DET PSYKOLOGISKE FAKULTET

Ψ

The impact of gaze communication on team functioning: A pilot study of eye-tracking in a simulator environment

HOVEDOPPGAVE

profesjonsstudiet i psykologi

Gry Dammen Tyssebotn

Vår 2014

Veileder:
Bjørn Sætrevik

Forord

Jeg vil rette en takk til Sturle Tvedt for samarbeidet under planlegging av studien og innhenting av data. Jeg vil også takke undervisningspersonell ved høyskolen for at de tok så godt imot oss under datainnsamlingen, og forsøkspersonene for at de ønsket å delta i studien. Jeg vil også takke Mark Price for at jeg fikk låne laben for å prosessere dataene og Ragnhild Holm for gjennomlesing av oppgaven.

En spesiell takk rettes til veileder Bjørn Sætrevik for støtte og innspill underveis i prosessen og for tidligere års samarbeid om andre prosjekter. Takk for at du hele veien har vært inkluderende og lyttende samt for engasjementet du har vist for eye-tracking. Jeg håper at denne pilotstudien vil komme til nytte i arbeidet med fremtidige studier som bruker eye-tracking som metode.

Bergen, 13.04.14

Abstract

This pilot study aimed to examine the feasibility of a future main experiment that will address the role of gaze communication in team functioning in a simulator environment. Feasibility was operationalized in terms of whether or not eight specified criteria were met, and resources needed to preprocess eye-tracking data were estimated. A matched-subjects between-subjects design is planned for the main study to ensure homogeneity between participants in experimental groups, while the pilot study is a within-subjects design. Students of maritime navigation were recruited to participate in the study during their scheduled simulator training. The eye movements of one of the participants in a two-person team were recorded during planning and navigation of a given task, and an expert rating of the team's performance was collected. Each team was recorded at two time-points with an intervening treatment (gaze communication training), and questionnaires measuring shared mental models and performance were distributed six times. The study showed that too few feasibility criteria were met for the main study to be considered feasible if conducted as originally planned, and that considerable resource use should be expected for preprocessing of eye-tracking data. Improvements to the design and sample size considerations are discussed based on the pilot study and literature review.

Keywords: pilot, simulator, experiment, eye-tracking, gaze, team performance, shared mental models.

Sammendrag

Denne pilotstudien søkte å undersøke gjennomførbarheten av et fremtidig hovedeksperiment som vil adressere hvilken betydning blikkommunikasjon kan ha for teamets fungering i et simulatormiljø. Gjennomførbarhet ble operasjonalisert som oppnåelse av åtte spesifiserte kriterier, og det ble estimert ressurser nødvendig for å preprosessere blikkdataene. For å sikre homogenitet mellom deltagerne i eksperimentgruppene i hovedstudien er det planlagt at et balansert mellomgruppedesign vil bli benyttet, mens denne pilotstudien benyttet et innengruppedesign. Nautikkstudenter ble rekruttert til å delta i en studie som ble gjennomført innenfor rammene av deres ordinære simulatortrening. Øyebevegelsene til en av deltakerne i et team bestående av to personer ble målt under planlegging og navigering av en gitt oppgave, og en ekspert vurderte lagets prestasjon. Hvert team ble målt før og etter en intervensjon (trening i blikkommunikasjon), og spørreskjema som måler delte mentale modeller og prestasjoner ble distribuert seks ganger. Resultatene viste at hovedstudien ikke er gjennomførbar som opprinnelig planlagt, og at preprosessering av blikkdataene vil kreve betydelige ressurser. Forslag til designforbedringer og betraktninger rundt utvalgsstørrelse er diskutert basert på pilotstudien og gjennomgang av litteratur.

Table of contents

List of tables and figures

Maritime navigation can be hazardous. During 2010, a total of 644 vessels were reported to have been involved in significant accidents in the European Union and neighboring waters (European Maritime Safety Agency, 2011). The vast majority (75-96%) of maritime accidents is estimated to be partly or fully accounted for by human errors (Rothblum, 2000). Among the most common human factors associated with such accidents are failures of situation awareness, decision-making, teamwork, and communication (see Hetherington, Flin, & Mearns, 2006, for a review). These aspects are typically emphasized in maritime training programs developed with intent to enhance the leadership and watch keeping skills of bridge personnel. The significance of communication processes is stressed in both training programs and in the literature, but there is a tendency to focus mainly on the verbal aspect of communication. It is possible that more attention should be directed toward nonverbal communication, such as gaze communication, in these areas. Gaze communication can be important to maritime safety as it is possible that observing what a team member is looking at might indicate the target of his or her attention. This information can have consequences for one's assessment of the situation and may constitute the basis for further choice of action. In turn, these actions (or lack of actions) can have consequences for the safety on board the vessel. Studies examining the relationship between gaze and safety should therefore be of interest to both developers of training programs and the maritime safety research community.

The relationship between gaze communication and maritime safety can be explored in a field experiment on a simulator bridge. This setting would allow for manipulation of the scenario in order to draw causal conclusions while ensuring ecological validity. Gaze patterns can be measured objectively by the use of eye-tracking equipment, and in a field study it would be important that such equipment is mobile and non-obtrusive. A suitable option is therefore to use eye-tracking glasses. However, few studies have applied eye-tracking glasses so far, as equipment of sufficient quality only recently became commercially available and the methodology requires considerable resource use. The objective of this pilot study is therefore to assess the feasibility of a future main study using eye-tracking in a simulator environment in an educational setting. The main study will explore (a) the impact of gaze on shared mental models (SMM) and team performance, and (b) whether or not gaze communication can be improved through team training. Thus, this pilot study will assess the feasibility of the main study in terms of participant flow through the study, data quality, and required resources.

Teams

A team may be defined as “two or more individuals with specified roles interacting adaptively, interdependently, and dynamically toward a common or valued goal” (Salas, Sims, & Burke, 2005, pp. 559, 562). Salas, Burke, and Cannon-Bowers (2000) have pointed out three of the most essential characteristics associated with teams. The first characteristic concerns that working toward a shared goal makes the team members interdependent of each other; a process that requires the team members to coordinate their actions (Salas et al., 2000). Coordination can be defined as “the process of orchestrating the sequence and timing of interdependent actions” (Marks, Mathieu, & Zaccaro, 2001, p. 367). The second characteristic is the necessity for team members to adapt to complex work settings to function efficiently. This process requires team members to have sufficient knowledge and skills to engage in dynamic exchanges of information and resources (Salas et al., 2000). The third and final characteristic concerns that team members must hold an attitude of working toward the team’s common goal as opposed to their individual goals (Salas et al., 2000). This is also one of the important aspects of the process of collaboration. Collaboration has been defined as “an evolving process whereby two or more social entities actively and reciprocally engage in joint activities aimed at achieving at least one shared goal” (Bedwell et al., 2012, p. 130).

Common to all three characteristics is their relationship to communication. Efficient communication has been proposed to be important to successfully accomplish coordination (MacMillan, Entin, & Serfaty, 2004) and collaboration (Bedwell et al., 2012). Furthermore, Salas et al. (2000) pointed out communication as one of the means available to achieve team adaptation. Thus, it seems evident that communication is vital to team functioning.

Communication

There are a range of different definitions of the term communication. Mainly, these definitions differ on the three dimensions inclusiveness, efficiency, and intentionality (Littlejohn & Foss, 2008). In the present study, communication will be understood as “the transmission of information” (Berelson & Steiner, 1964, p. 254). This definition was selected as it encompasses nonverbal communication, does not require the communication to be efficient (Littlejohn & Foss, 2008), nor limit communication to messages sent with intent. The two latter aspects were considered important as information about efficiency and intentionality will not be available to researchers using objective measures to assess gaze communication (e.g., eye-tracking glasses).

The majority of research conducted on the relationship between team performance and communication has focused on verbal communication (e.g., Mesmer-Magnus & DeChurch, 2009; Patrashkova-Volzdoska, McComb, Green, & Compton, 2003; Urban, Bowers, Monday, & Morgan, 1995). As a result of such studies, it has been established that there exists a strong relationship between verbal communication and team performance. However, advantages of studying variation in verbal communication are likely to be limited in situations with little or highly structured verbal communication (e.g., firefighting or maritime navigation). As such situations allow for nonverbal communication it may be valuable to increase our understanding of this concepts and its relationship to team functioning.

The current paper will focus on gaze as a range of studies have indicated its significant role in social interaction (e.g., Ellsworth & Ludwig, 1972; Emery, 2000; Frischen & Tipper, 2004; Kleinke, 1986; Pfeiffer et al., 2012; Symons, Hains, & Muir, 1998). Gaze can be argued to have several communicative properties associated with it both in terms of sending and interpretation of messages. In this manner, gaze might be an important aspect of communication in teams. Insight into gaze processes is therefore considered important to attain a more comprehensive understanding of human communication and team functioning.

Gaze

The role of gaze in social interaction has received increased interest in the last decades (see Frischen, Bayliss, & Tipper, 2007, for a review). In this paper, the term gaze will be used to refer to the process where “an individual [is] looking at another person” (Knapp, Hall, & Horgan, 2013, p. 296). Gaze can also be defined more specifically according to whether it is directed toward the face or eyes of another person, which is referred to as face-gaze or eye-gaze, respectively (Kleinke, 1986). Gaze can be quantified in several ways of which three of the most widely used measures in psychological research are fixations, saccades, and dwells. Fixations can be understood as the period of time of which eye movements almost pause at a stimulus, whereas saccades refer to rapid, direction-changing eye movements (Land, 2011). A dwell is often of longer duration and can be defined as the length of a visit within a predefined visual region (Holmqvist et al., 2011).

Objective measures of gaze allow researchers to examine the role of gaze communication in processes such as SMM and team performance. In the current study, the term gaze communication is defined by combining elements from the presented definitions of communication and gaze. Thus, gaze communication will be understood as a transmitted message that, regardless of intention and effect, results from directing visual attention toward another person’s face. According to this definition, gaze communication is important because it allows us to make inferences about the internal states of others based on the extent to which they look at us. Gaze direction (direct or averted) has been found to influence the assessments of others' trustworthiness (Bayliss & Tipper, 2006; Wyland & Forgas, 2010), friendliness (Wyland & Forgas, 2010), likeability, and attractiveness (Mason, Tatkow, & Macrae, 2005). Moreover, the extent to which angry and fearful faces elicit amygdala activation in observers appears to depend on whether the gaze is direct or averted (Adams, Gordon, Baird, Ambady, & Kleck, 2003). Reasonably, gaze has been proven significant not only to the interpretation of others' internal states, but to the expression of internal states as well. For instance, the importance of the eye region has been recognized for expression of a wide array of emotions such as joy, fear, anger, sadness, surprise, and disgust (see Itier & Batty, 2009, for a review).

Such studies indicate that eye-gaze is of particular importance in social interaction. This assumption is further supported by a range of studies demonstrating sensitivity to gaze early in life (e.g., Batki, Baron-Cohen, Wheelwright, Connellan, & Ahluwalia, 2000; Farroni, Menon, & Johnson, 2006; Hains & Muir, 1996). For instance, there is evidence that infants from birth prefer direct rather than averted gaze (Farroni, Csibra, Simion, & Johnson, 2002; Farroni et al., 2006). Drawing on similar and further research, Baron-Cohen (1995) suggested that humans have an innate gaze module, namely the eye-direction detector. The eye-direction detector is proposed to identify eyes in the surroundings, calculate whether the gaze orientation is direct or averted, and imply the objective of others' attention and perceptual state of mind. According to Baron-Cohen (1995), the eye-direction detector is one of several components enabling individuals to develop a theory of mind. Theory of mind is the ability to ascribe mental states (e.g., intentions, knowledge, and thoughts) to oneself and others and can provide the basis for predictions about others' behavior (Premack & Woodruff, 1978).

Baron-Cohen's (1995) proposition about the enabling role of the eye-direction detector is supported by a positron emission tomography study indicating that the same brain areas are involved in both gaze processing and exposure for tasks designed to measure theory of mind (Calder et al., 2002). A review by Itier and Batty (2009) further points out that the existence of an eye-direction detector is supported by evidence indicating that neuronal activation depends on whether the gaze is directed toward the eyes or face, and that a visual scanning of the face starts by directing attention toward the eyes. Baron-Cohen's (1995) theory has, however, been criticized for underestimating the importance of other bodily clues when encoding perceptual states (Langton, Watt, & Bruce, 2000). As the provided data are ambiguous, and few studies specifically have been examining the brain's response to eyes in isolation, it can be concluded that we do not yet have sufficient evidence to determine whether or not an eye-direction detector exists (Itier & Batty, 2009).

Although the existence of the eye-direction detector remains inconclusive, there is substantial evidence of a relationship between gaze and attention. For instance, researchers have observed that gaze may effectuate an attentional shift toward the same stimulus in the observer (Driver et al., 1999; Friesen & Kingstone, 1998; Friesen, Moore, & Kingstone, 2005). This is often referred to as the gaze cuing effect which is associated with several factors such as gender (Bayliss, Pellegrino, & Tipper, 2005) and empathy (Alwall, Johansson, & Hansen, 2010). The relationship between eye movements and attention is also recognized in the eye-mind hypothesis. The eye-mind hypothesis suggests that attention is directed toward the stimulus for as long as the fixation endures, and the fixation is therefore assumed to be indicative of the time required to process visual information (Just & Carpenter, 1980). However, although several studies suggest that visual attention and eye movements are strongly related, these processes do not appear to be perfectly synchronized (Baldauf & Deubel, 2008; Deubel, 2008; Kowler, Anderson, Dosher, & Blaser, 1995). As Just and Carpenter (1980) held that "there is no appreciable lag between what is being fixated and what is being processed" (p. 331), the eye-mind hypothesis shows signs of being too rigid.

However, there is little doubt that eye movements and attention are closely related processes, and although the order of these processes is debated (see Hutton, 2008, for a review), the underlying assumption of the eye-mind hypothesis appears to be sound. It is, however, important to bear in mind that there is no one-to-one relationship between gaze and attention. This statement can be illustrated by studies showing that we are able to let our minds wander away from stimuli in the physical environment (Schooler et al., 2011) and to continue the processing of information no longer visible to us (e.g., visual memory, Sperling, 1960).

Team cognition

Team cognition can be defined as "cognitive activity that occurs at a team level" (Cooke, Gorman, & Winner, 2007, p.240). According to Cooke et al. (2007), the team cognition concepts that have received most attention during the last decades are team situation awareness and SMM or team mental models (TMM). For a discussion of the conceptualization of SMM and TMM, see Klimoski and Mohammed (1994; Mohammed, Ferzandi, & Hamilton, 2010). SMM differ from other team cognition concepts by encompassing both taskwork and teamwork processes (Mohammed & Dumville, 2001). The term taskwork is used to describe the team's utilization of equipment and systems to accomplish a task (Bowers, Braun, & Morgan, 1997; Marks et al., 2001), whereas teamwork can be defined as the "members' interdependent acts that convert inputs to outcomes through cognitive, verbal, and behavioral activities directed toward organizing taskwork to achieve collective goals" (Marks et al., 2001, p. 357). As other team cognition concepts usually focus on either teamwork or taskwork, it can be argued that the SMM construct is more inclusive (Mohammed & Dumville, 2001). The current study will therefore focus on SMM in the further presentation of team cognition.

The SMM construct was presented by Cannon-Bowers and Salas in 1990 as an attempt to identify one of the factors that account for differences in team performance

(Mohammed & Dumville, 2001). In the years that followed, numerous studies recognized the association between SMM and team performance (e.g., Edwards, Day, Arthur, & Bell, 2006; Lim & Klein, 2006; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000; Rouse, Cannon-Bowers, & Salas, 1992; Stout, Cannon-Bowers, Salas, & Milanovich, 1999). A mental model can be defined as the “internal representations that individual cognitive systems create to interpret the environment” (Denzau & North, 1994, p. 4). However, SMM do not equal the sum of the team members’ mental models, but develop as a result of interactions between team members (Cooke et al., 2007). SMM can be defined 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 team members” (Cannon-Bowers, Salas, & Converse, 1993, p. 228). In essence, SMM are the team’s common and organized understanding of information relevant to solve the task (Mohammed & Dumville, 2001). This understanding is based on predictions of the team members’ planned actions and need for resources, which in turn allow them to act in a consistent and coordinated manner (Mathieu et al., 2000).

Although the general conception of SMM appears to be consistent across scholars, there is no universal agreement on how many and which factors that are encompassed by SMM (Mohammed, Klimoski, & Rentsch, 2000). For instance, Cannon-Bowers et al. (1993) suggested that SMM reflect the extent to which team members hold shared knowledge about task, equipment, team interaction, and team members. Mathieu et al. (2000), however, argued that these four factors can be reduced to two major factors, namely taskwork and teamwork. Conversely, Johnson et al. (2007) argued that SMM consist of a total of five factors: knowledge of team and task, task and communication skills, attitudes toward team and task, team dynamics and interaction, and team resources and working environments. These factors were suggested on the basis of exploratory and confirmatory factor analyses which included components from different SMM frameworks in the literature (Johnson et al., 2007). To summarize, although a relationship between SMM and performance has been established, the comprehension and application of the concept appear to vary between frameworks.

A related criticism is directed toward the issue of measuring SMM (Mohammed et al., 2000), supported by the fact that no consistently applied methodology exists (see Mohammed et al., 2010, for a review). Mohammed et al. (2000) reviewed four of the instruments used to measure SMM (pathfinder, multidimensional scaling, causal cognitive mapping, and text-based cause mapping) and concluded that none of these clearly stood out as the best choice of technique. Instead, they recommended that researchers make their choice of technique based on qualities characterizing the relevant research question and context (Mohammed et al., 2000). Therefore, in the current study, the instrument provided by Johnson et al. (2007) will be used as this is a general, empirically developed measure on SMM.

The relationships between gaze, SMM, and team performance

Studies explicitly examining the relationship between gaze and SMM are scarce, but one study examining this link was identified. Feldman, Lum, Sims, Smith-Jentsch, and Lagatutta (2008) conducted a simulator study where eye-tracking glasses were used to record fixations patterns and durations. The mental model similarity (i.e. SMM) was operationalized as the strength of the correlation between two participants' card sort. Feldman et al. (2008) showed a significant relationship between similarity in fixation patterns and SMM, and concluded that individuals whose gaze patterns are similar when solving a task have a higher probability of recalling events in the same manner.

The study by Feldman et al. (2008) suggested that accessing the same visual information might provide a basis for, or at least enhance the probability of, developing SMM. In an ecological environment visual stimuli will vary in importance and value. The importance of the stimulus can be communicated through gaze cueing, which has been shown to elicit a shift in visual attention toward the same stimulus in the observer (Driver et al., 1999; Friesen & Kingstone, 1998; Friesen et al., 2005). The gaze cueing effect might consequently facilitate shared attention and increase similarity in gaze patterns between team members. This process can in turn lead to greater SMM (cf. Feldman et al., 2008). The gaze cueing effect has been supported by a range of studies that apply variations of the Posner cueing paradigm in laboratory experiments (for a review, see Frischen et al., 2007). However, the studies' real world applicability has been questioned (Kingstone, Smilek, Ristic, Friesen, & Eastwood, 2003), and the ecological validity of such findings has, to my knowledge, yet to be proved in field studies. It could be that the gaze cueing effect will not occur as frequently in a field setting as in a laboratory setting. Attention is a selective process (e.g., Broadbent, 1958; Deutsch & Deutsch, 1963; Treisman, 1969), and it seems reasonable to assume that gaze cueing is elicited by a notable stimulus such as the team member's rapid shift in visual attention or prolonged focus on a visual stimulus. Hence, it is possible that gaze cueing will occur more frequently under circumstances that facilitate face-gazing. Gaze cueing can be understood within the framework of Baron-Cohen's (1995) eye-direction detector, and it is thus possible that gaze cueing might enable a theory of mind through shared attention.

Gaze might further promote a theory of mind by communicating internal states (Adams et al., 2003; Bayliss & Tipper, 2006; Itier & Batty, 2009; Mason et al., 2005; Wyland & Forgas, 2010) that are not openly discussed in a team. By providing an additional information channel, the interpretation and sending of verbal information can be supported and supplemented. For instance, the amygdala has been shown to respond to eyes that convey fear through an increased area of eye whites (Whalen et al., 2004). Consequently, a team member might be able to understand that the other is experiencing fear even if this is not expressed verbally. This insight into another's internal states has been proposed to enable prediction of behavior (Premack & Woodruff, 1978) and might enable operators in teams to adapt to the situation and coordinate their behaviors (cf. Cannon-Bowers et al., 1993). Thus, it is possible that there exists a relationship between gaze and SMM. To illustrate, knowing whether a team member is anxious about performing a task aspect might be valuable in order to make accurate predictions of the team member's behavior and adequately identify situations that might require the team to offer support. Adequate team interventions might enhance team performance, and it is therefore possible that there is a relationship between gaze communication and team performance as well. Assuming that the proposed relationship exists, it is possible that SMM constitute a mediating variable (see Figure 1). This assumption is grounded in the fact that studies consistently have shown evidence of a positive relationship between SMM and team performance (e.g., Edwards et al., 2006; Lim & Klein, 2006; Mathieu et al., 2000; Rouse et al., 1992; Stout et al., 1999).

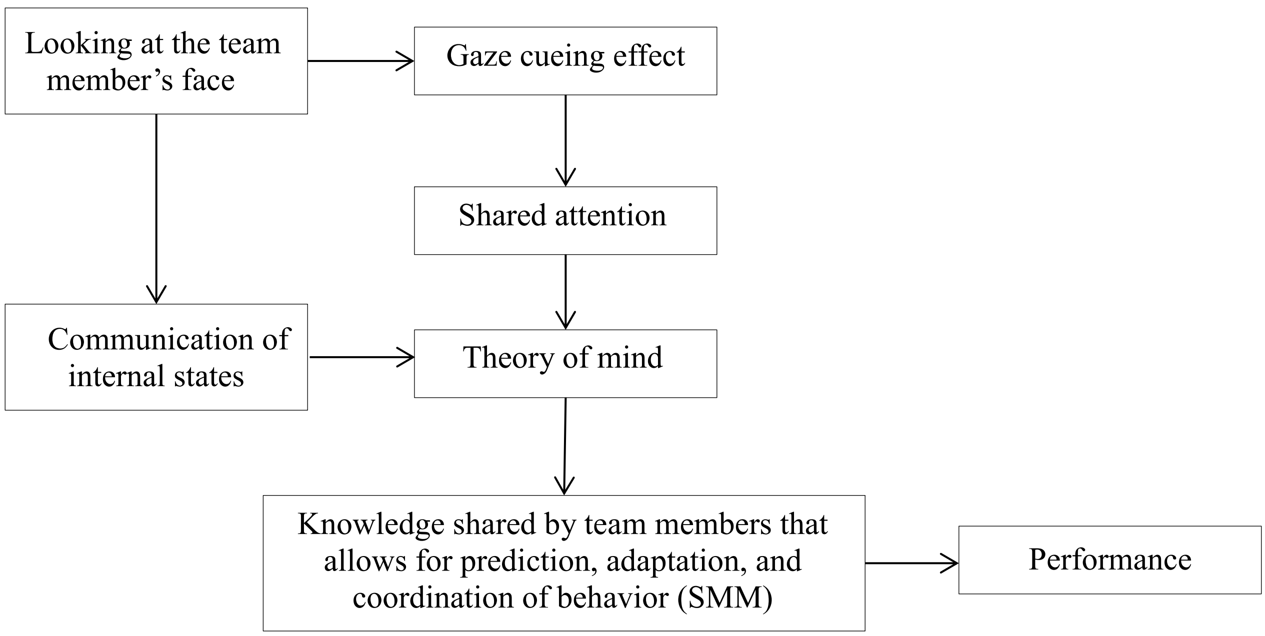

Figure 1. The model shows the suggested relationships between gaze, SMM, and team performance.

Team training

As noted in the opening, human errors are a significant source of accidents in the maritime industry (Rothblum, 2000). This notion has given rise to training programs aiming to enhance the non-technical skills of operators. Non-technical skills may be defined as “the cognitive, social, and personal resource skills that complement technical skills, and contribute to safe and efficient task performance” (Flin, O’Connor, & Crichton, 2008, p. 1). Hence, the goal of enhancing the non-technical skills of operators through team training programs is to increase the safety on board the vessel.

Team training can be understood as the use of a set of evidence-based tools and methods to facilitate team members’ (a) learning about team relevant aspects, such as taskwork and role expectations, and (b) understanding and practice of required competencies (Salas & Cannon-Bowers, 1997). There exists a range of different team training interventions that endeavor to enhance team effectiveness (for a review, see Buljac-Samardzic, Dekker-van Doorn, van Winjgaarden, & van Wijk, 2010) and two examples are Crew Resource Management (CRM) and Bridge Resource Management (BRM). CRM is a training program which aims to enhance the non-technical skills of flight crews, and was developed as a consequence of the acknowledgment that human errors could be the leading cause of aviation accidents (Flin, O’Connor, & Mearns, 2002). Factors included in CRM-based training are typically teamwork, leadership, situation awareness, decision-making, communication, and personal limitations (Flin et al., 2002). A review study assessing 58 studies indicated that CRM aviation training was associated with positive effects on behavior (Salas, Burke, Bowers, & Wilson, 2001).

In the maritime industry, the counterpart to CRM is BRM. So far, few studies have examined the effect of BRM training (Hetherington et al., 2006; O’Connor, 2011). O’Connor (2011) identified and reviewed three studies on the effectiveness of BRM, and concluded that BRM training was less effective in improving human factors (i.e. knowledge and attitudes) than CRM training had been reported to be. This discrepancy can possibly be explained by the lack of knowledge about human factors in the maritime domain (O'Connor, 2011). A meta-analysis that investigated the effectiveness of team training interventions in general concluded that there was a positive moderate relationship between training and outcomes, particularly concerning team processes (Salas et al., 2008). To summarize, these findings suggest that non-technical skills can be learned through the use of training programs, but that the program content should be based on research carried out in the relevant domain. Thus, it appears possible that gaze patterns can be altered through training as long as the training course is relevant to the arena in which the skill will be practiced.

Rationale for the pilot study

This pilot study aimed to examine the feasibility of a main eye-tracking study that will examine the relationship between gaze communication and team functioning in an ecological setting. The rationale for performing a pilot study rather than a large-scale study was twofold. First, to attain the power needed to be able to test the relationship between gaze communication and team functioning, a sufficient sample size would be required. Given the restricted availability of eye-tracking equipment and the time involved in setting up and preprocessing the recordings, a large-scale data collection in an ecological environment would be resource-demanding. Second, the research group had limited experience with the methodology, which is a factor that may negatively influence the quality of gaze data (Holmqvist et al., 2011). The study was therefore at risk of not yielding testable data, and it was consequently judged not justifiable to conduct such a resource-demanding large-scale study. The research group therefore chose to conduct a pilot study in the context of a scheduled navigation course for students at a university college in order to assess the feasibility of a main study planned for the future.

Objectives for the main study

General setting. As it is expensive to develop scenarios and use simulators to recreate real situations and challenges with trained personnel for research purposes, the data collection for the main study will be done in a training program for bridge personnel. Two-person teams composed of one captain and one second officer will plan and perform a navigation task during which the captain's eye movements will be recorded. As this task will be part of the training program, the research aims of the data collection must be balanced against educational aims. For instance, while it from a research perspective would be important to maintain scenarios and tasks unchanged throughout the data collection, a common course objective would be that students should gain experience with different scenarios and tasks. A further challenge will be the presence of confounding variables. In this setting, the passage of time has the potential to be one of the most significant confounding variables. First, the extent to which participants have experience with the simulator may lead to variations in performance between different measurement points. Second, an intervening treatment is planned for the main study, and difference in the time passed between the treatment and measurement may lead to variations in the treatment's effectiveness. Thus, the study design must be able to control for effects of time by using several measurement points and keeping the timespan between study components constant for all teams.

Hypotheses. Three pairs of hypotheses are planned for the main study. In order to examine the impact of both frequency and duration of gaze, the wording will be similar for each pair of hypotheses with the exception of the operationalization of the amount of gaze. The first two hypotheses are based on the studies presented above that have linked gaze and SMM. These hypotheses state that face-gaze frequency (Hypothesis 1) or face-gaze duration (Hypothesis 2) will be associated with higher SMM as measured by the similarity between team members' responses on a self-report questionnaire. The second pair of hypotheses is based on the presented literature concerning the established relationship between SMM and performance. Thus, these hypotheses state that face-gaze frequency (Hypothesis 3) or face-gaze duration (Hypothesis 4) will be associated with higher team performance. The third pair of hypotheses is based on the presented literature concerning team training, which suggests that non-technical skills can be acquired through such training. This will be tested by giving a brief lecture about the significance and function of gaze communication. These hypotheses state that training in gaze communication will be associated with increased face-gaze frequency (Hypothesis 5) or face-gaze duration (Hypothesis 6).

Analytic approach. All of the proposed hypotheses for the main study will be tested by measuring the frequency and duration of gaze with eye-tracking glasses. In the main study, SMM will be measured by calculating the average of the team members' responses on a self-report measure for each of the five factors (Johnson et al., 2007; Johnson, Top, & Yukselturk, 2011). Team performance will be operationalized in terms of expert ratings, simulator data, self-reported quality of communication, and the frequency of unwanted incidents (e.g., collisions and groundings).

Hypotheses 1 through 4 predict that the extent to which the captain face-gazes at the team member will be related to an outcome measure. These hypotheses will be examined by conducting simple linear regression analyses to test the relationship between gaze communication and SMM, and gaze communication and performance, respectively. Face-gaze will constitute the predictor variable in these hypotheses, while SMM and team performance will represent the effect variable for their respective hypotheses. While the two first hypotheses will be tested with one predictor variable and one effect variable, the two latter hypotheses will be tested with several effect variables. The relationships between face-gaze and performance measures can be tested by running several simple regression analyses as suggested or, alternatively, a summarized score for all performance variables can be calculated and treated as one effect variable.

The matched-subjects between-subjects design allows for the examination of a training effect to test hypotheses 5 and 6 while avoiding variability between groups to be explained by individual differences. The participants will be tested before (baseline) and after (posttest) an intervening treatment to assess whether training will have an effect on gaze communication in teams. One may expect considerable individual variation in baseline measurement of face-gaze (see Frischen et al., 2007, for a review). In order to prevent random differences from hiding group differences due to few data points, participants will be assigned to a treatment group or a control group according to their baseline eye-tracking scores (a matched group approach). The treatment group will attend a brief lecture explaining the benefits and functions of gaze in team functioning, while the control group will attend a comparable lecture on navigational issues not related to communication. Hypothesis 5 and 6 will be tested by running two one-way analyses of variance (ANOVA). Two factors will be included in these analyses: The respective gaze measure (frequency or duration) and group affiliation (treatment group or control group). This approach will allow the researchers to test whether there are statistically significant differences in face-gaze communication between the two conditions.

Study design for the pilot study

The primary goal for the present pilot study was to assess the feasibility of the main study in terms of participant flow in the pilot study and the resources required to carry out the main study. As recommended by Thabane et al. (2010), outcome measures and feasibility criteria have been specified in the Methods section. Feasibility has been operationalized as whether or not the specified criteria were met. The Discussion section covers explanations for the results, a discussion of scenarios and the use of eye-tracking equipment, a sample size estimation, and suggested improvements to the study design.

Methods

The data collection was carried out in collaboration with the supervisor and a scientist with affiliation to the university college at which the data collection took place.

Data collection setting

The pilot study data were collected in a mandatory training course in maritime navigation, and data were collected over seven days. The university college had divided the class of 39 students into three exercise groups of 12-13 students due to capacity constraints. The three exercise groups performed the same navigation task consecutively on the same day. This schedule allowed us to record one team for each exercise group during one day of data collection which equals a maximum of three teams to be recorded a day. As we had planned to record each team twice, and we had seven days to collect the data, a maximum of nine teams could be eye-tracked during the data collection period.

The organization of the course was as follows: First, and prior to the navigation exercise, educational personnel briefed the class about the scenario and gave practical advice on how to accomplish the task successfully. This session lasted for approximately 10 minutes. Second, students coupled up with one fellow student of their own choice, and decided for themselves who would be acting captain and who would be second officer. For students participating in the study, teams and roles would be constant throughout the data collection period. The captain was responsible for operations and decisions at sea while the second officer was responsible for the navigation. Third, students were expected to accomplish a task on the simulator bridge that included both preparation of the navigation route (the planning phase) and maneuvering through the exercise (the navigation phase). When the simulation was over, the class was gathered for a debrief that lasted for about 10 minutes. The total duration of the exercise for each group (including briefing and debriefing) was two full hours.

Participants

Fourteen students enrolled in a navigation course during their second year of a Bachelor's degree program in maritime navigation were recruited to participate in the eye-tracking experiment. A further eight participants were recruited from the same population to complete surveys only. The sample was predominately male (86.4%), and the participants were between 19 and 40 years of age (M = 24.91, SD = 6.47). All participants had completed simulator training involving both navigational instruments (10 ECTS) and navigation (5 ECTS). There was no monetary incentive for participating. Debrief included a summary of the findings on a group level.

Eligibility

Assessment of eligibility was limited to determining whether or not students interested in participating in the simulator experiment wore glasses. This was done because glasses may complicate the eye-tracking of participants (Holmqvist et al., 2011). Also, instituting this exclusion criterion was practical and not resource-demanding. This procedure is congruent with the screening process planned for the main study. There are several additional individual characteristics that may complicate the eye-tracking of participants, e.g., mascara, droopy eyelids, and contact lenses (Holmqvist et al., 2011). However, a complete screening process of the participants' eye health was regarded as too intrusive for the purpose of the main study, and the screening was thus limited to whether or not students wore glasses.

Procedure

All students enrolled in the navigation course were given the opportunity of participating in the study. The students were presented with the aim of the study and other practical information by our collaborating partner employed at the university college.

Interested students were invited to sign up for the study. Students wishing to postpone their decision were invited to sign up at their earliest convenience. The study was also presented verbally by the researchers on the first day of data collection. This presentation took place during the briefing with the participants' instructors present. At this point, a written informed consent form was distributed to students interested in participating. The students could choose to participate in the field experiment with subsequent completion of questionnaires or to complete questionnaires only. Students were offered the latter possibility to increase the testability of a feasibility criterion regarding questionnaires, and the opportunity to complete questionnaires only will not be part of the procedure for the main study.

To maximize the utilization of available resources, we needed to recruit nine teams to participate in the eye-tracking experiment. As fewer teams than necessary initially expressed their interest, students were also offered the opportunity to sign up for the study on each new day of data collection until all teams had been through the baseline assessment. In the end, seven teams (A-G) volunteered to participate in the study. Data collection was conducted one day a week over seven of nine consecutive weeks. Data were not collected on the second and seventh week of data collection as training was not arranged those weeks. On the first day of data collection a test trial was performed to examine practical issues such as timing, calibration, and lighting. An overview of the time schedule for collection of eye-tracking data and the time points for the intervening treatment is provided in Table 1. The first two weeks are excluded as these are not considered relevant for the purpose of this table.

Table 1
The table shows the time schedule for the pilot data collection

Week	Baseline recording	Treatment	Posttest recording
Week 3	Teams D-E		
Week 4	Teams F-G		
Week 5	Teams A-C	Teams D-E	
Week 6		Teams A-C, teams F-G	Teams D-E
Week 7	*No data collection*	*No data collection*	*No data collection*
Week 8			Teams F-G
Week 9			Teams A-C

Based on our assessments of the test trial, in which three of the teams participated in addition to their ordinary assessments, we concluded that there were challenges related to the dim lighting in the simulator. Due to difficulties in identifying the second officer's face in the recordings, we decided to equip these participants with glow sticks to increase the visibility of the face area. As the luminosity was considered to be a problem even after the introduction of glow sticks, a floor lamp with a red light bulb was placed in the simulator on the fourth day of measurement. Both the glow sticks and the lamp were retained throughout the study. Further, we noticed that the teams appeared to gaze longer and more frequently at each other during the planning compared to navigation, and decided to eye-track the planning phase in addition to the navigation phase. No further changes were made to the planned procedure based on the test trial.

Before the planning phase commenced, the participant playing the role of the captain was outfitted with eye-tracking glasses and underwent a nine-point calibration in the hallway to calculate the direction of gaze. The participant who was second officer was equipped with a glow stick necklace. Recording started after the calibration was completed, and the team began planning the navigation on the simulator bridge. The same simulator was used for eye-tracking throughout the study. The planning phase was expected to last between 10 to 20 minutes. Educational personnel communicated the beginning of the navigation phase over radio. The captain's eye movements were recorded for approximately the first 30 minutes of this phase and the team's navigation performance was rated by an expert. A few of the expert ratings were vague, and in such cases we repeated the question intended to assess performance to obtain a clear response from the expert. The questionnaire, which takes about five minutes to complete, was handed out in the debriefing. (See Appendix 1 for the complete questionnaire). All participants were expected to complete the questionnaire subsequent to each scenario exercise (six questionnaires in total). The students returned the questionnaires before leaving the area.

In the current pilot study, the procedure deviated from that of the main study in some regards. First, the aim was to introduce the course two weeks after the baseline test and one week previous to the posttest, but due to reorganization of classes these standards could not be complied with in the pilot study. For three of the teams the course was arranged one week after the baseline and three weeks before the posttest. For two of the teams the course was arranged two weeks previous to the posttest. Besides from these deviations, the data collection was performed in accordance with the planned time schedule. Second, participants were not assigned to one treatment group and one control group as specified for the main study. The aim of this pilot study was to examine the feasibility of the main study, and assignment of students to different experimental groups was not necessary to reach this goal. Thus, while the main study will have a between-group design, the pilot study applied a within-subjects design in which all teams received the same treatment and were measured repeatedly.

Scenarios

The data collection was done in mandatory classes on the navigation course, and for educational purposes the scenarios were different for each week. Scenarios were developed and administered by the educational staff. The educators estimated the overall cognitive and team workload of the six scenarios to be comparable. Scenarios varied in terms of factors such as task, sea, traffic density, and time of day (see Table 2). Although the teams performed the same task in the same environment simultaneously, the teams did not receive data about the positioning of the other teams' vessels, with the exception of Scenario 5, in which the teams had to communicate with each other to solve the task.

Table 2.
The table shows characteristics of the scenarios used in the data collection.

Nr.	Task	Sea	Traffic density	Time of day	Other
1	Cross a traffic artery, and cast anchor.	Calm	High	Midday	*N/A*
2	Sail into harbor, navigate a canal, and dock the vessel.	Calm	Low	Morning	*N/A*
3	Navigate buoys, sail into harbor, and dock the vessel twice.	Calm	High	Night	*N/A*
4	Sail into harbor, dock the vessel, and leave the dock.	Moderate	Low	Morning	Reefs
5	Respond to mayday, make contact over radio, and search for man overboard.	Moderate	Low	Morning	Fog
6	Sail into harbor, and dock the vessel twice.	Calm	High	Midday	*N/A*

Note. N/A = not applicable

Measures

Eye-tracking. Eye movements were measured using the Tobii Glasses Eye Tracker; a monocular (right eye) eye-tracker with a sampling frequency of 30 hz. This involves the eye-tracker registering the gaze direction of the participant’s right eye 30 times per second. Available systems range from approximately 30 Hz and up to 2000 Hz, and the applied eye-tracker can consequently be considered a relatively slow system. According to the Nyquist-Shannon sampling theorem, the sampling frequency should be at least twice the length of the movement that is to be measured (Holmqvist et al., 2011). Consequently, the glasses were not considered sufficiently sensitive to register commonly used measures such as saccades (30-80 ms) and fixations (200-300 ms). However, while fast systems and lab settings are required for studying cognitive processes in the millisecond range, comparably slow systems may be sufficient for team processes in the tenth of a second range in noisy field environments. The selected eye-tracking equipment was therefore considered suitable to measure the frequency and duration of dwells.

SMM. The extent to which team members had SMM was measured using a scale adapted from Johnson et al. (2007). The original instrument consisted of 42 items that have been identified as comprising five factors through exploratory and confirmatory factor analyses (Johnson et al., 2007). As the scale was too long to be included in our questionnaire, and is not in wide use in its current form, several of the items were excluded. Items that had a factor loading below .65 or were irrelevant for the current setting were excluded. Consequently, 22 items were retained. These items were translated into Norwegian, and wording was adapted to the study's population and context.

The factor *general task and team knowledge* originally consisted of nine items, and five of these were included in our survey. An example item is "My team knows the general process involved in conducting a given task." The factor *general task and communication skills* was originally composed of seven items, of which four were retained in the current study. An example item is "My team informally communicates with one another throughout various team tasks." The factor *attitude toward teammates and task* originally consisted of eight items, and five of these were included in the study. An example item is "My team is committed to the team goal." The factor *team dynamics and interactions* was composed of nine items, and five of these were retained. An example item is "My team understands how they can exchange information for doing various tasks." The final factor *team resources and work environment* also originally consisted of nine items, but was reduced to three items in the current study. An example item is "My team has a positive team climate." The responses were rated on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Communication. Communication was measured using a self-composed scale consisting of eight items. A validated communication scale was not used in thus study as an instrument suitable for teams consisting of two inexperienced team members could not be identified. Most communication scales applicable in a team setting are developed for use in larger, constant teams that have a regular set of assignments. Therefore, a scale based on a review of the communication literature was developed. The items were adapted to fit the participants and the context being studied. Responses were rated on a Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Accidents. Accidents and risk of accidents were measured by adapted versions of items repeatedly used in surveys by the Operational Psychology Research Group (University of Bergen). The items “How likely do you think it is that some of your operations during the exercise could have led to a personal injury?” and “How likely do you think it is that some of your operations during the exercise could have led to an accident with material damage?” were rated on a 7-point Likert scale ranging from 1 (*very unlikely*) to 7 (*very likely*). The third item “Was the vessel involved in an accident?” was answered with “Yes” or “No.”

Expert ratings. An expert was defined as educational personnel involved in the simulator training that had been employed as a captain at sea for at least 20 years. Each team was rated by one expert, and three different experts performed the ratings during the data collection. The following question was asked: “How do you rate the team’s goal accomplishment on a scale from 1 to 10, where 1 is the worst possible performance and 10 is the best possible performance?”

Simulator data. Objective information about the vessel’s navigation route, speed, turns, and number of accidents were planned to be collected from the simulator. Such data might be an objective indicator of performance efficiency if criteria are set according to the scenario’s goals. The issue of retrieving these data was discussed between researchers and educational personnel before the data collection commenced. However, due to missing information and absence of qualified personnel, the necessary criteria could not be predefined. As it was not customary to automatically store simulator data, these data were not retrieved in the present data collection.

Processing of eye-tracking data

Eye-tracking data was processed using Tobii Studio 3.2.1. This software provides information that can be used to assess the trackability of recordings. Trackability can be defined as “the proportion of raw data samples that are lost during a recording” (Blignaut & Wium, 2014, p. 68). Hence, the trackability decreases as the amount of data loss increases. In the same way that a small sample size is known to be unreliable (Tversky & Kahneman, 1971), a recording that has low trackability might provide false results (Holmqvist et al., 2011). Therefore, it is important to exclusively retain data that are considered valid, i.e. data that give an adequate impression of the participant’s visual focus. However, there exists no generally accepted criterion for deciding which recordings that can be considered valid (Holmqvist et al., 2011). In this study, the percentage of the recording that contains gaze data was used as an indicator of validity, as this information is provided by the software. Furthermore, no generally accepted cut-off for whether a recording should be included or excluded from analyses was identified. Other studies have excluded recordings which retained less than 80% of the gaze data (Komogortsev, Gobert, Jayarathna, Koh, & Gowda, 2010) or where less than 75% of trials (4 seconds long, 272 trials) were tracked (Isaacowitz, Toner, Goren, & Wilson, 2008). The study by Isaacowitz et al. (2008) differs from the current pilot study as it applies the cut-off to the percentage of trials considered successful for each participant. Nevertheless, since the trials were of high numbers, the study was considered worthy of comparison. In this pilot study, recordings were included if they met the criterion of containing at least 60% of gaze data. The basis for deciding a lower cut-off was that inexperienced use of eye-tracking equipment could compromise the calibration quality (Holmqvist et al., 2011), which in turn could influence the trackability.

The preprocessing of eye-tracking data was done manually. That is, the region that is of relevance to the research question, the area of interest (AoI; Holmqvist et al., 2011), was marked frame by frame by the analyst. The material was coded using dynamic AoIs since the relevant stimulus (i.e. the second officer's face) moved. Therefore, it was necessary to specify the shape, size, and position of the stimulus for each new frame (Papenmeier & Huff, 2010). The analyst can be considered "blinded" to the target stimulus during the preprocessing of eye-tracking data, as the gaze indicator provided by the software is inactive when AoIs are assigned.

Two different coding schemes were tested. The most straightforward coding scheme ("face area") placed an AoI over the team member's face as accurately as possible (i.e. without including any of the additional surroundings) whenever the face was visible in the field of view. However, based on our test trial, it was anticipated that there would be instances where the team member's face was not clearly discernible in the recording due to dim lighting, but was likely to be within tracked area. An example of this would be when the wearer of the eye-tracker turns his or her head quickly, recording a brief silhouette that may or may not be the head of the team member, thus making it difficult to mark it with the "face area" AoI with any certainty. Ensuring that the second officer's face was covered in this situation would have involved that an AoI significantly larger than the actual face was defined. In the course of the first run at preprocessing, it was decided to assign a dedicated AoI for such cases, called "possible face area". This allowed for a more honest preprocessing, yet maintained all the data, and allowed us to run analyses with or without including gaze at the "possible face area" AoI. However, the significant difference in reliability between these two coding schemes should be kept in mind when interpreting the data. Subsequent to the preprocessing of eye-tracking, the software was used to analyze the recordings for the frequency and duration of gaze directed toward the AoI (i.e. dwells).

In the preprocessing of eye-tracking data, the planning phase was defined as the period of time from when the participant had undergone calibration until the educational staff communicated that the exercise had commenced over the radio. The navigation phase was defined as the period of time from when the exercise had been announced as initiated until the researchers collected the equipment (approximately after half an hour of navigation). However, it was suspected that captains gazed more at the second officer during planning than during navigation. Therefore, eye-tracking data were calculated for both planning and navigation phases. Values were adjusted for different lengths of recordings, and outliers (± 2 SD from the mean) were removed. A confidence level of p <.05 will be used throughout the present study. A two-tailed, paired t-test was carried out to compare dwells during planning and navigation. Results showed a significant difference in the total duration of dwells for planning (M = 14.68, SD = 14.04) and navigation (M = 4.38, SD = 3.57) phases (t (7) = 2.46, p <.05). The same tendency, although not statistically significant, was shown for frequency of gaze during planning (M = 23.57, SD = 25.53) and navigation (M = 12.57, SD = 12.23) phases (t (6) = 1.63, p = .16). As a result of these findings, only the planning phase was included when the valid recordings were coded for the second time.

Outcomes

As the current work was a pilot feasibility study, feasibility criteria were established. Calculations were intended to evaluate to what extent the criteria were met. The outcome measures used to assess feasibility were recruitment rate, participation rates at different stages of the study, proportion of surveys fully completed, expert ratings rate, and trackability rate. The main study was considered feasible if the eight criteria that follow were met.

To maximize the use of available recourses 18 eligible students should be recruited to participate in the field experiment. Therefore, the first feasibility criterion was that 18 of the 39 students in class (i.e. 46.15% ≈ 45%) should be recruited to the pilot study for the main study to be considered feasible. The participation rates at different stages of the study were expected to be high as data were collected during mandatory classes, and participation in the study required limited extra effort. Based on experiences from other studies, 90% of students were expected to participate in the baseline recording, surveys, course, and posttest, respectively. Related to these estimates was the expectation that 100% of expert ratings would be collected since the exercise would not be performed if educational personnel were not present. Another aspect relevant for the main study was the extent to which the length and content of the planned questionnaire was feasible in an educational setting. Including the scale items only, 90% of the questionnaires were expected to be fully completed. Items assessing background information were not included as these questions will probably be adjusted for the main study. The final feasibility criterion considered the trackability rate. Trackability rate was defined as the percentage of the collected recordings that were considered valid. The estimate of the expected trackability rate was based on a study that used participants of similar age, of which 85% of the participants were found to be trackable (Isaacowitz et al., 2008). As their cut-off for validity ((≥ 75 %) was considerably higher than the one used in our study (≥ 60%), the trackability rate was expected to be more than 90% in the current study. See Table 3 in the Results section for a summary of the presented criteria.

The management of resources spent on the study was not quantified by feasibility criteria as the other factors. This was due to the fact that economic allocations for the main study had not yet been made when the pilot study commenced. Further, I did not have information about the resource-demands to be expected from specific processes of the main study besides unsystematic observations. This was a particular concern for the preprocessing of eye-tracking data. Therefore, the present study aimed to examine two aspects of this process. The first aspect was to examine the extent to which training in preprocessing of eye-tracking data is necessary to obtain data that holds sufficient quality to be included in subsequent analyses of gaze patterns. The second aspect concerned the time needed for preprocessing of eye-tracking data when the analyst was experienced in handling the data.

The first aspect was examined by testing whether experience with coding AoIs would impact the results. Valid recordings of the planning phase were therefore preprocessed twice by the same analyst, and frequency and duration of dwells were calculated for each participant. Data from the first and second run at preprocessing were included in a two-tailed, paired t-test to examine whether there was a statistically significant difference between the two runs. None of the recordings were excluded from the analysis. To enable the calculation of the time needed to preprocess eye-tracking data when the analyst was experienced, time needs were documented for each recording when they were preprocessed for the second time. The duration of each recording was also registered to allow for a calculation of the ratio between recording duration and time required to define AoIs. The calculation was made by dividing the total duration of the recordings by the total time required to conduct the preprocessing of eye-tracking data, and 1 was divided by the resulting quotient. Some recordings were aborted due to technical difficulties. Only recordings of at least five minutes were used in the analyses. Data analyses were performed in Excel 2010 and SPSS version 21.

Ethical aspects

This pilot study was registered with the Norwegian Social Science Data Service and conducted in accordance with their guidelines. Our most salient ethical concerns were related to the relationship between students and the university college. First, the study was presented by the educators that would grade the students' final exam, and there was thus a possibility that they would feel pressured to participate in the study. Second, students completed questionnaires on which they rated the quality on their teamwork and communication, and thus provided information that could reveal negative aspects of their own performance. These concerns made it essential to sufficiently inform the students that participation in the study was voluntary and would not influence their course grades. Students were also informed that individual test scores would not be provided to the college or used in other settings.

Results

Seventeen students (43.59%) consented to participate in the eye-tracking experiment. Of these, 14 students approached us to become part of an eye-tracking team and the finite recruitment rate was 35.9%. Figure 2 describes the participant flow through the experiment.

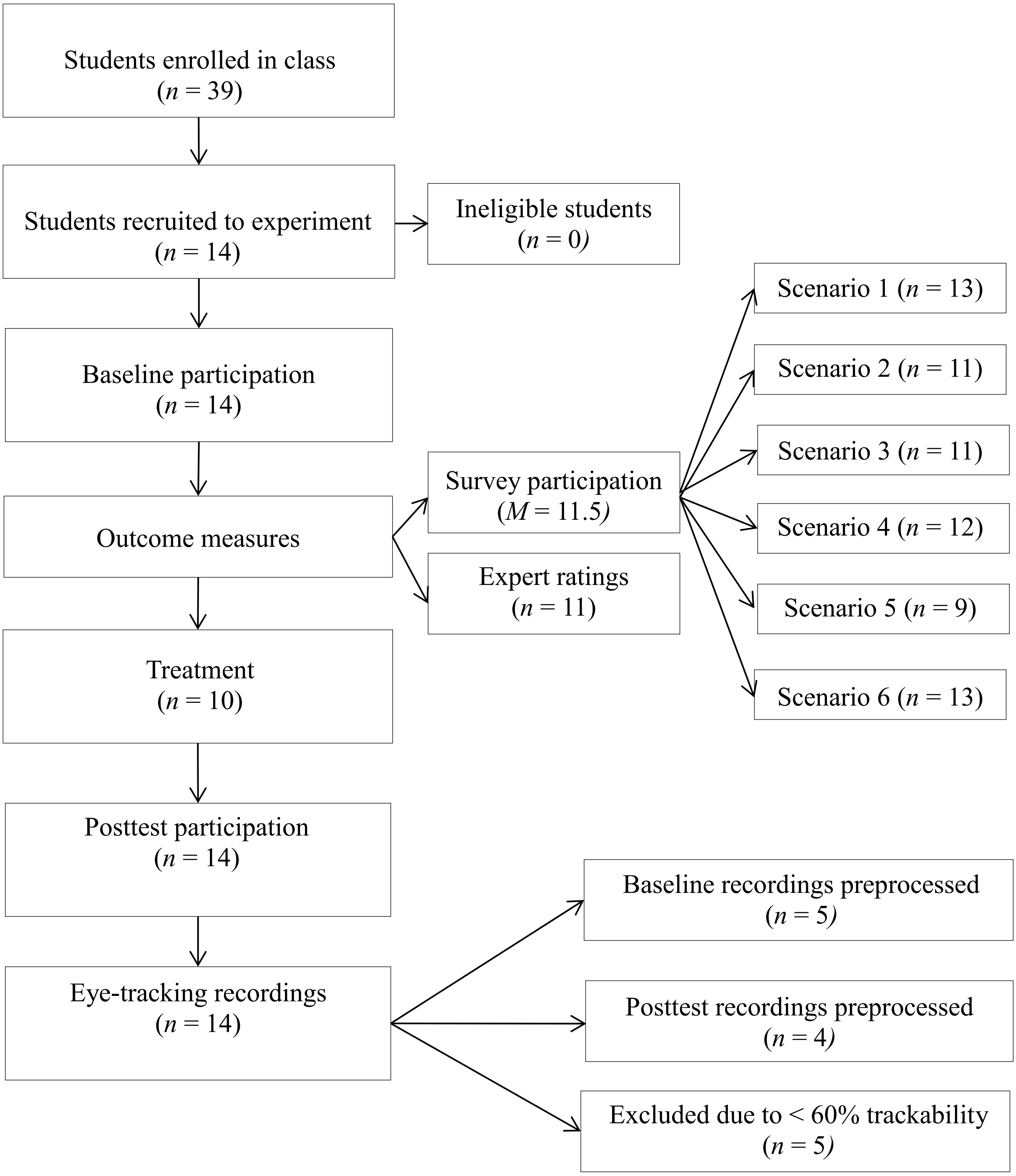

Figure 2. The figure shows the participant flow through the field experiment.

A further eight students (20.51%) chose to fill out questionnaires only. In summary, more than half of the class (56.41%) participated in the study. The total participation rate was thus higher than the target value (> 45%) which may imply that students in general were willing to participate in the study. All participants involved in the field experiment completed both the baseline recording and the posttest. The participation rates for the other study components, however, were lower. Among participants in the field experiment the average survey participation rate for all six scenarios amounted to 82.14% (*M* = 11.5, *SD* = 1.52). The corresponding average rate for participants completing surveys only was 52.13% (*M* = 4.17, *SD* = 1.72). All 22 participants included, 93.81% of the questionnaires were fully completed. The participation rate for course attendance was calculated to 71.43%, and approximately four-fifths (78.57%) of the expert ratings were collected. The finding that differed from the expected value the most was the trackability rate (64.29%). An overview of the extent to which the feasibility criteria were met is provided in Table 3.

Table 3.

The table shows the extent to which the feasibility criteria were met.

Nr.	Feasibility criterion	Attained %	Demand met
1	More than 45% of the students were recruited to participate in the field experiment	35.9%	No
2	More than 90% of the participants participated in the baseline recording	100%	Yes
3	More than 90% of the participants in the field experiment completed the surveys	82.14%	No
4	More than 90% of the participants attended the training course	71.43%	No
5	More than 90% of the participants participated in the posttest	100%	Yes
6	More than 90% of all collected questionnaires were fully completed	93.81%	Yes
7	All of the participants were rated by an expert after the exercise	78.57%	No
8	More than 90% of the collected eye-tracking recordings were valid	64.29%	No

A two-tailed, paired t-test revealed significant differences in the total duration of dwells according to whether the analyst was inexperienced (*M* = 15.27, *SD* = 13.26) or experienced (*M* = 8.14, *SD* = 7.57) in preprocessing data (t (8) = 2.33, p <.05). Further, borderline significant results were identified for the frequency of dwells between recordings coded for the first (*M* = 32.44, *SD* = 33.92) and second (*M* = 17.56, *SD* = 17.13) time (t (8) = 2.15, p = .06). A closer look at the coding pattern revealed that the assessment of sequences where the classification of AoIs was open to interpretation changed from the first to the second run at preprocessing the eye-tracking data. While the “possible face area” coding scheme was applied for two of the recordings in the first run at preprocessing data, this scheme was used for seven of the recordings in the second run. These results indicate that there was a training effect in assigning AoIs.

The ratio between duration of recording and time used to define AoIs was calculated to 1:2.86 ≈ 1:3. The recordings varied from 6.25 minutes to 24.3 minutes in duration (*M* = 16.70, *SD* = 7.33) and the minimum and maximum time used for preprocessing amounted to 15 minutes and 103 minutes (*M* = 48, *SD* = 29.04), respectively.

Discussion

This study aimed to examine the feasibility of a main study of which the objective is to explore the relationship between gaze communication and team functioning. Feasibility was quantified in terms of eight feasibility criteria, and the present study showed that three of the eight target values were attained. This includes participation rates for baseline recording, posttest, and extent of survey completion. However, feasibility criteria for recruitment and trackability rates, survey completion, course attendance, and expert ratings were not met. The present study further examined the resources necessary to use the eye-tracking methodology in the main study. The results showed that there was an effect of experience in defining AoIs, and that the ratio between duration of recordings and time used for preprocessing of eye-tracking data was calculated to 1:3. Although the study did not meet all specified target values, the study did provide insight into potential sources of errors that can be addressed and improved.

This section will cover three main topics. First, possible explanations for difficulties in meeting the feasibility criteria will be discussed. Also, implications and specific suggestions for study improvement will be pointed out. Second, other aspects such as scenarios, sample size, and choice of equipment will be addressed. Third, the feasibility of the study will be discussed, and based on what I have learned by performing this study, suggestions to an improved study design will be proposed.

Feasibility criteria

Recruitment and participation rates. While 35.9% of students in the navigation course participated in the field experiment, an additional one-fifth of the class participated in survey completion only. As about half of the class participated in either the experiment or the survey, it seems as though students in general were willing to participate in the study. However, fewer students than anticipated were recruited to the field experiment. This finding indicates that the study could have been carried out differently to increase the students' motivation to participate.

First, characteristics of the study procedure might have affected the recruitment rate. Some students might have expected the recording of eye-movements to interfere with their performance during navigation (for example by being time-consuming or stressful). If that is the case, it seems reasonable to assume that some students might have decided to refrain from participating in the study for that reason. This explanation is partly supported by the fact that one-fifth of the students decided only to participate in a part of the study (i.e. completing questionnaires) that was arranged subsequent to the navigation exercise. Another aspect could be that characteristics of the structural framework for the data collection might have had impact on the students' access to participate in the experiment. Due to the maximum eye-tracking capacity of three teams per exercise group, it is possible that students could not be included as the upper limit was reached. This hypothesis is supported by the observation that three students consented to being eye-tracked, but failed to come forward at the time of data collection.

Second, the recruitment procedure could have been performed differently. In the initial presentation of the study, students were told that they could sign up for the study at a later date. This means that students would have to sign up for the study at their own initiative. It is possible that the recruitment rate would have been higher if effort had been invested in increasing the students' feeling of commitment to the study. For instance, Lipsitz, Kallmeyer, Ferguson, and Abas (1989) showed that reminders of blood donations that ended with "We'll count on seeing you then, OK?," and awaited a response from the individual, were significantly more effective compared to remainders that took participation for granted.

The participation rates for students that had been successfully recruited to the experiment varied considerably between study components. Whereas the participation rates for both baseline recording and posttest met the target values, fewer participants than expected completed the questionnaires. It is unlikely that the length or content of the questionnaire affected the response rate, as more than 90% of all collected questionnaires were fully completed. However, we were unable to keep all teams constant throughout the data collection, and it is thus possible that students refrained from filling out the surveys as they, contrary to our intentions, had completed the exercise as part of a different team. Further, students were asked to complete the same questionnaire six times. It is possible that students found it difficult to motivate themselves to complete the considerable number of identical surveys. The response pattern indicated that the participation rate for participants recruited to the experiment was highest (92.86%) on days when researchers informed the full class about the project (Scenario 1 and Scenario 6) compared to 76.79% on days without further presentation of the study. This implies that information from the researchers, or the order of the scenarios, affected the students' motivation to participate. Moreover, it seems as though experiencing a feeling of commitment to the study is of major importance, as the average participation rate was considerably lower for participants not included in the field experiment (52.13%).

The component that was found to have the lowest student participation rate was course attendance (71.43%). It was anticipated that arranging the course subsequent to mandatory lectures would facilitate study participation, and additional measures were not taken to increase the attendance rate. Whereas some students might have been absent due to valid reasons, it is possible that others had reasons for absence that could have been addressed in the study (e.g., forgetfulness or insufficient motivation).

Means to increase participation. To address the above listed challenges, it is essential to comprehend their common theme. Such understanding allows for prediction, and potentially avoidance, of similar challenges in the main study. With some exceptions the suggested explanations are, in large, concerned about the informational flow between researchers and students. The majority of the mentioned challenges could probably have been eased by adjusting communication channels, motivating techniques, and information about practical issues.

Communication channels. The presentation of the study was performed by a local educator whom the students trusted. However, it may have been beneficial if the researchers also had been present at the initial presentation of the study. By personally showing up at the college university, researchers would have signified the importance of the study and their appreciation of student participation. Further, the presence of the researchers could have put weight to the statement that individual data were not to be provided to the university college despite the college being a collaborating partner. Also, a different recruitment procedure could have been used that allowed researchers to personally contact each participant to inform them about the dates and locations for baseline recording, course, and posttest. This procedure might have increased the feeling of responsibility for the study which in turn could have increased the participation rates (cf. Lipsitz et al., 1989).

Motivation. Effort was put into making the study relevant to maritime navigation as people in general are more likely to participate in a study if they find the issue directly relevant to their own lives (Galea & Tracy, 2007). However, it is possible that other incentives should have been considered as well. Economic compensation could have had an effect on participation rates (Edwards, Cooper, Roberts, & Frost, 2005). However, incentives should be used with caution as such practice might be ethically questionable in cases where there may be high levels of risk or aversion, a dependency relationship between participants and researcher, or poverty (Grant & Sugarman, 2004). The two latter factors are the most relevant in the present study. As the collaborating researcher who informed about the study was a local educator, it could be argued that there exists a dependency relationship between the two parties. However, this aspect is only considered an issue if participants are given a reward that might apply to their relationship (e.g., assigning course credits to students that might influence their grades, Grant & Sugarman, 2004), which is not the case with a monetary reward. Another issue concerns students' tendency to have a low monthly income (Otnes, Thorsen, & Vaage, 2011), which might make this group vulnerable to exploitation. However, monetary incentives in populations of low socioeconomic status are not normally questioned if the research is not burdensome and might be of interest to the population (Grant & Sugarman, 2004). Hence, the use of monetary incentives to increase participant motivation in the main study appears to be sound as long as the amount is not too high.

Practical information. It is possible that the informed consent form and the verbal presentation of the study could have been more detailed. It is suspected that some students refrained from participating due to expectations of interference. Therefore, it could have been beneficial to emphasize practical information that targeted such aspects, e.g. how the eye-tracking glasses work, the time required to perform the calibration, the exact design of the study, and the professional confidentiality concerning the students' performance. Instructions about how to participate in the study and the importance of keeping the teams constant throughout the data collection could also have been stated more clearly.

Expert ratings. The pilot study design assumed expert ratings would be collected from all scenarios. However, three of 14 participants were not assessed by an expert due to protocol errors and miscommunication among the researchers. Further, when providing a rating our impression was that the instructors were uncertain and uncomfortable in providing a rating, and appeared to use arbitrary evaluation criteria. At times they expressed that the performance was evaluated according to the scenario's level of difficulty, which complicates comparisons between different scenarios. Thus, it is likely that the inter-rater reliability would have been low if tested between multiple raters. It would therefore be prudent to develop an alternative approach for the main study.

An alternative to expert ratings could be to use objective simulator data to evaluate performance, e.g., speed and deviations from the optimal course. The attempt to use such data in the pilot study was unsuccessful due to lack of preparation of simulators and adapted scenarios. Based on what has been learned in this study, it appears likely that the scenario must be carefully designed to allow for collection of such data. Also, it may be beneficial to invite navigation experts and technical staff to participate in a discussion about the accuracy and contributions of simulator data.

Trackability. A surprising result was that the attained trackability rate (i.e. the percentage of valid recordings) did not meet the feasibility criterion. These results were unexpected as other studies have attained a considerably higher trackability rate than the attained 64.29% despite using a higher cut-off score to determine validity. For example, Isaacowitz et al. (2008) applied a cut-off score of 75% and attained a trackability rate of 85% for adults from 18-25 years of age (*M* = 19.72, *SD* = 1.82). Although not identical, this age distribution is comparable to the one found in our sample (*M* = 23.57, *SD* = 5.29). Possible explanations for this finding will be addressed in the following section.

First, Isaacowitz et al. (2008) eye-tracked 85 young adults compared to seven students (tracked twice) in the present pilot study. According to the law of large numbers large samples have a higher representative value compared to small samples (Tversky & Kahneman, 1971). That is, random occurrence of low trackability individuals in the sample could have significantly more impact on the results in a small-scale study compared to a large-scale study, and this factor may be one of the explanations for the observed differences between the studies. Second, the researchers had limited experience in performing the calibration and some challenges were encountered during this procedure. In order to complete the nine-point calibration, several participants required instructions to further open their eyes as recommended by the supplier (T. Strandvall, personal communication, April 19, 2013). However, these participants are unlikely to have held their eyes equally open during the time period that followed. A more stringent calibration or selection procedure would have been preferable, but would be difficult to combine with the educational goals of the course.

Another issue could be that the calibration was performed in the hallway outside the simulator. The lighting in the simulator is dimmed as the visibility of the navigation instruments is enhanced in faint lighting. Consequently, a considerable difference in luminosity between the hallway and the simulator could be observed. Prior to the pilot study, a test calibration was performed at both places of which the conclusion was that more problems were experienced during calibration in dim lighting (simulator) compared to calibration in normal lighting (hallway). The supplier recommended that the lighting be held constant during calibration and recording (T. Strandvall, personal communication, April 19, 2013). However, as the test trial indicated that calibration in the hallway was smoother, and the time to perform the calibration was limited, the participants were calibrated in the hallway. This issue would have been avoided if the luminosity in the simulator were increased. Also, it is possible that the dim lighting in the simulator did not allow for high-quality eye-tracking. To identify a compromise that might have ensured both trackability and the team's performance, experts on navigation could have been consulted. To summarize, it is likely that the tracking quality would have been far better if the experimenters had more training, the sample had undergone a thorough selection procedure, there was more time to perform the calibration, and the lighting in the simulator was better.

Resource indicators of feasibility

In order to map the resources needed to carry out the main study, two aspects of the preprocessing of eye-tracking data were assessed. These were the importance of experience in handling eye-tracking data and the time needed to perform the preprocessing when the analyst was experienced. Results showed that there was a significant difference between data resulting from the first and second run at assigning AoIs. This finding indicates that there was a learning effect of preprocessing the recordings. A change in strategy between the two runs might be partly accountable for this observation. Due to the increase in experience the use of the category that indicates more uncertainty was applied more frequently, and the AoI that was used in clear cases was placed more accurately over the face (i.e. included less redundant area). As the present study indicates that the level of experience is of importance to the data material that will be used in further analyses, it can be argued that resources should be assigned to ensure that the analyst holds sufficient competency in preprocessing eye-tracking data. Having experience with similar material is likely to increase the consistency of assigning AoIs in dubious cases, and may be important to ensure the instrument's reliability. However, the possibility that some of these differences might have been reduced if the lighting was better should be noted. It is likely that good visibility would have eased the assigning of AoIs, which in turn could have increased the intra-rater reliability.

The study further estimated the ratio between recording duration and time needed to define AoIs to 1:3, i.e. that for every minute of eye-tracking recording, three minutes of pre-processing would be necessary. It is estimated that required preprocessing time would be shorter with better lighting conditions. On the other hand, required preprocessing time would be longer with a more complex coding scheme and if there had been more movement in the field-of-view. Further, most of the recordings did not include events to be assigned AoIs, so the task for the analyst was to carefully look through the recording waiting for an event needing to be assigned an AoI. A task where AoIs were assigned more frequently would require more relative preprocessing time.

Knowledge of the relative time needed to preprocess the data will be valuable in resource planning for the main study. If a similar ratio can be expected, collecting 40 recordings that are 5 minutes long each will require 10 hours of preprocessing, while using 20 minutes long recordings will require 40 hours of preprocessing. The duration of the eye-tracking period would influence the efficiency of the data collection, as shorter recordings would allow more participants in a given time span.

Scenarios

In the current pilot study, scenarios were changed weekly and the scenario characteristics varied considerably due to the educational setting. Therefore, a quantitative assessment of the extent to which the changing scenarios might account for potential main effects detected in the main study would have been preferable. Unfortunately, the low sample size in the present study did not allow for performing statistical tests on subgroups representing the different scenarios. Nevertheless, some observations were made during the data collection. The different scenarios affected the team members' positioning on the simulator bridge and the distance between team members varied significantly. For instance, participants could be positioned close to one another when discussing the course using a map. Increased distance, however, could occur if the captain watched the instruments in the front of the simulator bridge while the second officer was at the helm in the back of the simulator bridge. The latter example is further illustrative of situations where the team members were positioned away from each other. It is likely that such factors will affect the extent to which the captain looks at the second officer's face, and if so, hypothesis testing in the main study will be affected by qualities of the scenarios. Due to the probable impact of scenarios on team functioning, scenarios should have been kept constant during the data collection.

Sample size

A sufficient sample size is essential to attain the statistical power necessary for rejecting the null hypothesis (Cohen, 1988). Low power leads to an increased risk of failing to detect an effect that exists in the population (i.e. a Type II error) and of overestimating the strength of a true effect (Button et al., 2013). Conventional practice is to use a statistical power of .80 and to set the probability of Type I error to 5 % (Cohen, 1992). These precautions should reduce the risk of making faulty assumptions about the meaning of the results. However, a statistically significant result does not indicate the strength of a detected effect. This aspect can be examined by calculating the effect size, which is a standardized measure of "the degree to which the phenomenon is present in the population" (Cohen, 1988, p. 9). Effect sizes can be expressed in terms of Pearson's r, and there are conventions for which strength of association is considered a small (r = .10), moderate (r = .30), or large effect (r = .50, Cohen, 1988). The present pilot study should attempt to determine the required sample size necessary for confidently rejecting the null hypotheses in the main study.

Based on the effect size of the eye-tracking results in the pilot study, one may calculate the required sample size for the main study. As the second run at preprocessing eye-tracking data was assessed most accurately, these data were applied in the current sample size calculation, using an online power and sample size calculator (HyLown Consulting LLC, 2013). Since both frequency and duration of dwells have been examined with disparate effect sizes, one calculation was made for each. The estimate based on duration data (baseline: M = 8.46, SD = 7.62; effect: M = 7.75, SD = 8.66) had a small effect and thus indicated a required sample size of 1171 participants in each experiment group. The frequency data (baseline: M = 23, SD = 19.16; effect: M = 10.75, SD = 13.5) had a large effect and thus indicated a need for 19 participants in each group. As there is no a posteriori reason to assume that one of these measures better reflects our outcome variable, it appears that pilot study effect sizes are an uncertain basis for estimating main study sample size. Eye-tracking values varied greatly within the sample, and this is probably the reason for the observed disparity in estimated sample size. This issue might have been avoided if the pilot study had used a matched-subjects between-subjects design as is planned for the main study. The inaccuracy associated with estimating sample size from pilot studies has been recognized by Thabane et al. (2010), who have suggested that such calculations might result in misleading recommendations of sample size.

Another approach to calculating sample size is to estimate an effect size for the main study based on the effects seen in comparable studies described in the literature. For instance, Feldman et al. (2008) found that there was a positive relationship between SMM and fixation similarity (r = .55). Using Cohen's (1988) conventions, this effect size can be characterized as large. Further, the main study design resembles the design used by Wieser, Pauli, Alpers, and Mühlberger (2009), who included one free viewing task (i.e. no instructions about gaze behavior) and one eye-contact task (i.e. instructions about gaze behavior) and found large effect sizes (r = .53, r = .58, for different gaze directions). Based on these two studies, it seems reasonable to expect a moderate to large effect size for the main study.

It has been suggested that the hypotheses for the main study could be tested by conducting simple regression analyses and a one-way ANOVA, of which the ANOVA will require the largest sample size. If the effect size can be expected to be moderate, an ANOVA with two groups will require a sample of 64 participants in each group (a total *N* of 128; Cohen, 1992). As indicated above, one may argue for expecting a large effect size, in which case one would require a sample size of 26 in each group (a total *N* of 52). The more conservative first estimate will be challenging to adhere to, considering the resource costs associated with the use of eye-tracking methodology. The fundamental problem is therefore whether or not a study should be performed when it is uncertain whether or not the sample size is sufficient to attain the power necessary to reliably reject the null hypothesis. Whereas conducting a study of which the sample size can be questioned could lead to a faulty denial of the existence of a phenomenon, refraining from carrying out the study could hinder the progression of the research field. A brief literature search for studies using eye-tracking equipment to experimentally examine the role of gaze in social interaction identified examples of eye-tracking studies with two groups using 30 (Horley, Williams, Gonsalvez, & Gordon, 2003; Klin, Jones, Schultz, Volkmar, & Cohen, 2002) and 40 (Garner, Mogg, & Bradley, 2006) participants. Wieser et al. (2009) included three groups, with a combined sample size of 56 participants. It thus seems as though comparable eye-tracking studies include about 15-20 participants in each experiment group. This is an estimate which is considered feasible in a main study, although the risk of performing a Type II error should be kept in mind when interpreting the results.

Alternatives to eye-tracking equipment

In the main study, eye-tracking is planned to be used as a means to examine the role of gaze in team functioning. However, as only one participant could be tracked at a time, this process proved to be time-consuming. This issue could have been solved by acquiring a second pair of eye-tracking glasses. Alternatively, more affordable methods than eye tracking glasses could be considered. Having access to two recording units would allow for assessment of several teams in parallel and in turn make the data collection more effective. Another possibility could be to assess both members of the same team. This procedure would allow for a wider scope of testable hypotheses, e.g., examining the extent to which both team members looking at the same object at the same time (i.e. at a given piece of equipment or at each other) would predict SMM.

An example of an alternative to eye-tracking equipment could be “consumer grade action cameras,” typically used in nonprofessional sports. These are light-weight cameras that can be head-mounted, and the price allows for purchase of several cameras. Behavioral data could then be quantified by using categories such as “head turned toward person A’s face,” “head turned toward the map,” and “head turned toward person A’s hand gestures.” However, based on our experiences in defining AoIs, it was evident that although the AoI is within a visual field, gaze can be focused at other stimuli. Hence, use of consumer grade action cameras is likely to decrease the reliability of the measure, as gaze directed toward neutral stimuli is likely to be coded as directed toward stimuli of interest if such stimuli are part of the visual field.

Assessment of feasibility

The feasibility of the main study was operationalized in terms of the extent to which the specified feasibility criteria were met in the pilot study. As less than half of the target values were attained, the main study cannot be considered feasible if conducted as originally planned. Several suggestions to specific improvements that might increase the feasibility of the main study have been proposed, such as introducing economic incentives and adjusting the lighting conditions during recording of eye movements. However, these suggestions may not address more fundamental underlying weaknesses of the design, and design alternatives will therefore be considered in the following section.

Alternatives to the main study design

The design weaknesses of the pilot study discussed above may have resulted from trying to achieve three aims in the same design: doing the study in an ecologically valid setting (a maritime simulator), maintaining experimental control (repeated measures to control for individual differences in gaze and simulator experience), and using an ecologically valid sample (maritime personnel in training). In order for the planned main study to avoid some of the observed difficulties it might be necessary to compromise on either ecological validity or control. Three approaches to increasing the feasibility of the main study will be discussed in this section.

One approach to testing the proposed hypotheses could be to perform the experiment in a laboratory setting that models some of the task aspects, although not being as similar to reality as a simulator. A laboratory setting would have the advantage of increasing the experimental control while retaining the matched-subjects between-subjects design. As a consequence, a range of the experienced issues could have been avoided, such as inconstant scenarios, variations in the duration of the planning phase, insufficient lighting, technical difficulties in running the simulator scenarios, and challenges in getting measures of performance. A laboratory experiment can be designed in several ways depending on whether the planning phase or the navigation phase is of most interest. To focus on the planning phase, the team could be asked to plan a course using a navigational chart, information about the vessel, and mock-ups for the navigation instruments. To focus on the navigation phase, one could use an existing or custom designed computer game about ship navigation (with less fidelity than an actual simulator) that requires coordination between two team members. Performance may be assessed by examining the extent to which the team's chosen course deviates from an optimal route predefined by experts. Such an approach would be more affordable and easier to coordinate. This approach would further have higher reliability than the outcome measures in the pilot study due to standardized scenarios, objective performance measures, and a consistent time span between treatment and measurement. However, the simplified setting may raise questions about the generalizability of laboratory findings to an actual ship's bridge.

A second approach would involve relinquishing some of the experimental control in exchange for more data points. In the present study each team was recorded twice to cancel out individual variations in gaze communication across conditions. Further, teams were surveyed six times during the data collection to register the variation in self-reported SMM and performance over time. If each team member only had to participate in one scenario recording and one questionnaire, this would release considerable resources by allowing for recording of twice as many teams in the same time. Such a between-subjects design would involve participants being randomized to the conditions, and treatment being given prior to the scenario recording. The lower resource requirements in such a design could increase the feasibility of the main study. However, reducing the ability to control for factors such as individual differences in gaze communication and time might weaken the internal validity of the findings. Without a baseline recording of eye-movements, counterbalancing of participants will not be possible. Findings from the pilot study showed that there were significant individual differences in gaze communication, and this aspect should be considered in the main study design due to its potential influence on validity. Further, by measuring self-reported SMM and performance only once, there will be a possibility that findings can be explained by differences in experience between teams. Consequently, random or systematic baseline differences may be confounded with the effect of training in gaze communication. Thus, for the current research question, multiple assessments of performance should be made in order to control for the impact of simulator experience.

A third approach is to retain the simulator setting and two time-points for eye-tracking recordings, but to perform the study in a setting where it is easier to retain control over structural factors. This procedure would allow for application of a matched-subjects between-subjects design. The basis for this suggestion is that most of the challenges to the feasibility of the main study (e.g., difficulties in recruiting participants, time pressure, dim lighting, different lengths of the planning phase, scenario inconsistencies) were due to the limitations of doing the data collection in an educational setting. Thus, these challenges are likely to persist if the main data collection is carried out during maritime training for experienced captains as planned. Further, it is not likely that a typical course length of four to five days would be sufficient to perform both baseline recording and posttest for all teams. For these reasons, it should be considered whether the main study could be carried out in a setting with more control of structural factors.

One possibility is to enter into a close collaboration with a maritime academy that has access to bridge simulators and technical staff to operate the simulators. As several problems related to lack of experimental control were experienced in the present study, it might be beneficial to motivate students to participate in the study outside classes (e.g., by considering a monetary reward). This approach has the advantage of not limiting the population to students of a particular course. However, the expenses involved in renting simulator time exclusively for research purposes would in most cases be prohibitive. An alternative option could be to cooperate with a naval academy where the data could be collected during the training of cadets with some maritime navigation experience. One may expect these participants to have a higher attendance rate and to be more willing to adapt their training program for research purposes. A dedicated data collection will to a greater extent let researchers control factors that might represent essential threats to the validity of the study (such as scenario inconstancy, time pressure, and different lengths of the planning phase). For instance, the scenario could be selected according to criteria such as task, level of difficulty, and length. The duration of the scenario would be important as a standard duration of recordings would make the performance comparable across teams. In the pilot data collection it was observed that students differed significantly in the time they used to perform the task, and it therefore appears most viable to specify a standard length for the recording (rather than waiting for the students to complete the task). Further, this flexibility would allow for shorter duration of recordings which would entail less time being used to record each team and fewer resources being required to preprocess eye-tracking data. To summarize, a dedicated data collection might reduce threats to validity, make the data collection more efficient, and increase the feasibility of the main study.

Conclusion

The present simulator study has assessed the feasibility of a main study aiming to examine the relationship between gaze communication, SMM, and team performance. Few studies have examined the link between these variables and a hypothetical model for their relationships has therefore been proposed. The aim of the study was to assess the feasibility of the study rather than to test the hypotheses, and eight feasibility criteria were specified. The results showed that only three of these target values were met, and the main study was assessed as not feasible if performed as originally planned. The criteria that were not met were related to recruitment, participation, performance ratings, and trackability. In addition to examining the specified feasibility criteria, the study further addressed the issue of resources needed to preprocess the collected eye-tracking data. The findings indicated that training in preprocessing eye-tracking data is necessary, and that considerable resources should be assigned to the preprocessing of eye-tracking data. This pilot study provided the opportunity to learn from experience and prevented the waste of considerable resources on a project that was unlikely to succeed. The knowledge gained from our experiences led to the suggestion of both specific and general improvements to the study design. Of the three presented approaches to improvement of the study design, the most viable option appears to be to perform the data collection outside an educational setting in order to increase experimental control. This study gives some insight into issues worthy of attention and thus contributes to a better foundation for conducting the main study. Hopefully this study may contribute to facilitating the use of this methodology to test the relationships between gaze communication, SMM, and team performance as well as in other fields of psychological research.

References

Adams, R. B., Gordon, H. L., Baird, A. A., Ambady, N., & Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. *Science, 300*(5625), 1536.

Alwall, N., Johansson, D., & Hansen, S. (2010). The gender difference in gaze-cueing: Associations with empathizing and systemizing. *Personality and Individual Differences, 49*(7), 729-732.

Baldauf, D., & Deubel, H. (2008). Properties of attentional selection during the preparation of sequential saccades. *Experimental Brain Research, 184*(3), 411-425.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge: MIT Press.

Batki, A., Baron-Cohen, S., Wheelwright, S., Connellan, J., & Ahluwalia, J. (2000). Is there an innate gaze module? Evidence from human neonates. *Infant Behavior & Development, 23*(2), 223-229.

Bayliss, A. P., Pellegrino, G., & Tipper, S. P. (2005). Sex differences in eye gaze and symbolic cueing of attention. *The Quarterly Journal of Experimental Psychology: 58*(4), 631-650.

Bayliss, A. P., & Tipper, S. P. (2006). Predicitive gaze cues and personality judgments: Should eye trust you? *Psychological Science, 17*(6), 514-520.

Bedwell, W. L., Wildman, J. L., DiazGranados, D., Salazar, M., Kramer, W. S., & Salas, E. (2012). Collaboration at work: An integrative multilevel conceptualization. *Human Resource Management Review, 22*(2), 128-145.

Berelson, B., & Steiner, G. A. (1964). *Human behavior: An inventory of scientific findings.* New York, NY: Harcourt, Brace & World.

Blignaut, P., & Wium, D. (2014). Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior Research Methods, 46*(1), 67-80.

Bowers, C. A., Braun, C. C., & Morgan, B. B., Jr. (1997). Team workload: Its meaning and measurement. In M. T. Brannick, E. Salas, & C. W. Prince (Eds.)., *Team performance assessment and measurements: Theory, methods, and applications* (pp. 85-108). Mahwah, NJ: Lawrence Erlbaum Associates.

Broadbent, D. E. (1958). *Perception and communication.* London, England: Pergamon Press.

Buljac-Samardzic, M., Dekker-van Doorn, C. M., van Wijngaarden, J. D. H., & van Wijk, K. P. (2010). Interventions to improve team effectiveness: A systematic review. *Health Policy, 94*(3), 183-195.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365-376.

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye-gaze. *Neuropsychologia, 40*(8), 1129-1138.

Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), *Individual and group decision making: Current issues* (pp. 221-246). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (*2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Cooke, N. J., Gorman, J. C., & Winner, J. L. (2007). Team cognition. In F. T. Durso, R. S. Nickerson, S. Dumais, S. Lewandowsky, & T. J. Perfect (Eds.), *Handbook of applied cognition* (2nd ed., pp. 239-268). West Sussex, England: Wiley.

Denzau, A. T., & North, D. C. (1994). Shared mental models: Ideologies and institutions. *Kyklos, 47*(1), 3-31.

Deubel, H. (2008). The time course of presaccadic attention shifts. *Psychological Research, 72*(6), 630-640.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*(1), 80-90.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze Perception triggers reflexive visuospatial orienting. *Visual Cognition, 6*(5), 509-540.

Edwards, B. D., Day, E. A., Arthur, W., Jr., & Bell, T. (2006). Relationships among team ability composition, team mental models, and team performance. *Journal of Applied Psychology, 91*(3), 727-736.

Edwards, P., Cooper, R., Roberts, I., & Frost, C. (2005). Meta-analysis of randomized trials of monetary incentives and response to mailed questionnaires. *Journal of Epidemiology and Community Health, 59*(11), 987-999.

Ellsworth, P. C., & Ludwig, L. M. (1972). Visual behavior in social interaction. *The Journal of Communication, 22*(4), 375-403.

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24*(6), 581-604.

European Maritime Safety Agency (2011). *Maritime Accident Review 2010.*

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences of the United States of America, 99*(14), 9602-9605.

Farroni, T., Menon, E., & Johnson, M. H. (2006). Factors influencing newborns' preference for faces with eye contact. *Journal of Experimental Child Psychology, 95*(4), 298-308.

Feldman, M., Lum, H. C., Sims, V. K., Smith-Jentsch, K., & Lagattuta, N. (2008). Do you see what I see? Eye tracking and shared mental models. In *Proceedings of the 52th*

Human Factors and Ergonomics Society Annual Meeting. September 22-26, New York, NY.

Flin, R., O'Connor, P., & Crichton, M. (2008). *Safety at the sharp end: A guide to non-technical skills.* Aldershot, England: Ashgate Publishing.

Flin, R., O'Connor, P., & Mearns, K. (2002). Crew resource management: Improving team work in high reliability industries. *Team Performance Management, 8*(3/4), 68-78.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5*(3), 490-495.

Friesen, C. K., Moore, C., & Kingstone, A. (2005). Does gaze direction really trigger a reflexive shift of spatial attention? *Brain and Cognition, 57*(1), 66-69.

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin, 133*(4), 694-724.

Frischen, A., & Tipper, S. P. (2004). Orienting attention via observed gaze shift evokes longer term inhibitory effects: Implications for social interactions, attention, and memory. *Journal of Experimental Psychology: General, 133*(4), 516-533.

Galea, S. & Tracy, M. (2007). Participation rates in epidemiologic studies. *Annals of Epidemiology, 17*(9), 643-653.

Garner, M., Mogg, K., & Bradley, B. P. (2006). Orienting and maintenance of gaze to facial expressions in social anxiety. *Journal of Abnormal Psychology, 115*(4), 760-770.

Grant, R. W., & Sugarman, J. (2004). Ethics in human subject research: Do incentives matter? *Journal of Medicine and Philosophy, 29*(6), 717-738.

Hains, S. M. J., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development, 67*(5), 1940-1951.

Hetherington, C., Flin, R., & Mearns, K. (2006). Safety in shipping: The human element. *Journal of Safety Research, 37*(4), 401-411.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measurements.* New York, NY: Oxford University Press.

Horley, K., Williams, L. M., Gonsalvez, C., & Gordon, E. (2003). Social phobics do not see eye to eye: A visual scanpath study of emotional expression processing. *Journal of Anxiety Disorders, 17*(1), 33-34.

Hutton, S. B. (2008). Cognitive control of saccadic eye movements. *Brain and Cognition, 68*(3), 327-340.

HyLown Consulting LLC (2013). *Calculate Sample Size Needed to Compare 2 Means: 2 Sample, 1-Sided*. Retrieved March 10, 2014, from http://powerandsamplesize.com.

Isaacowitz, D. M., Toner, K., Goren, D., & Wilson, H. R. (2008). Looking while unhappy: Mood-congruent gaze in young adults, positive gaze in older adults. *Psychological Science, 19*(9), 848-853.

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience and Biobehavioral Reviews, 33*(6), 843-863.

Johnson, T. E., Lee, Y., Lee, M., O'Connor, D. L., Khalil, M. K., & Huang, X. (2007). Measuring sharedness of team-related knowledge: Design and validation of a shared mental model instrument. *Human Resource Development international, 10*(4), 437-454.

Johnson, T. E., Top, E., & Yukselturk, E. (2011). Team shared mental model as a contributing factor to team performance and students' course satisfaction in blended courses. *Computers in Human Behavior, 27*(6), 2330-2338.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354.

Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science*, *12*(5), 176-180.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, *100*(1),78-100.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, *20*(2), 403-437.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*(9), 809-816.

Knapp, M. L., Hall, J. A., & Horgan, T. G. (2013). *Nonverbal communication in human interaction* (8th ed.). Wadsworth, MA: Cengage Learning

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D. H., & Gowda, S. M. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Biomedical Engineering*, *57*(11), 2635-2645.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*(13), 1897-1916).

Land, M. F. (2011). Oculomotor behaviour in vertebrates and invertebrates. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 3-16). Oxford, NY: Oxford University Press.

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences, 4*(2), 50-59.

Lim, B., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behaviour, 27*(4), 403-418.

Lipsitz, A., Kallmeyer, K., Ferguson, M., & Abas, A. (1989). Counting on blood donors: Increasing the impact of reminder calls. *Journal of Applied Social Psychology*, *19*(13), 1057-1067.

Littlejohn, S. W., & Foss, K. A. (2008). *Theories of human communication* (9th ed.). Belmont, CA: Thomas/Wadsworth.

MacMillan, J., Entin, E. E., & Serfaty, D. (2004). Communication overhead: The hidden cost of team cognition. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 61-82). Washington, DC: American Psychological Association.

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

Mason, M. F., Tatkow, E. P., & Macrae, C. N. (2005). The look of love: Gaze shifts and person perception. *Psychological Science*, *16*(3), 236-239.

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

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology, 94*(2), 535-546.

Mohammed, S., & Dumville, B. C. (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior, 22*(2), 89-106.

Mohammed, S., Ferzandi, L., & Hamilton (2010). Metaphore no more: A 15-year review of the team mental model construct. *Journal of Management*, *36*(4), 876-910.

Mohammed, S., Klimoski, R., & Rentsch, J. R. (2000). The measurement of team mental models: We have no shared schema. *Organizational Research Methods, 3*(2), 123-165.

O'Connor, P. (2011). An evaluation of the effectiveness of bridge resource management training. *International Journal of Aviation Psychology*, *21*(4), 357-374.

Otnes, B., Thorsen, L. R., & Vaage, O. F. (2011). *Levekår blant studenter 2010.* Oslo, Norway: Statistisk sentralbyrå.

Papenmeier, F., & Huff, M. (2010). DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animations and movies. *Behavior Research Methods, 42*(1), 179-187.

Patrashkova-Volzdoska, R. R., McComb, S. A., Green, S. G., & Compton, W. D. (2003). Examining a curvilinear relationship between communication frequency and team performance in cross-functional project teams. *IEEE Transactions on Engineering Management, 50*(3), 262-269.

Pfeiffer, U. J., Schilbach, L., Jording, M., Timmermans, B., Bente, G., & Vogeley, K. (2012). Eyes on the mind: Investigating the influence of gaze dynamics on the perception of others in real-time social interaction. *Frontiers in Psychology, 3,* 1-11.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526.

Rothblum, A. R. (2000, October). *Human error and marine safety.* Paper presented at the meeting of National Safety Council Congress and Expo, Orlando, FL.

Rouse, W. B., Cannon-Bowers, J. A., & Salas, E. (1992). The role of mental models in complex systems. *IEEE Transactions on Systems, Man and Cybernetics 22*(6), 1296-1308).

Salas, E., Burke, C. S., Bowers, C. A., & Wilson, K. A. (2001). Team training in the skies: Does crew resource management (CRM) training work? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(4), 641-674.

Salas, E., Burke, C. S., & Cannon-Bowers, J. A. (2000). Teamwork: Emerging principles. *International Journal of Management Reviews, 2*(4), 339-356.

Salas, E., & Cannon-Bowers, J. A. (1997). Methods, tools, and strategies for team training. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace* (pp. 249-279). Washington, DC: American Psychological Association.

Salas, E., DiazGranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A meta-analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(6), 903-933.

Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a “Big Five” in teamwork? *Small Group Research, 36*(5), 555-599.

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences, 15*(7), 319-326.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11), 1-29.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & Milanovich, D. M. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors, 41*(1), 61-71.

Symons, L. A., Hains, S. M. J., & Muir, D. W. (1998). Look at me: Five-month-old infants’ sensitivity to very small deviations in eye-gaze during social interactions. *Infant Behavior and Development, 21*(3), 531-536.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L., ... & Goldsmith, C. (2010). A tutorial on pilot studies: The what, why and how. *BMC Medical Research Methodology*, *10*(1), 1-10.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review, 76*(3), 282-299.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105-110.

Urban, J. M., Bowers, C. A., Monday, S. D., & Morgan, B. B., Jr. (1995). Workload, team structure, and communication in team performance. *Military Psychology, 7*(2), 123-139.

Whalen, P. J., Kagan, J., Cook, R. G., Davis, F. C., Kim, H., Polis, S., ... & Johnstone, T. (2004). Human amygdala responsivity to masked fearful eye whites. *Science*, *306*(5704), 2061.

Wieser, M. J., Pauli, P., Alpers, G. W., & Mühlberger, A. (2009). Is eye to eye contact really threatening and avoided in social anxiety? – An eye-tracking and psychophysiology study. *Journal of Anxiety Disorders*, *23*(1), 93-103.

Wyland, C. L., & Forgas, J. P. (2010). Here’s looking at you kid: Mood effects on processing eye gaze as a heuristic cue. *Social cognition*, *28*(1), 133-144.

Appendix: Questionnaire

BAKGRUNN

1. Hvilket team var du i under øvelsen?

   Team-nummer __________________ Har ikke fått team-nummer: ☐

2. Når var du kaptein i denne øvelsen? **Først** ☐$_1$ **Sist** ☐$_2$ **Var ikke kaptein** ☐$_3$

3. Kjønn: **Kvinne** ☐$_1$ **Mann** ☐$_2$ Alder (fyll inn): __________________

SAMARBEID OG KOMMUNIKASJON

De neste spørsmålene handler om hvordan du opplevde at samarbeidet om de ulike oppgavene var under øvelsen. Sett ett kryss for hvert utsagn på skalaen.

I teamet mitt…	Helt uenig	Delvis uenig	Nøy-tral	Delvis enig	Helt enig
4. … visste vi hvordan de ulike oppgavene i øvelsen hang sammen	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
5. … vurderte vi hvilke begrensninger vi hadde i å utføre øvelsen	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
6. … hadde vi ikke blitt enige om hva målet skulle være	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
7. … diskuterte vi ulike måter å gjennomføre øvelsen på	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
8. … visste vi hva som måtte gjøres for å gjennomføre øvelsen	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
9. … kommuniserte vi med hverandre når vi utførte øvelsen	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
10. … var vi enige om hvordan vi skulle kommunisere for å utføre øvelsen	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
11. … kommuniserte vi med en avslappet tone	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
12. … lyttet vi alltid til hverandre	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
13. … likte vi å jobbe med alle oppgavene som øvelsen innebar	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
14. … oppmuntret vi hverandre til å prestere best mulig	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
15. … var vi stolte over arbeidet vi gjorde	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
16. … opplevde vi det som viktig å nå målet	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
17. … sa vi ikke alltid det vi egentlig mente	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
18. … forstod vi hvordan vi skulle få tak i den informasjonen vi trengte	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
19. … forstod vi hvordan vi skulle jobbe sammen	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
20. … klarte vi å tilpasse oss til de ulike rollene i teamet	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
21. … forstod vi hvordan vi effektivt skulle utveksle informasjon	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
22. … løste vi problemene som oppstod	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$

I teamet mitt…	**Helt uenig**	**Delvis uenig**	**Nøy-tral**	**Delvis enig**	**Helt enig**
23. … var det et trygt miljø der vi åpent diskuterte problemene i øvelsen	☐1	☐2	☐3	☐4	☐5
24. … var det god stemning mellom oss	☐1	☐2	☐3	☐4	☐5
25. … hadde vi tilstrekkelig erfaring til å gjennomføre øvelsen	☐1	☐2	☐3	☐4	☐5
26. … kjente vi til hverandres begrensninger for samarbeidet	☐1	☐2	☐3	☐4	☐5

De neste spørsmålene handler om hvordan du opplevde kommunikasjonen under øvelsen

	Helt uenig	**Delvis uenig**	**Nøy-tral**	**Delvis enig**	**Helt enig**
27. Det var god kommunikasjon mellom meg og samarbeidspartneren min	☐1	☐2	☐3	☐4	☐5
28. Samarbeidspartneren min forstod hva jeg mente	☐1	☐2	☐3	☐4	☐5
29. Vi snakket også om ting som ikke var relevante for å løse oppgaven	☐1	☐2	☐3	☐4	☐5
30. Samarbeidspartneren min fulgte med på hvordan jeg håndterte oppgavene	☐1	☐2	☐3	☐4	☐5
31. Samarbeidspartneren min kommuniserte det samme med kroppsspråk som med ord	☐1	☐2	☐3	☐4	☐5
32. Samarbeidspartneren min gav ikke tilbakemelding på hvordan jeg håndterte oppgavene	☐1	☐2	☐3	☐4	☐5
33. Vi var samkjørte når vi løste oppgavene	☐1	☐2	☐3	☐4	☐5
34. Når viktige beskjeder ble gitt, svarte vi ved å si hvordan vi hadde forstått beskjeden	☐1	☐2	☐3	☐4	☐5

ULYKKESRISIKO OG ULYKKER

De neste spørsmålene handler om ulykker. Svar som om simulatorøvelsen hadde vært en reell seilas.

	Liten sannsynlighet						**Stor sannsynlighet**
35. Hvor stor sannsynlighet var det for at noe av det dere gjorde i øvelsen kunne ha ført til en personskade?	☐1	☐2	☐3	☐4	☐5	☐6	☐7
36. Hvor stor sannsynlighet var det for at noe av det dere gjorde i øvelsen kunne ha ført til en ulykke med materiellskade?	☐1	☐2	☐3	☐4	☐5	☐6	☐7

	Ja	**Nei**
37. Var fartøyet involvert i en ulykke?	☐1	☐2

Takk for at du tok deg tid til å fylle ut skjemaet!