Can I check your gown please?”), and assisted in proper gowning. During the same case, the same nurse addressed the trainee again, for standing close behind the surgical registrar (“Hey [trainee name], [surgical registrar’s name] is not sterile at the back, so you have to be really careful, alright?”). She then approached the surgical registrar and tightened his gown at the back.

Examples of nurses protecting the sterile field in relation to their fellow nurses were also noted. In one case, the scrub nurse reacted to the circulating nurse coming too close to the surgical trolley ("Could you put some sterile gloves on, please?")

Nurses in established teams were also observed to not be afraid to question and speak up against the decisions of members of the other two OR subteams. In one case, the charge nurse argued against the surgical registrar wanting to change a patient's dressing (not related to the surgery site):

- Surgeon: "We need to change his dressing at the back."
- Nurse: "What for? It shouldn't be dislodged."

The surgeon lifts up the dressing; blood gushes out.

- Nurse [shouting]: "What are you doing???"

In another case, a charge nurse demonstrated a good knowledge of the surgeon's role and acted accordingly:

- Charge nurse: "You would do much better there if you took those top gloves off"
- Surgeon: "I would actually, yeah" [Laughs and takes top gloves off, tosses them aside].

In the cases observed, the surgical checklist ("sign in", "time out", and "sign out") was overseen by a circulating nurse, except for one long laparotomy case where the "sign in" was performed by an AT. In only two cases, the surgeon explicitly called for "time out" that was then performed by a circulating nurse. In one case, as he was leaving the OR, the consultant surgeon explicitly asked the surgical registrar if he would "do 'sign out' with the team".

Overall, the nurse seemed to be in charge of ensuring the surgical checklist progressed smoothly. For example, in one case, the charge nurse initiated time out; everyone said their names, except for the medical trainee who did not seem familiar with basic procedures. This was promptly corrected by the charge nurse (”You have to say your name, we’re introducing ourselves”).

The nurses were also observed to be the ones to let the coordinator know of the progress of the surgery and when the team was ready for the next case. The nursing team was also in charge of preventing unnecessary distractions, such as answering the surgeons’ private phone calls and taking messages.

One striking characteristic that was observed of the nursing role was multitasking. Examples of this include a scrub nurse in a vascular case chatting informally with the anaesthetist while simultaneously and effortlessly handing instruments to the surgeon; a scrub nurse handing instruments to the surgeon while performing an instrument count with the charge nurse; or the charge nurse being handed new supplies by the circulating nurse in the middle of a count. Closing up of the incision was observed to be the busiest time for the scrub nurse, and one with the highest requirement for multitasking and interaction with both the nursing and surgical teams.

b) Coordination of information and activities

Most procedures were characterised by a relaxed and often jovial atmosphere. This was especially the case when the team members appeared to know each other well.

In the case of one team, a scrub nurse, who appeared to be a trainee, stood out as an outsider, needing a lot of guidance from both the surgical and nursing teams, and eventually a charge nurse to step in as a second scrub to assist her. A lot of the instruments were not prepared in advance and had to be obtained from outside the OR by another nurse throughout the intraoperative phase. Consequently, this situation led to intensified intra-team coordination for nursing and some tension between the nursing and the surgical teams.

Explicit coordination

Occurrences of clear, explicit, closed-loop communication were observed on numerous occasions, characterised by the sender of the message making sure that the message was correctly received through explicit verbal confirmation, such as in the following examples:

Example 1:

- Charge nurse (CN):"Periportal tissue? [Holds up an open jar to scrub nurse].
- Scrub nurse (SN): "Periportal tissue? [Surgeon's name]?"
- Surgeon (S): "Yep, periportal tissue."

SN repeats this again, while being handed the tissue specimen by S, and drops it in the jar. CN closes it, and puts it on the trolley.

Example 2:

- CN: "Do you want to look at the specimen [S's name] or can [SN] hand it off?"
- S: "Can hand it off"
- CN: "So, that's a left?"
- S [mumbles]: "Ughmm."

CN waits

- S: "Yeah."

CN takes the specimen and puts the patient's label on the specimen container.

In another example, the circulating nurse (CcN) was checking about the patient's personal items on the computer file:

- CcN: "Personal items?"
- CN: "Just dentures."
- CcN: "Is it dentures?"
- CN: "Yeah."

CcN types this information in patient's file.

AT: "Dentures are here." Hands them to CcN who puts them on the computer table.

- CcN: "Thanks."

The usefulness of making coordination explicit, closing the communication loop, and the use of gestures to back up verbal communication in order to prevent potential adverse consequences was evident from the following example of the nursing team performing a swab count:

- SN [to CN]: "3 swabs."
- CN: "Where?"
- SN: "In the cavity."
- CN: "Is that in the cavity?" [writes it down on the board]
- SN: "Yeah."

A while later, SN calls out to CN: "There's 2" [lifts up 2 fingers].

- CN: "Not 3?" Goes to the board to change the record.

An attempt was made to classify specific examples of explicit coordination behaviours according to categories described by existing observation systems [23, 49, 113]. Examples of specific categories are provided as follows.

Explicit task management/coordination of actions

Task distribution - delegation:

Anaesthetist (A) [to AT]: "You concentrate on the airway, I'll do the bagging."

Task distribution – giving orders:

- S [to SN]: "Never load it like that. I can't put a needle in."
- "See those long ends [holds up sutures]? I want them all long ends. I'll reuse them."
- "Do the top one [drape] first. Change your gloves for me in the end."

Sequencing, task distribution/ delegation:

- A [to nurses]: "Can someone let them know we're going off to sleep, but I still have a central line to go in after that?"
- N: "I wouldn't take that [drapes] off quite yet cause there's still a bit of a wound there."

Initiating an action:

- "Shall we do a sign in?"
- "Should we do time out, team?"

Clarification/planning and procedural questions:

- CcN: "Alcohol?" [Holds up a bottle].

- CcN: “Does he [S]...? Likes it?”
- CN-CcN: “No, he likes ___”

“They know we’re sending down another one [sample], don’t they?”

“Is there specimen collection or not?”

“What other blood product do you want?”

Verbal requests for assistance/explicitly asking for help:

“Pack please, I need a pack.”

“Can we raise the table, please guys?”

“Gown off, please.”

<u>Explicit information management/exchange</u>

Decision making: make/state/(re-)evaluate a decision:

“Right. We are going to turn him on the side to change his dressing.”

“Alright, I think we’re in a position to intubate.”

“I think we should send it [blood product] back.”

“She should have full blood count repeated around 8-9 o’clock.”

“It’s not bad. In fact, what he should be doing is lay on his side.”

Decision making: stating and questioning a decision:

- S: “We need to change his dressing at the back.”

- N: “What for? It shouldn’t be dislodged.”

Information request and giving information after request:

- AT: “Where’s this patient going afterwards?”
- A: “Well, to be reviewed in PACU.”

Verifying information/confirmation /clarification:

“I can’t read your writing. Is that English? I’m just getting clarification.”

“This guy’s on the ward, right?”

Approval:

“Good job everybody!”

Situation assessment/verbalising interpretation of a situation:

- A: “So, the bed’s gonna have to go on this side.”
- A: “He’s not quite there yet.” [about the patient post-anaesthesia reversal]
- A: “It’s not the best of airways.”
- S: “There’s a very big loop over there. Very big.”

Teaching

“Watch the chest, if you squeeze too hard, it’s gonna leak.”

“You just hold it in such a way so they can just grab it” [shows with body language to another nurse how to hand over instruments to surgeons].

Implicit coordination

Common to all the observed cases was the tendency to make a verbal request without explicitly identifying the recipient. Rather than being directly addressed at a particular recipient, it appeared as if the content of the message, often accompanied by a slightly raised voice or a head turn in the general direction of the recipients, implied who it was addressed to. This was especially the case when one individual, most often the surgeon, wanted to address a member outside his immediate team. For example:

- S [with a slightly raised intonation]: “The specimen’s gone, yeah?”
- N: “Yeah.” [implicitly knows message is directed at her, due to the content of the message].

Or:

- S: “OK, continue screening, please.”

Radiographer (R): acknowledges, double-checks verbally [“Screen?”]

- S: “Yes, please.”

- S: “Can we have lights dimmed please?”
- CN: “Yes, of course.”

- S [not lifting his head up from the operating table]: “Hey, can you put this warmer down, guys?”
- A: “Yes.”

- S [raised voice]: “AM I ok to start?”
- A: “Yeah.”

Occasionally, the request would be met without verbal acknowledgement from the recipient:

S: “Can someone clamp the nephrostomy tube for me?” N does it.

A [to no one in particular]: “Head rest?” S hands over the head rest

The above form of coordination was often the norm in the coordination of activities between the surgeon and the scrub nurse, who in the intraoperative phase seemed to coordinate as a team of their own. Although there was a lot of closed-loop communication between the two, characterised by verbal requirements from the surgeon for instruments, followed by the scrub nurse responding by actions, and surgeon’s acknowledgements upon being handed them over (for example, “thanks”), the majority of coordination progressed in the implicit fashion. For example, the scrub nurse handed an instrument when requested, but also often held up another in anticipation of surgeon‘s future needs. Occasionally, the scrub nurse demonstrated the ability to anticipate the surgeon’s needs well ahead of time:

SN [to CcN]: “See if you can find scissors longer than that” [lifts up a pair of scissors for CcN to see].

SN: “How about we check for ultrasound?”

The scrub nurse rarely verbally verified the surgeon’s requests:

- SN: “ [names an instrument]?”
- S:” Yes, absolutely.”

The scrub nurse also acted as a backup for the surgical team, making sure the surgeons' requests had been heard and the nursing team responded correctly:

S asks for a piece of equipment. CN gets up, repeats what S said, walks over to the supplies cabinet. Then confused, walks to another cabinet.

- SN –CN: "What are you looking for?"

CN names equipment

- SN: "No, he asked for ____"
- CN: "Oh, I thought you said _ _ " [then finds the correct piece of equipment].

Occasionally, however, the scrub nurse was faced with ambiguous requests she found difficult to decipher:

S asks SN for an instrument [a long clamp].

SN hands him an instrument.

- S: "No." [motions with a twisting of the hand as if holding scissors and cutting]

SN hesitates

- S: "No, the green one."

CN goes through the packed instruments, finds the one the surgeon then confirms he asked for, and hands it to SN by opening its packaging.

In rare instances, the scrub nurse did not correctly anticipate the surgeon's request:

SN about to hand over retractors. Pauses.

S gestures to SN.

SN [chuckles]: "Oh" [Gets another instrument].

Surgeons' needs being incorrectly anticipated was most often associated with a scrub nurse being a trainee or being new to the team. In one observed case, for example, the surgeon silently showed his disapproval by getting an instrument from the surgical trolley himself without repeating his request to the scrub nurse who repeatedly held up the wrong instrument; or the scrub nurse failing to anticipate the surgeons' needs and subsequently failing to set up the correct instruments for the procedure, or failing to provide the required assistance:

- S:"Give me a Morrison"

SN hesitates

- S:"Do they not have any other retractor? Are they not set up for 2 surgeons to operate?"

- S [to SN]: "Suction please."

SN tries to hand him the suction.

- S : "Suck. You need to suck."

Implicit coordination often took a nonverbal form. An example of this is a charge nurse moving equipment out of the way in anticipation of the operating table being rotated, or a nurse moving the diathermy machine to make space for another surgical registrar who was about to step in – an assistive task management behaviour [23, 49], also referred to as "providing unsolicited task-relevant actions" [113]. Another common example of this type of implicit behaviour is nonverbal offering of assistance with gowning – for example, a nurse, who was working on the computer, noticing the surgical registrar putting on a gown and walking over to him to assist with tying it.

Similarly, examples of nonverbal requests for assistance [23, 49] were also observed:

SN stands in front of CN, turns her back to CN.

CN undoes SN's gown and walks off.

Understanding of each other's roles, equipment and procedural steps, both intra- and inter-team, were often evident from team members being able to decipher ambiguous or unfinished requests from their colleagues, or being able to offer assistance having correctly anticipated their needs:

- S: "Do you have one of those 10 french dilated things?"
- CN: "Yep" [leaves and brings the correct piece of equipment].

- SN [stands behind S]: "Are we putting …[pauses, unfinished]?"
- S2: "Dressing. Yes, dressing."

- N2 [addressing A]: "Do you want us to …. [unfinished sentence, pause]?"
- A: "Catheterise? Yes, sure."

- SN [addressing CN]: "Ann. Ann."

CN looks over. SN lifts a swab up for CN to see, wiggles it in the air.

CN nods. Writes something on board.

A new surgical registrar steps in next to the consultant surgeon and immediately starts suctioning, with no verbal communication.

Numerous examples of offering unsolicited assistance verbally, often in anticipation of others‘ needs were also observed:

- CN [to A]: “Do you think she needs some cuffs on her legs?”
- CN [to SN]: “[Name], do you want some saline?”
- CN [to SN]: “[Name], are you alright? Do you want me to take over?”

Examples of providing information without request, or unsolicited task-relevant information were also observed:

- S: “Hey, this guy, you can’t roll him on his left, can only roll on his right.”
- N [to AT]: “I’m just gonna get some iv dressings.”

Team members occasionally also provided unsolicited information about themselves or declared their own needs:

- “This is my first case of the day, I want to do this right.”
- “I did nothing today!”
- “Haha, I need a foot stool.”

Coordination by “monitoring other crew members” [49] was a common occurrence, and this was especially amplified if a member of the same subteam was a trainee. The monitoring that occurred implicitly (i.e., watching a colleague perform an action) was usually confirmed verbally:

CcN watches SN and CN performing the initial instrument count, and finally comments: “Artery opening?”

- SN: “I don’t have artery opening.”
- CcN: “That’s fine.”

…and in the same case:

- CN-SN: “I think you have one that is not working well.”

“Coordination by listening to communication of others”[49] was also observed. For example, during one case, a charge nurse observed an inexperienced scrub nurse handing instruments to the surgeons:

S asks SN for an instrument [assistance]

- SN: “Sorry?” [clarification]

CN repeats the name of the instrument that S asked for to SN

Or:

- S-SN: “Self-retainer?”
- CN: “Yeah, it’s coming.”

S [looks at SN, who just handed him the incorrect instrument]: “No, no.”

- CN [at SN]: “No, no that one. Have you had a mental block or something?”

Examples of “correcting behaviours of other team members”[113] mainly through monitoring of their activity, and in order to prevent potential adverse consequences was also observed:

- “You have some of your instruments mixed with your swabs. Just make sure you don’t throw them out.”

Or as in the following examples:

- S [to SN, upon being handed an instrument]: “This is not correct.”
- CN[to SN]: “Do you actually have 2 in your set? You have 2, use the smaller one.”

- CN[to SN, during a swab count]: “No, I can’t see. You have to lift them up and show them to me.”

- CN [to SN]:“Did you count your sets?”
- SN: “No.”
- CN: “Well, we probably should have, cause they’re going into the groin.”

- N1 to N2: “There’s blood on the floor. Hey [N2’s name], just a thought. You shouldn’t push the trolley out until the case is finished. The blood canister is still attached.”
- N2: “Where is it?”
- N1: “You left it in the corridor.”
- N2: “I’ll go and get it.”
- N1: “At least the canister – there’s still blood in it.”

In another example, one nurse reprimanded another (who was an outsider to the team) for not having washed her hands again after touching the supplies cabinet and then handling sterile equipment. In a subsequent episode, the new nurse touched the sterile drapes with no gloves on while being monitored by the same nurse, and was reproached again:

- N [to new N]: "Why did you just do that? Because you just touched that before you put your gloves on, so that's now rubbish."

Monitoring for the purpose of being ready to step in and provide back up, was also observed. For example:

- CN [to SN, after having observed her at work for a while]: "[Name], are you alright? Do you want me to take over?"

- CN [to SN]:"Do you want me to get another one?"

In one case, a charge nurse provided back up to the scrub nurse by clarifying with the surgeon the name of the instrument he asked for:

- S [to SN]:"Please [name of the instrument]."

SN hands him an instrument. S pauses, looks at CN, who is monitoring the exchange, and passes the instrument back to SN:

- CN [to S]:"Did you say [repeats the name of instrument]?"
- S:"Yes." SN hands S the correct instrument.

Monitoring for the purpose of starting a task dependent on a successful completion of a prior task, often performed by a different team member, was also observed. Examples of this behaviour include an anaesthetist observing the surgeons' progress in order to identify how soon to start reversing the anaesthesia, or a charge nurse waiting behind the sterile trolley for the scrub nurse to obtain a tissue specimen from the surgeon, so that she can hand it to the circulating nurse who then needs to label it and send it for processing.

Examples of "coordination via artefacts"[23, 49] were also observed, such as nurses rechecking the instrument count written on the white board, followed by the following process information transfer:

- CN [looks at the white board, to SN]: "[Name], are those [points at writing] out?"
- SN: "No."

In another case, one nurse noticed another nurse had left a machine power cord sitting in a water puddle on the floor. She moved the cord, got a towel and wiped the floor.

Coordination breakdowns

Problems in team coordination were often observed in those cases where there was 'an outsider' present in a well-established team, or when there was a trainee present. In those cases, it appeared that the process was interrupted due to those involved not having a shared understanding of the situation and one team member not being able to correctly anticipate the needs of another. Apart from the example of the scrub nurse disrupting the flow of surgery by handing incorrect instruments to the surgeon, several examples of breakdowns in coordination were observed within the nursing team. In one episode, a circulating nurse walked into the OR holding a bottle of saline. She then looked at the charge nurse and the scrub nurse while opening the bottle, as if seeking confirmation nonverbally. After not having received any acknowledgement from either of the two nurses, she finally approached the scrub nurse and diverted her attention by offering her the saline:

- SN [to CcN]: “Don’t need another one, sorry.”

CcN leaves the open bottle on the computer table without telling anyone and walks out.

In other examples, the scrub nurse incorrectly assumed that the charge nurse would implicitly understand what she meant:

- SN-CN: “2 steri strips, please.”
- CN brings and opens a package. SN [looks at them]: “Oh, not these - the blue ones.”

SN [lifts up a piece of dirty gauze, looks at CN]:

- N3 [walks over to supply cabinet]: “How many pieces?”
- SN-N3: “Oh, no – I meant swab count.”

Disruptions to the coordination process were also observed to occur due to a team member being reluctant to engage in inter-team coordination, engaging the members of his own subteam instead:

- S: “Can we open the nephrology tube?”
- CN looks at SN, reaches behind for a catheter. Holds it up for SN to see.
- S: “No, no, no. The NEPH tube. It’s in the kidney. It’s been turned off.”

During a long case, the circulating nurse tries to convey a telephone message meant for the surgeons via the scrub nurse:

CcN gets off the phone, stands behind the surgeon and calls out to the scrub nurse.

- S [to CcN]: "Why are you calling her when there are two consultants and a reg in the room? Speak!"

In another case, a circulating nurse was also reluctant to talk to the surgeon directly:

- CcN: "Alcohol?" [Holds a bottle up, turned towards surgeon but looking at the CN].
- CcN: "Does he [S]...? Likes it?"
- CN-CcN:"No, he likes ---"

In one case, the surgeons passed on a message to the circulating nurse about which instruments they will need, assuming this will be relayed to the SN:

- CcN [to SN]: "Oh, didn't they tell you about it? I told them, don't tell me to relay the message – you tell them."

The reverse was also observed, where the scrub nurse relied on the circulating nurse to answer a question intended for the surgeon:

- SN looks at CcN: "Can I paint now?"
- CcN: "Sure you can paint."

Examples of communication related distractions, or "case-irrelevant communication"[236], or non-coordination related communication or "informal chatting"[49] were also observed. These often involved requesting information about another patient or telephone calls to surgeons' private phones – distractions that the circulating nurse often prevented by answering them and taking messages she conveyed to the intended recipient only if urgent or case-related. In other cases, interruptions to the case were due to the coordination of the overall work process which involved talks about the changes to the schedule, cancellations of the last case on the list, discussions on other surgeons taking over the subsequent case, or requesting updates on the current case from outside of the OR. This often involved people outside the OR team coming into the OR and disrupting the coordination flow.

7.3. Appendix 3. Case briefing sheets

Scenario 1 – Ian Peterson (abdominal stab wound)

Briefing sheet for surgeon A

It’s 0800h on a weekend. You have just turned up at work for a ward round and have been called by the on-call night registrar. He advises you about a case requiring urgent attention. Ian Peterson aged 48 came in about an hour ago with a stab to the abdomen. He was seen in ED as a trauma call. The injury seems isolated, the patient is shocked but stable with modest fluid requirements. He’s on the methadone program. Others were brought in with less severe injuries. The surgical registrar has sent him straight up to the OR and called to ask if you can do the exploratory laparotomy.

Briefing sheet for surgeon B

It’s 0800h on a weekend. You have just got to work and been called by your colleague to assist in theatre. A victim was brought in with a stab wound to the abdomen after some sort of incident at a “P” users’ (methamphetamines) party. He is shocked but stable with modest fluid requirements and is coming up to OR from ED. The rest of the theatre staff is on the way to theatre and will meet you here.

Briefing sheet for nurse A (CN)

It’s 0800h on a weekend. You’ve just arrived at work as acute OR charge nurse for the day. The nursing coordinator tells you there’s an abdominal stab patient coming up from ED for urgent laparotomy. The knife is still in situ and the coordinator relays a message from the police to cover and preserve fingerprints on the knife. The night team has already begun setting up for the case and you are asked to go to OR to help.

Briefing sheet for nurse B (scrub nurse)

It’s 0800 on a weekend and you’ve just arrived at work. The nursing coordinator calls you and tells you there’s an abdominal stab patient with hepatitis C coming up to OR for urgent laparotomy. The night team has already begun setting up because of the urgency. She asks if you could go quickly to scrub for the case.

Briefing sheet for anaesthetist

It’s 0800h on a weekend and you’ve just arrived at work. The nursing coordinator tells you there’s a male patient with an abdominal stab wound for urgent laparotomy. The ED reg has brought him up and they’re in OR waiting for you. He’s shocked but stable with modest fluid requirements with knife in situ. The only records in the system on him are dispensing records as per this printout. Dispensing notes dated 9 months previously: Warfarin, cilazapril, paracetamol, simvastatin.

Briefing sheet for anaesthetic technician

It's 0800 on a weekend and you've just arrived at work. You heard there's an urgent case coming up to your theatre - an unknown male stabbed in the abdomen. The night tech tells you they have just completed a level 2 check of the machine, the patient has an A line, and the Belmont is ready to go outside OR. Blood bank rang through to say they were processing the sample and the massive transfusion protocol had not been activated. You are asked to go to OR and start the case.

Scenario 2 – Brian Richards – suspected perforated viscus

Briefing sheet for surgeon A

It is 0800h on a weekend. You have just turned up at work for a post-acute ward round and have been called by the night registrar. He advises you about a case requiring urgent laparotomy for suspected perforated viscus. The patient has had 24 hours abdominal pain, worsening overnight and he came in about three hours ago to ED quite unwell. He was seen at White Cross two days ago complaining of a sore leg after arriving from London. O/E guarding, febrile and x-ray showed free air under the diaphragm. Theatre nurses have sent for him and he should be in theatre. Please could you go there to perform the operation.

Briefing sheet for surgeon B

It is 0800h on a weekend. You have just turned up at work and you are called by your consultant to assist in theatre. A patient, Brian Richards, needs an urgent laparotomy for suspected perforated viscus. He is quite unwell and has medical problems. A second group and screen sample is needed as the first sample was not labelled correctly. Other theatre staff are on the way to theatre and will meet you there.

Briefing sheet for nurse A (CN)

It is 0800h on a weekend. You have just turned up at work and are the charge nurse in acute OR for the day. The theatre coordinator has sent for a patient, Brian Richards for an urgent laparotomy for suspected perforated viscus. She also mentions that he is quite unwell, and that CSSD rang a short while ago to say that there's a problem with the steam – it's out of action for now. Other staff are on the way to theatre and will meet you there.

Briefing sheet for nurse B (scrub)

It is 0800 on a weekend. You have just turned up at work. The floor coordinator has sent for a patient, Brian Richards is ready to go for an urgent laparotomy for suspected perforated viscus. She also mentions that he is quite unwell, and that ED told her he hasn't had his metronidazole infusion yet as they ran out of time. Other staff are on the way to theatre and will meet you there – you will be scrubbing for this case.

Briefing sheet for anaesthetist

It is 0800h on a weekend. You have just turned up at work. The theatre coordinator has sent for a patient, Brian Richards for an urgent laparotomy for suspected perforated viscus. She mentions that he is quite unwell. As you were on your way to theatre, you read on the booking form as follows:

BOOKING SLIP

IHD, hypertension, mild asthma

Patient taking gingko and garlic*, inhalers, aspirin, diltiazem.*

** these can affect coagulation*

Other staff are on the way to theatre and will meet you there.

Briefing sheet for anaesthetic technician

It is 0800h on a weekend. You have just turned up at work. The theatre coordinator has sent for a patient, Brian Richards, for an urgent laparotomy for perforated viscus. She mentions that the patient is quite unwell, someone from ICU rang to say they don't have a bed for this patient. The machine has had a level 2 check by your colleague but he had to leave. The rest of the staff are on the way to theatre and will meet you there.

7.4. Appendix 4. On-screen instructions for performing the Momento task sort

CARD SORTING EXERCISE

In the following exercise, you will be sorting actions relevant for a successful completion of **an urgent laparotomy for suspected perforated viscus [management of an abdominal stab patient] once the patient is in the OR**, in the order in which *you* believe they should be performed and by which team (i.e., anaesthesia, surgical, nursing) *you* believe they are *most likely* to be performed.

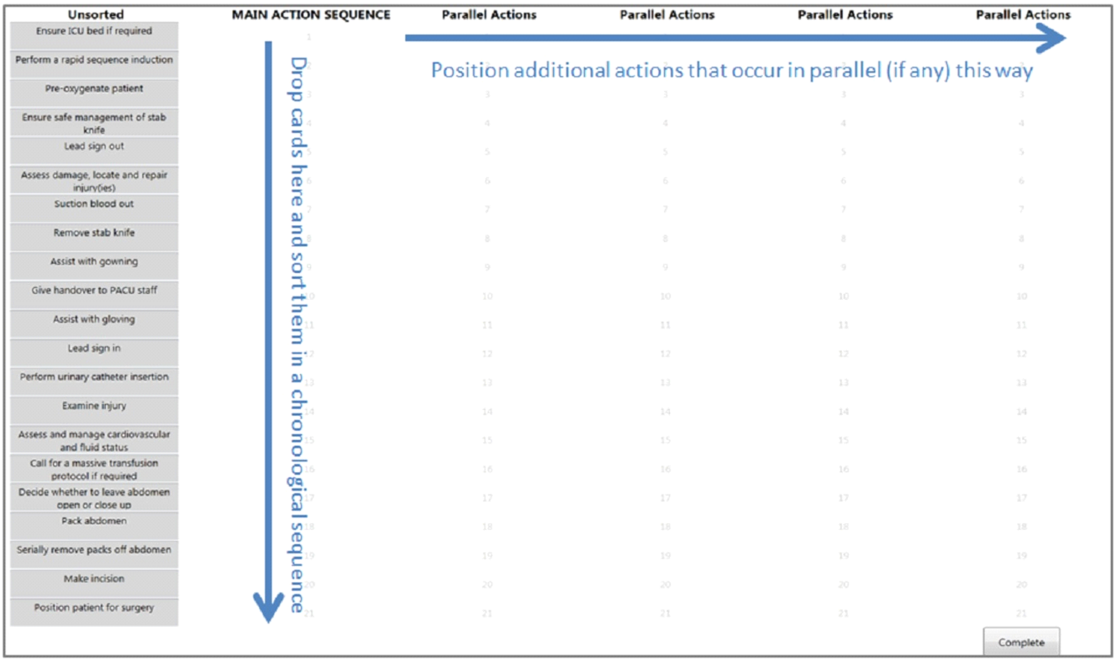
NOTE: The actions depicted on the cards are only a subset of steps that may be characteristic of the procedure and the list is not intended to be exhaustive.

Please read the following instructions carefully.

Instructions for sorting cards in a chronological sequence:

Drag and drop cards representing actions to be completed during the above procedure from the column labelled 'Unsorted' on the left of the screen to the column labelled 'Main Action Sequence' (see the image below) . Sort the cards in a chronological order from top to bottom, *down* the 'Main Action Sequence' column in which you believe actions represented on cards should be performed throughout the procedure (*regardless* of who you think should perform them) . Once positioned, individual cards can be rearranged within the sequence by dragging and dropping.

If you believe additional actions should be performed at the same time as the chosen action, drag the card representing the parallel action across to the first column labelled 'Parallel Actions' and drop it adjacent to the chosen action in the 'Main Action Sequence' column.



You can drag and drop up to 4 cards *adjacent* to the 'Main Action Sequence' column to depict actions that you believe should occur in parallel to any step in the sequence. The order in which you sort parallel actions is not important.

NOTE: You do not have to sort actions in parallel if you believe they can all be performed in a sequence. In that case, simply order them chronologically down the 'Main Action Sequence' column.

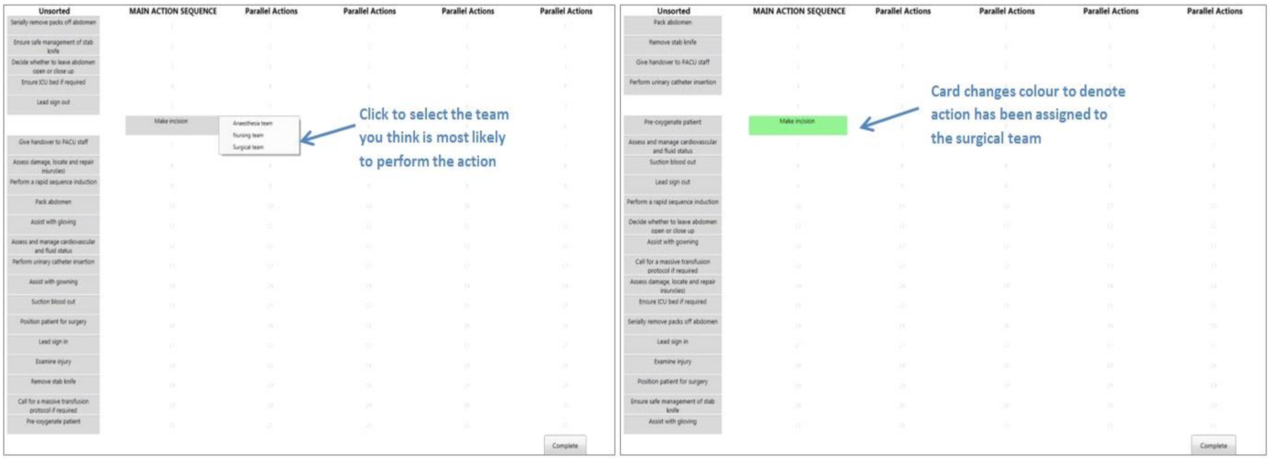
Each individual card can be placed in only one step of the sequence. The number of cards (i.e., actions) you place in the sequence (i.e., down the 'Main Action Sequence' column) is entirely up to you.

If *you believe* an action is <u>not required</u> for the management of *this particular patient*, leave the card depicting that action in the 'Unsorted' column. In this case, at the end of the task you will be prompted to confirm that the action(s) on card(s) left in the 'Unsorted' column are not required, as opposed to overlooked.

Instructions for assigning actions to teams:

As soon as you drag and drop a card from the 'Unsorted' column, a window will pop up next to the card asking you to select the team (i.e., surgical, nursing, or anaesthesia) <u>you believe is most likely</u> to perform the action depicted on the card. Point to the appropriate team on the list and click to make a selection. You can choose to assign an action to a team or change your selection of a team at any point in time by right-clicking on an individual card. You will know that an action has been assigned a team when the card depicting the action changes colour (as in the image below) . Each of the 3 teams is denoted by a different colour:

Surgical = GREEN Anaesthesia = RED Nursing = BLUE



NOTE: Assign an action to a team even if you feel it is most likely to be performed by only one of the team members (e.g., only the scrub nurse for nursing team, an anaesthetic technician for the anaesthesia team, etc.) .

At the end of the task, you should have sorted cards showing actions you believe are required for the management of this case from the 'Unsorted' column down the 'Main Action Sequence' column and should you choose, across in the 'Parallel Actions' column(s) . Cards depicting actions you believe are *not* required for the management of this particular patient should remain in the 'Unsorted' column. You should have also assigned a team to each sorted action to denote who you think is most likely to perform it during the procedure.

When finished, click the COMPLETE button at the bottom right of the screen.

8. REFERENCES

1. Lingard L, Espin S, Whyte S, Regehr G, Baker GR, Reznick R, et al. Communication failures in the operating room: an observational classification of recurrent types and effects. Qual Saf Health Care. 2004:13(5):330.

2. Helmreich RL, Schaefer HG. Team performance in the operating room. In: Bogner MS, editor. Human error in medicine. Hillsdale, NJ: Erlbaum; 1995. p. 225-53.

3. Manser T, Howard SK, Gaba DM. Identifying characteristics of effective teamwork in complex medical work environments: adaptive crew coordination in anaesthesia. In: Flin RH, Mitchell L, editors. Safer surgery: analysing behaviour in the operating theatre. Aldershot, UK: Ashgate; 2009. p. 223–39.

4. Rosen MA, Salas E, Wilson KA, King HB, Salisbury M, Augenstein JS, et al. Measuring team performance in simulation-based training: adopting best practices for healthcare. Simul in Healthc. 2008;3(1):33-41.

5. Sexton JB, Thomas EJ, Helmreich RL. Error, stress, and teamwork in medicine and aviation: cross sectional surveys. BMJ. 2000;320(7237):745-9.

6. Wilson RM, Runciman WB, Gibberd RW, Harrison BT, Newby L, Hamilton JD. The Quality in Australian health care study. Med J Aust. 1996;163:458-71.

7. Davis P, Lay-yee R, Briant R, Ali W, Scott A, Schug S. Adverse events in New Zealand public hospitals I: occurrence and impact. NZ Med J. 2002;115(1167):U271.

8. Davis P, Lay-yee R, Briant R, Schug S, Scott A, Johnson S, et al. Adverse events in New Zealand public hospitals: principal findings from a national survey. Wellington, NZ: Ministry of Health; 2001.

9. Baker G. The Canadian adverse events study. Can J Anesth. 2004;51(1):R5-R.

10. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. BMJ. 2001;322(7285):517.

11. Brennan TA, Leape LL, Laird NM, Hebert L, Localio AR, Lawthers AG, et al. Incidence of adverse events and negligence in hospitalized patients : results of the Harvard Medical Practice Study I. N Engl J Med. 1991;324(6):370-6.

12. Wilson RM, Michel P, Olsen S, Gibberd RW, Vincent C, El-Assady R, et al. Patient safety in developing countries: retrospective estimation of scale and nature of harm to patients in hospitals. BMJ. 2012;344(7850):20.

13. Arbous MS, Grobbee DE, van Kleef JW, de Lange JJ, Spoormans HH, Touw P, et al. Mortality associated with anaesthesia: a qualitative analysis to identify risk factors. Anaesthesia. 2001;56(12):1141.

14. Sutcliffe KM, Lewton E, Rosenthal ME. Communication failures: an insidious contributor to medical mishaps. Acad Med. 2004;79:186-94.

15. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An evaluation of outcome from intensive care in major medical centers. Ann Intern Med. 1986;104(3):410-8.

16. Gittell HJ, Fairfield MK, Bierbaum MB, Head MW, Jackson MR, Kelly MM, et al. Impact of relational coordination on quality of care, postoperative pain and functioning, and length of stay: a nine-hospital study of surgical patients. Med Care. 2000;38(8):807-19.

17. Christian CK, Gustafson ML, Roth EM, Sheridan TB, Gandhi TK, Dwyer K, et al. A prospective study of patient safety in the operating room. Surgery. 2006;139(2):159-73.

18. de Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. Qual Saf Health Care. 2008;17(3):216.

19. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The nature of adverse events in hospitalized patients : results of the Harvard Medical Practice Study II. N Engl J Med. 1991;324(6):377-84.

20. Mazzocco K, Petitti DB, Fong KT, Bonacum D, Brookey J, Graham S, et al. Surgical team behaviors and patient outcomes. Am J Surg. 2009;197(5):678-85.

21. Williams AL, Lasky RE, Dannemiller JL, Andrei AM, Thomas EJ. Teamwork behaviours and errors during neonatal resuscitation. Qual Saf Health Care. 2010;19(1):60.

22. Makary MA, Sexton JB, Freischlag JA, Holzmueller CG, Millman EA, Rowen L, et al. Operating room teamwork among physicians and nurses: teamwork in the eye of the beholder. J Am Coll Surgeons. 2006;202(5):746-52.

23. 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(8):1153-78.

24. Heinemann GD. Teams in health care settings. In: Heinemann GD, Zeiss AM, editors. Team performance in health care: assessment and development. New York: Kluwer Academic; 2002. p. 3-17.

25. Hall P. Interprofessional teamwork: professional cultures as barriers. J Interprof Care. 2005;19 Suppl 1:188-96.

26. Petrie HG. Do you see what I see? The epistemology of interdisciplinary inquiry. Educ Res. 1976;5(2):9-14.

27. Johnson-Laird PN. A computational analysis of consciousness. Cognition Brain Theor. 1983;6:499-508.

28. Cannon-Bowers JA, Salas E. Cognitive psychology and team training: shared mental models in complex systems. The Annual meeting of the Society for Industrial and Organizational Psychology; Miami, Florida 1990.

29. Salas E, Sims DE, Burke CS. Is there a "Big Five" in teamwork? Small Gr Res. 2005;36(5):555-99.

30. Cannon-Bowers JA, Salas E, Converse SA. Shared mental models in expert team decision making. In: Castellan NJ, editor. Individual and group decision making: current issues. Hillsdale, NJ, England: Lawrence Erlbaum Associates; 1993. p. 221–46.

31. Klimoski R, Mohammed S. Team mental model: construct or metaphor? J Manage. 1994;20(2):403-37.

32. Mohammed S, Ferzandi L, Hamilton K. Metaphor no more: a 15-year review of the team mental model construct. J Manage. 2010;36(4):876.

33. Reimer T, Park ES, Hinsz VB. Shared and coordinated cognition in competitive and dynamic task environments: an information-processing perspective for team sports. Int J Sport Exe Psychol. 2006;4:376-400.

34. Rasker PC, Post WM, Schraagen JMC. Effects of two types of intra-team feedback on developing a shared mental model in Command & Control teams. Ergonomics. 2000;43(8):1167-89.

35. Eccles DW, Tenenbaum G. Why an expert team is more than a team of experts: a social-cognitive conceptualization of team coordination and communication in sport. J Sport Exe Psychol. 2004;26(4):542-60.

36. Leach LS, Myrtle RC, Weaver FA. Surgical teams: role perspectives and role dynamics in the operating room. Health Serv Manage Res. 2011;24(2):81-90.

37. Mathieu JE, Hefner TS, Goodwin GF, Cannon-Bowers JA, Salas E. Scaling the quality of teammates' mental models: equifinality and normative comparisons. J Org Beh. 2005;26(1):37-56.

38. Stout RJ, Cannon-Bowers JA, Salas E, Milanovich DM. Planning, shared mental models, and coordinated performance: an empirical link is established. Hum Factors. 1999;41(1):61-71.

39. DeChurch LA, Mesmer-Magnus JR. Measuring shared team mental models: a meta-analysis. Group Dyn: Theor Res. 2010;14(1):1-14.

40. Marks MA, Zaccaro SJ, Mathieu JE. Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. J Appl Psychol. 2000;85(6):971-86.

41. Burtscher MJ, Manser T. Team mental models and their potential to improve teamwork and safety: a review and implications for future research in healthcare. Safety Sci. 2012;50(5):1344-54.

42. Salas E, Fowlkes JE, Stout RJ, Milanovich DM, Prince C. Does CRM training improve teamwork skills in the cockpit? Two evaluation studies. Hum Factors. 1999;41(2):326.

43. Swain K, Mills V. Implicit communication in novice and expert teams. Australia: DSTO Systems Sciences Laboratory; 2003.

44. Serfaty DS, Entin EE, Deckert JC. Implicit coordination in command teams. In: Levis AH, Levis IS, editors. Science of command and control: Part III coping with change: Aip Information Systems; 1994. p. 87-94.

45. Entin EE, Serfaty D. Adaptive team coordination. Hum Factors. 1999;41(2):312.

46. Shah J, Breazeal C. An empirical analysis of team coordination behaviors and action planning with application to human–robot teaming. Hum Factors. 2010;52(2):234-45.

47. 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(2):273.

48. Burtscher MJ, Manser T, Kolbe M, Grote G, Grande B, Spahn DR, et al. Adaptation in anaesthesia team coordination in response to a simulated critical event and its relationship to clinical performance. Brit J Anaesth. 2011;106(6):801-6.

49. Manser T, Harrison TK, Gaba DM, Howard SK. Coordination patterns related to high clinical performance in a simulated anesthetic crisis. Anesth Analg. 2009;108(5):1606-15.

50. Xiao Y, Hunter WA, Mackenzie CF, Jefferies NJ, Horst RL. Task complexity in emergency medical care and its implications for team coordination. LOTAS Group. Level One Trauma Anesthesia Simulation. Hum Factors. 1996;38(4):636-45.

51. Undre S, Sevdalis N, Healey AN, Darzi SA, Vincent CA. Teamwork in the operating theatre: cohesion or confusion? J Eval Clin Pract. 2006;12(2):182-9.

52. Langan-Fox J, Code S, Langfield-Smith K. Team mental models: techniques, methods, and analytic approaches. Hum Factors. 2000;42(2):242-71.

53. Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. Qual Saf Health Care. 2009;18(2):104.

54. Ellis APJ. System breakdown: the role of mental models and transactive memory in the relationship between acute stress and team performance. Acad Manage J. 2006;49(3):576.

55. Novak JD. Concept maps and Venn diagrams: two metacognitive tools to facilitate meaningful learning. Instr Sci. 1990;19:29-52.

56. 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(3):257-69.

57. Salas E, Dickinson TL, Converse S, Tannenbaum SI. Toward an understanding of team performance and training In: Swezey RW, Salas E, editors. Teams: their training and performance. Norwood: Ablex; 1992. p. 3-29.

58. Paris CR, Salas E, Cannon-Bowers JA. Teamwork in multi- person systems: a review and analysis. Ergonomics. 2000;43(8):1052-75.

59. Cannon-Bowers JA, Salas E. Team performance and training in complex environments: recent findings from applied research. Curr Dir Psychol Sci. 1998;7(3):83-7.

60. Baker DP, Day R, Salas E. Teamwork as an essential component of high- reliability organizations. Health Serv Res. 2006;41(4):1576-98.

61. Rousseau V, Aubé C, Savoie A. Teamwork behaviors: a review and an integration of frameworks. Small Gr Res. 2006;37(5):540.

62. Healey AN, Undre S, Sevdalis N, Koutantji M, Vincent CA. The complexity of measuring interprofessional teamwork in the operating theatre. J Interprof Care. 2006;20(5):485.

63. Reader TW, Flin R, Mearns K, Cuthbertson BH. Developing a team performance framework for the intensive care unit. Crit Care Med. 2009;37(5):1787-93.

64. Ilgen DR, Hollenbeck JR, Johnson M, Jundt D. Teams in organizations: from input-process-output models to IMOI models. Ann Rev Psychol. 2005;56:517.

65. Orasanu J. Shared mental models and crew decision making. CLS Report no. 46. Princeton, NJ: Princeton University Cognitive Science Laboratory, 1990.

66. Waller MJ, Gupta N, Giambatista RC. Effects of adaptive behaviors and shared mental models on control crew performance. Manage Sci. 2004;50(11):1534-44.

67. Sundstrom E, McIntyre M, Halfhill T, Richards H, Forsyth DR. Work groups: from the hawthorne studies to work teams of the 1990s and beyond. Group Dyn: Theor Res. 2000;4(1):44-67.

68. Hollenbeck JR, Derue DS, Guzzo R, Edwards JE. Bridging the gap between I/ O research and HR practice: improving team composition, team training, and team task design. Hum Resour Manage. 2004;43(4):353-66.

69. Helmreich RL. On error management: lessons from aviation. BMJ. 2000;320(7237):781-5.

70. Grote G, Zala-Mezö E, Grommes P. The effects of different forms of coordination in coping with work load: cockpit versus operating theatre. Ashgate, Aldershot, UK: GIHRE-Kolleg (Group Interaction in High-Risk Environments) of the Daimler-Benz-Foundation; 2004.

71. Lingard L, Regehr G, Espin S, Devito I, Whyte S, Buller D, et al. Perceptions of operating room tension across professions: building generalizable evidence and educational resources. Acad Med. 2005;80(10):S75-S9.

72. Lingard L, R. Forming professional identities on the health care team: discursive constructions of the ‘other’ in the operating room. Med Educ. 2002;36(8):728-34.

73. Sexton JB. Teamwork in the operating room: frontline perspectives among hospitals and operating room personnel. Anesthesiol. 2006;105(5):877.

74. Miller PA. Nurse-physician collaboration in an intensive care unit. Am J Crit Care. 2001;10(5):341.

75. Hojat M, Gonnella JS, Nasca TJ, Fields SK, Cicchetti A, Scalzo AL, et al. Comparisons of American, Israeli, Italian and Mexican physicians and nurses on the total and factor scores of the Jefferson scale of attitudes toward physician–nurse collaborative relationships. International J Nurs Stud. 2003;40(4):427-35.

76. Thomas EJ, Sexton JB, Helmreich RL. Discrepant attitudes about teamwork among critical care nurses and physicians. Crit Care Med. 2003;31(3):956.

77. Fleming M, Smith S, Slaunwhite J, Sullivan J. Investigating interpersonal competencies of cardiac surgery teams. Can J Surg. 2006;49(1):22-31.

78. Flin R, Yule S, McKenzie L, Paterson-Brown S, Maran N. Attitudes to teamwork and safety in the operating theatre. Surg. 2006;4(3):145-51.

79. Ummenhofer W, Amsler F, Sutter PM, Martina B, Martin J, Scheidegger D. Team performance in the emergency room: assessment of inter-disciplinary attitudes. Resusc. 2001;49(1):39-46.

80. Horsburgh M, Perkins R, Coyle B, Degeling P. The professional subcultures of students entering medicine, nursing and pharmacy programmes. J Interprof Care. 2006;20(4):425.

81. Mills P, Neily J, Dunn E. Teamwork and communication in surgical teams: implications for patient safety. J Am Coll Surg. 2008;206(1):107-12.

82. Wheelan SA, Burchill CN, Tilin F. The link between teamwork and patients' outcomes in intensive care units. Am J Crit Care. 2003;12(6):527.

83. Manser T. Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. Acta Anaesth Scand. 2009;53(2):143-51.

84. Manser T. Team performance assessment in healthcare: facing the challenge. Sim Healthc. 2008;3(1):1-3.

85. Bartlett FC. Remembering : a study in experimental and social psychology. Cambridge; New York: Cambridge University Press; 1995.

86. Schank RC, Abelson RP. Scripts, plans, goals, and understanding : an inquiry into human knowledge structures. Hillsdale, NJ; New York: Halsted Press Division of John Wiley and Sons; 1977.

87. Minsky M. A framework for representing knowledge. In: Winston PH, editor. The psychology of computer vision. New York: McGraw-Hill; 1975. p. 211-77.

88. Fillmore C, editor An alternative to Checklist Theories of Meaning. Proceedings of the 1st Annual Meeting of the Berkeley Linguistics Society; 1975.

89. Holland DC, Quinn N. Cultural models in language and thought. Cambridge; New York: Cambridge University Press; 1987.

90. D'Andrade RG, Strauss C. Human motives and cultural models. Cambridge; New York: Cambridge University Press; 1992.

91. Johnson - Laird PN, Byrne RMJ, Schaeken W. Why models rather than rules give a better account of propositional reasoning: a reply to Bonatti and to O'Brien, Braine, and Yang. Psychol Rev. 1994;101(4):734.

92. Johnson-Laird PN. Mental models in cognitive science. Cog Sci. 1980;4(1):71-115.

93. Held C, Knauff M, Vosgerau G. Mental models and the mind: current developments in cognitive psychology, neuroscience, and philosophy of mind. Amsterdam; Boston: Elsevier; 2006.

94. Newman WM, Lamming MG. Interactive system design. Reading, MA: Addison-Wesley; 1995.

95. Norman DA. Cognitive engineering. In: Norman DA, Draper W, editors. User centered systems design. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986. p. 31-61.

96. Rouse WB, Cannon - Bowers JA, Salas E, Cannon-Bowers JA. The role of mental models in team performance in complex systems. IEEE T Syst Man Cyb. 1992;22(6):1296-308.

97. Kieras DE, Bovair S. The role of a mental model in learning to operate a device. Cog Sci. 1984;8(3):255-73.

98. Bourbousson J, P. Description of dynamic shared knowledge: an exploratory study during a competitive team sports interaction. Ergonomics. 2011;54(2):120-38.

99. Mohammed S, Dumville BC. Team mental models in a team knowledge framework: expanding theory and measurement across disciplinary boundaries. J Org Beh. 2001;22:89-106.

100. Rouse WB, Morris NM. On looking into the black box. Psychol Bull. 1986;100(3):349-63.

101. Yen J, Fan X, Sun S, Wang R, Chen C, Kamali K, et al. Implementing shared mental models for collaborative teamwork. Workshop on Collaboration Agents: autonomous agents for collaborative environments The IEEE/WIC Intelligent Agent Technology Conference; 2003; Halifax, Canada. p. 115-26.

102. Webster CS. Why anaesthetising a patient is more prone to failure than flying a plane. Anaesthesia. 2002;57(8):818-38.

103. Cooke NJ, Salas E, Cannon-Bowers JA, Stout RJ. Measuring team knowledge. Human Factors. 2000;42(1):151-73.

104. Johnson T, Lee Y, Lee M, O'Connor D, Khalil M, Huang X. Measuring sharedness of team-related knowledge: design and validation of a shared mental model instrument. Hum Res Develop Intern. 2007;10(4):437-54.

105. Edwards BD, Day EA, Arthur W, Bell ST, Zedeck S. Relationships among team ability composition, team mental models, and team performance. J Appl Psych. 2006;91(3):727-36.

106. Serfaty DS, Entin EE, Volpe C, editors. Adaptation to stress in team decision making and coordination. Proceedings of the 37th Human Factors and Ergonomics Society Annual Meeting; 1993 October.

107. Calland J, Guerlain S, Adams R, Tribble C, Foley E, Chekan E. A systems approach to surgical safety. Surg Endosc. 2002;16(6):1005-14.

108. Wittenbaum GM, Stasser G, Merry CJ. Tacit coordination in anticipation of small group task completion. J Exp Soc Psychol. 1996;32(2):129-52.

109. Espinosa A, Lerch FJ, Kraut RE. Explicit vs. implicit coordination mechanisms and task dependencies: one size does not fit all. In: Salas E, Fiore SM, editors. Team cognition: understanding the factors that drive process and performance. Washington, DC: American Psychological Association; 2004. p. 107-29.

110. Kolbe M, Burtscher MJ, Manser T, Künzle B, Grote G. The role of coordination in preventing harm in healthcare groups: research examples from anesthesia and an integrated model of coordination for action teams in healthcare. In: Boos M, Kolbe M, Kappeler P, Ellwart T, editors. Coordination in human and primate groups. Heidelberg, Germany: Springer; 2011. p. 75-92.

111. Wittenbaum GM, Vaughan SI, Stasser G. Coordination in task-performing groups. In: Tindale RS, Heath L, Edwards J, Posavac EJ, Bryant FB, Suarez-Balcazar Y, et al., editors. Theory and research on small groups. New York: Plenum; 1998. p. 177–204.

112. Serfaty DS, Entin EE, Johnston JH. Team coordination training. In: Cannon-Bowers J, Salas E, editors. Making decisions under stress: Implications for training and simulation. Washington, DC: American Psychological Association; 1998. p. 221-45.

113. Kolbe M, Künzle B, Zala-Mezö E, Wacker J, Grote G. Measuring coordination behaviour in anaesthesia teams during induction of general anaesthetics. In: Flin RH, Mitchell L, editors. Safer surgery: analysing behaviour in the operating theatre. England ; Burlington, VT: Ashgate; 2009. p. 203-21.

114. Burtscher MJ, Wacker J, Grote G, Manser T. Managing non-routine events in anesthesia: the role of adaptive coordination. Hum Factors. 2010;52(2):282-94.

115. Rosen MA, Dietz AS, Yang T, Priebe CE, Pronovost PJ. An integrative framework for sensor-based measurement of teamwork in healthcare. JAMIA. 2015;22(1):11-8.

116. Healey AN, Undre S, Vincent CA. Developing observational measures of performance in surgical teams. Qual Saf Health Care. 2004;13(Suppl. 1):i33.

117. Rentsch JR, Heffner TS, Duffy LT. What you know is what you get from experience: team experience related to teamwork schemas. Group Organ Manage. 1994;19(4):450-74.

118. Beckman CM. The influence of founding team company affiliations on firm behavior. Acad Manage J. 2006;49(4):741-58.

119. Rico R, Sanchez-Manzanares M, Gil F, Gibson C. Team implicit coordination processes: a team knowledge-based approach. Acad Manage Rev. 2008;33(1):163-84.

120. Blickensderfer E, Reynolds R, Salas E, Cannon-Bowers J. Shared expectations and implicit coordination in tennis doubles teams. J Appl Sport Psychol. 2010;22(4):486-99.

121. Ellis APJ, Bell BS. Capacity, collaboration, and commonality. A framework for uderstanding team learning. In: Neider LL, Schriesheim C, editors. Understanding teams. USA: IAP; 2005. p. 1-25.

122. Eby LT, Meade AW, Parisi AG, Douthitt SS. The development of an individual-level teamwork expectations measure and the application of a within-group agreement statistic to assess shared expectations for teamwork. Organ Res Method. 1999;2(4):366-94.

123. DeChurch LA, Mesmer-Magnus JR. The cognitive underpinnings of effective teamwork. J Appl Psychol. 2010;95(1):32-53.

124. Lim Bc, Klein KJ. Team mental models and team performance: a field study of the effects of team mental model similarity and accuracy. J Organ Behav. 2006;27(4):403-18.

125. Smith-Jentsch K, Zedeck S. Investigating linear and interactive effects of shared mental models on safety and efficiency in a field setting. J Appl Psychol. 2005;90(3):523-35.

126. Marks MA, Sabella MJ, Burke CS, Zaccaro SJ. The impact of cross-training on team effectiveness. J Appl Psychol. 2002;87(1):3-13.

127. Mathieu J, Rapp T, Maynard MT, Mangos P. Interactive effects of team and task shared mental models as related to air traffic controllers' collective efficacy and effectiveness. Hum Perform. 2009;23(1):22-40.

128. Gurtner A, Tschan F, Semmer NK, Nagele C. Getting groups to develop good strategies: effects of reflexivity interventions on team process, team performance, and shared mental models. Organ Behav Hum Dec. 2007;102(2):127-42.

129. Miles JR, Kivlighan DM, Parks CD. Team cognition in group interventions: the relation between coleaders' shared mental models and group climate. Group Dyn: Theor Res. 2008;12(3):191-209.

130. Rentsch JR, Klimoski RJ. Why do ‘great minds’ think alike?: antecedents of team member schema agreement. J Organ Behav. 2001;22(2):107-20.

131. Langan-Fox J. Skill acquisition and the development of the team mental model: an integrative approach to analyzing organizational teams. In: West M, Tjosvold D, Smith KG, editors. International handbook of organizational teamwork and co-operative working. London: Wiley; 2003. p. 321-60.

132. Kellermanns FW, Floyd SW, Pearson AW, Spencer B. The contingent effect of constructive confrontation on the relationship between shared mental models and decision quality. J Organ Behav. 2008;29(1):119-37.

133. Smith-Jentsch KA, Campbell GE, Milanovich DM, Reynolds AM, Cannon-bowers JA, Salas E. Measuring teamwork mental models to support training needs assessment, development, and evaluation: two empirical studies. J Organ Behav. 2001;22(2):179-94.

134. Cooke NJ, Kiekel PA, Salas E, Bowers C, Stout R, Cannon-Bowers J. Measuring team knowledge. Group Dyn: Theor Res. 2003;7(3):179-99.

135. Zheng Y. Unlocking founding team prior shared experience: a transactive memory system perspective. J Bus Venturing. 2012;27(5):577-91.

136. Espinosa JA, Slaughter SA, Kraut RE, Herbsleb JD. Familiarity, complexity, and team performance in geographically distributed software development. Organ Sci. 2007;18(4):613-30.

137. Shah P, Jehn K. Do friends perform better than acquaintances? The interaction of friendship, conflict, and task. Group Decis Negot. 1993;2(2):149-65.

138. Harrison DA, Mohammed S, McGrath JE, Florey AT, Vanderstoep SW. Time matters in team performance: effects of member familiarity, entrainment, and task discontinuity on speed and quality. Pers Psychol. 2003;56(3):633-69.

139. Resick CJ. An investigation of the antecedents and consequences of shared mental models in teams. Dissertation Abstracts International: Section B: The Sciences and Engineering. 2005;65(10-B):5447.

140. Kor YY. Experience-based top management team competence and sustained growth. Organ Sci. 2003;14(6):707-19.

141. Zenger TR, Lawrence BS. Organizational demography: the differential effects of age and tenure distributions on technical communication. Acad Manage J. 1989;32(2):353-76.

142. Gruenfeld DH, Mannix EA, Williams KY, Neale MA. Group composition and decision making: how member familiarity and information distribution affect process and performance. Organ Behav Hum Dec. 1996;67(1):1-15.

143. Wildman JL, Salas E, Scott CPR. Measuring cognition in teams. Hum Factors. 2014;56(5):911-41.

144. Rugg G, McGeorge P. The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. Expert Syst. 2005;22(3):94-107.

145. Mohammed S, Klimoski R, Rentsch JR. The measurement of team mental models: we have no shared schema. Organ Res Methods. 2000;3(2):123-65.

146. Johnson TE, O'Connor DL, Yoon SW. Measuring team shared understanding using the analysis-constructed shared mental model methodology. PIQ. 2008;21(3):113-34.

147. Schvaneveldt RW, Durso FT, Dearholt DW. Network structures in proximity data. Psychol Learn Motiv. 1989;24:249-84.

148. Gaba DM, Howard SK, Fish KJ, Smith BE, Sowb YA. Simulation-based training in anesthesia crisis resource management (ACRM): a decade of experience. Simulat Gaming. 2001;32(2):175-93.

149. Manser T, Wehner T, Rall M. Analysing action sequences in anaesthesia. Eur J Anaesth. 2000;17(8):526.

150. Guise JM, Deering SH, Kanki BG, Osterweil P, Li H, Mori M, et al. Validation of a tool to measure and promote clinical teamwork. Simul Healthc. 2008;3(4):217-23.

151. Webber SS, Chen G, Payne SC, Marsh SM, Zaccaro SJ. Enhancing team mental model measurement with performance appraisal practices. Organ Res Methods. 2000;3(4):307-22.

152. Kozlowski SWJ, Hattrup K. A disagreement about within-group agreement: disentangling issues of consistency versus consensus. J Appl Psychol. 1992;77(2):161-7.

153. Kane MT. Current concerns in validity theory. J Educ Meas. 2001;38(4):319-42.

154. Downing SM. Validity: on the meaningful interpretation of assessment data. Med Educ. 2003;37(9):830-7.

155. Cronbach LJ. Test validation. In: Thorndike RL, editor. Educational measurement. 2nd ed. Washington, DC: American Council of Education; 1971. p. 443-507.

156. Schwab DP. Construct validity in organization behavior. In: Staw BM, Cummings LL, editors. Research in organizational behavior. Greenwich, CT: JAI Press; 1980. p. 3-43.

157. Carmines EG, Zeller RA. Reliability and validity assessment. Beverly Hills, CA: Sage Publications; 1979.

158. Nunnally JC, Bernstein IH. Psychometric theory. New York: McGraw-Hill; 1994.

159. Trochim WMK, Donnelly JP. The research methods knowledge base: measurement validity types: Atomic Dog/Cengage Learning; 2006 [cited 2012 20 October]. Available from: http://www.socialresearchmethods.net/kb/constval.php.

160. Campbell GE. Assessing structural knowledge In: Klimoski R, editor. Knowledge structures and training: measurement and validity issues. The11th annual meeting of the Society for Industrial and Organizational Psychology; San Diego; 1996.

161. Webster CS, Larsson L, Frampton CM, Weller J, McKenzie A, Cumin D, et al. Clinical assessment of a new anaesthetic drug administration system: a prospective, controlled, longitudinal incident monitoring study. Anaesthesia. 2010;65(5):490-9.

162. Weller JM, Torrie J, Boyd M, Frengley R, Garden A, Ng WL, Frampton CM, Myles PS. Improving team information sharing with a structured call- out in anaesthetic emergencies: a randomized controlled trial. Brit J Anaesth. 2014;112(6):1042-9.

163. Weller JM, Merry AF, Robinson BJ, Warman GR, Janssen A. The impact of trained assistance on error rates in anaesthesia: a simulation-based randomised controlled trial. Anaesthesia. 2009;64(2):126-30.

164. Merry AF, Webster CS, Mathew DJ. A new, safety-oriented, integrated drug administration and automated anesthesia record system. Anesth Analges. 2001;93:385-90.

165. Merry AF, Weller JM, Robinson BJ, Warman GR, Davies E, Shaw J, et al. A simulation design for research evaluating safety innovations in anaesthesia. Anaesthesia. 2008;63(12):1349-57.

166. Weller J, Frengley R, Torrie J, Shulruf B, Jolly B, Hopley L, et al. Evaluation of an instrument to measure teamwork in multidisciplinary critical care teams. BMJ Qual Saf. 2011;20(3):216-22.

167. Weller J, Cumin D, Torrie J, Boyd M, Civil I, Madell D, et al. Multidisciplinary operating room simulation-based team training to reduce treatment errors: a feasibility study in New Zealand hospitals. NZ Med J. 2015;128(1418):40-51.

168. Leach LS, Myrtle RC, Weaver FA, Dasu S. Assessing the performance of surgical teams. Health Care Manage Rev. 2009;34(1):29-41.

169. Yule S, Flin RH, Maran N, Rowley D, Youngson G, Duncan J, et al. Development and evaluation of the NOTSS behaviour rating system for intraoperative surgery (2003-2008). In: Flin RH, Mitchell L, editors. Safer surgery: analysing behaviour in the operating theatre. Farnham, England ; Burlington, VT: Ashgate; 2009. p. 7-25.

170. Müller MP. Six steps from head to hand: a simulator based transfer oriented psychological training to improve patient safety. Resuscitation. 2007;73(1):137.

171. Smith-Jentsch K. Measuring team-related cognition: The devil is in the details. In: Salas E, Goodwin GF, Burke CS, editors. The organizational frontiers series. New York, NY: Routledge/Taylor & Francis Group; 2009. p. 491-508.

172. Kumar R, Vassilvitskii S, editors. Generalized distances between rankings. Proceedings of the 19th International Conference on World Wide Web, WWW '10; 2010.

173. Schimke S, Vielhauer C. Similarity searching for on-line handwritten documents. J Multimod User Interfaces. 2007;1(2):49-54.

174. Yujian L. A normalized Levenshtein distance metric. IEEE T Pattern Anal. 2007;29(6):1091-5.

175. Kohonen T. Median strings. Pattern Recogn Letters. 1985;3(5):309-13.

176. Haldar R, Mukhopadhyay D, editors. Levenshtein distance technique in dictionary lookup methods: an improved approach. CoRR [Internet]. 2011 [cited 2014 March 12]; January. Available from: http://arxiv.org/abs/1101.1232

177. Deibel K, Anderson R, Anderson R. Using edit distance to analyze card sorts. Expert Syst. 2005;22(3):129-38.

178. Bhamidipati NL, Pal SK. Comparing scores intended for ranking. IEEE T Knowl Data En. 2009;21(1):21-35.

179. Dwork C, Kumar Z, Naor M, Sivakumar D, editors. Rank aggregation methods for the Web. Proceedings of the Tenth International World Wide Web Conference WWW '01; 2001 May 1-5; Hong Kong: ACM. p. 613-22.

180. Scott J. Social network analysis: a handbook. London: SAGE Publications; 2000.

181. Harrison DAKKJ. What's the difference? Diversity constructs as separation, variety, or disparity in organizations. Acad Manage Rev. 2007;32(4):1199-228.

182. Chou LF, W. Shared work values and team member effectiveness: the mediation of trustfulness and trustworthiness. Hum Relat. 2008;61(12):1713-42.

183. Krippendorff K. Content analysis: an introduction to its methodology. Los Angeles; London: SAGE; 2013.

184. Krippendorff K. Computing Krippendorff 's alpha-reliability. 2011 [cited 2013 June 22]. Available from: http://repository.upenn.edu/asc_papers/43.

185. Watkins M, Pacheco M. Interobserver agreement in behavioral research: importance and calculation. J Behav Educ. 2000;10(4):205-12.

186. Stemler SE. A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. Prac Ass Res Eval [Internet]. 2004 [cited 2013 July 22]; 9(4). Available from: http://PAREonline.net/getvn.asp?v=9&n=4

187. Uebersax JS. Raw agreement indices 2014 [cited 2014 August 13]. Available from: http://www.john-uebersax.com/stat/raw.htm.

188. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005;85(3):257.

189. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Measur. 1960;20(1):37-46.

190. Gwet KL. Handbook of inter-rater reliability : the definitive guide to measuring the extent of agreement among raters. Gaithersburg, MD: Advanced Analytics; 2012.

191. Warrens MJ. Inequalities between multi-rater kappas. Theor Meth Appl Data Sci. 2010;4(4):271-86.

192. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med. 2005;37(5):360.

193. Maxwell AE. Coefficients of agreement between observers and their interpretation. Brit J Psych. 1977;130:79.

194. Uebersax JS, Masters JC. Diversity of decision-making models and the measurement of interrater agreement. Psychol Bull. 1987;101(1):140-6.

195. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. Am J Epidemiol. 1987;126(2):161.

196. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971;76(5):378-82.

197. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Commun Meth Meas. 2007;1(1):77-89.

198. King JE. Software solutions for obtaining a kappa-type statistic for use with multiple raters. Paper presented at the annual meeting of the Southwest Educational Research Association; Dallas, TX: 2004.

199. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52(3):591-611.

200. Shapiro SS, Wilk MB, Chen HJ. A comparative study of various tests for normality. J Am Stat Ass. 1968;63(324):1343-72.

201. Kraiger K, Wenzel HL. Conceptual development and empirical evaluation of measures of shared mental models as indicators of team effectiveness. In: Brannick MT, Salas E, Prince C, editors. Team performance assessment and measurement: theory, methods, and applications: Psychology Press; 1997. p. 63-82.

202. Mitchell MM, Jolley JM. Research design explained. Belmont, CA: Wadsworth; 2010.

203. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum; 1988.

204. Borgstede JP, Lewis RS, Bhargavan M, Sunshine JH. RADPEER quality assurance program: a multifacility study of interpretive disagreement rates. J Am Coll Radiol. 2004;1(1):59.

205. Filippi CG, Schneider B, Burbank HN, Alsofrom GF, Linnell G, Ratkovits B. Discrepancy rates of radiology resident interpretations of on-call neuroradiology MR imaging studies. Radiology. 2008;249(3):972.

206. Babiarz LS, Yousem DM. Quality control in neuroradiology: discrepancies in image interpretation among academic neuroradiologists. Am J Neuroradiol. 2012;33(1):37.

207. Dolatabadi AA, Baratloo A, Rouhipour A, Abdalvand A, Hatamabadi H, Forouzanfar M, et al. Interpretation of computed tomography of the head: emergency physicians versus radiologists. Trauma Mon. 2013;18(2):86-9.

208. Nolan TM, Oberklaid F, Boldt D. Radiological services in a hospital emergency department - an evaluation of service delivery and radiograph interpretation. J Paed Child Heal. 1984;20(2):109-12.

209. Minnes BG, Sutcliffe T, Klassen TP. Agreement in the interpretation of extremity radiographs of injured children and adolescents. Acad Emerg Med. 1995;2(9):826-30.

210. Cao J, Zhang S. Multiple comparison procedures. JAMA. 2014;312(5):543-5.

211. Shaffer JP. Multiple hypothesis testing. Ann Rev Psychol. 1995;46(1):561-84.

212. Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. Stat Med. 1990;9(7):811-8.

213. Perneger TV. What's wrong with Bonferroni adjustments. BMJ. 1998;316(7139):1236-8.

214. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. J Clin Epidemiol. 2014;67(8):850-587.

215. Curran-Everett D. Multiple comparisons: philosophies and illustrations. American J Physiol. 2000;279(1):R1-8.

216. Bender R, Lange S. Adjusting for multiple testing-when and how? J Clin Epidemiol. 2001;54(4):343-9.

217. Sainani KL. The problem of multiple testing. PM&R. 2009;1(12):1098-103.

218. Light RJ, Pillemer DB. Summing up: the science of reviewing research. Cambridge, MA: Harvard University Press; 1984.

219. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology. 1990;1(1):43-6.

220. Paige JT, Aaron DL, Yang T, Howell DS, Hilton CW, Cohn I, et al. Implementation of a preoperative briefing protocol improves accuracy of teamwork assessment in the operating room. Am Surg. 2008;74(9):817.

221. Paige J, Aaron D, Yang T, Howell D, Chauvin S. Improved operating room teamwork via SAFETY Prep: a rural community hospital's experience. World J Surg. 2009;33(6):1181-7.

222. Lingard L, Regehr G, Orser B, Reznick R, Baker GR, Doran D, et al. Evaluation of a preoperative checklist and team briefing among surgeons, nurses, and anesthesiologists to reduce failures in communication. Arch Surg. 2008;143(1):12.

223. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AHS, Dellinger EP, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. New Engl J Med. 2009;360(5):491-9.

224. Fudickar A, Hörle K, Wiltfang J, Bein B. The effect of the WHO Surgical Safety Checklist on complication rate and communication. Dtsch Arztebl Int. 2012;109(42):695.

225. Steiner ID. Group process and productivity. New York: Academic Press; 1972.

226. Milliken F, Martins L. Searching for common threads: understanding the multiple effects of diversity in organizational groups. Acad Manage Rev. 1996;21(2):402-33.

227. Boos M. Optimal sharedness of mental models for effective group performance. CoDesign. 2007;3(1):21-8.

228. LeBreton JM, Senter JL. Answers to 20 questions about interrater reliability and interrater agreement. Organ Res Methods. 2008;11(4):815-52.

229. Cannon-Bowers JA, Tannenbaum SI, Salas E, Volpe CE. Defining competencies and establishing team training requirements. In: Guzzo R, Salas E, editors. Team effectiveness and decision making in organizations. San Francisco: Jossey-Bass; 1995. p. 333-80.

230. Bedwell WL, Ramsay P, Salas E. Helping fluid teams work: a research agenda for effective team adaptation in healthcare. Transl Behav Med. 2012;2(4):504-9.

231. Baker DP, Gustafson S, Beaubien JM, Salas E, Barach P. Medical team training programs in health care. Rockville, MD: Agency for Healthcare Research And Quality; 2005.

232. Schaefer HG, Helmreich RL. The importance of human factors in the operating room. Anesthesiol. 1994;80(2):479.

233. Cumin D, Boyd M, Gurusinghe N, Civil I, Guthrie W, Johnston S, et al. Creating a multidisciplinary operating room simulation course (MORSim) for training and research: overcoming barriers. Simul Healthc. 2014;9(1):73.

234. King HB, Battles J, Baker DP, Alonso A, Salas E, Webster J, et al. TeamSTEPPS™: team strategies and tools to enhance performance and patient safety. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. Advances in patient safety: new directions and alternative approaches (Vol 3: performance and tools). Rockville,MD: Agency for Healthcare Research and Quality; 2008. p. 5-20.

235. Thomas EJ, Mazzocco K, Graham S, Petitti D, Fong K, Bonacum D, et al. An empiric study of surgical team behaviours, patient outcomes, and a programme based on its results. In: Flin RH, Mitchell L, editors. Safer surgery: analysing behaviour in the operating theatre. England ; Burlington, VT: Ashgate; 2009. p. 261-82.

236. Sevdalis N, Healey AN, Vincent CA. Distracting communications in the operating theatre. J Eval Clin Prac. 2007;13(3):390-4.