Exploring the potential of video technologies for collaboration in emergency medical care. Part II: Task performance

Hanna M. Söderholm, Diane H. Sonnenwald*
Swedish School of Library and Information Science
Göteborg University & University College of Borås, SE-50190 Borås, Sweden
E-mail: {hanna.maurin, diane.sonnenwald}@hb.se
Telephone: +46 (0)33 435 59 89
Fax: +46 (0)33 435 40 05

James E. Manning, Bruce Cairns
School of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA
E-mail: {james_e_manning, bruce_cairns}@med.unc.edu
Telephone: +1 191 966 4131
Fax: +1 919 966 3049

Greg Welch, Henry Fuchs
Department of Computer Science, University of North Carolina, Chapel Hill, NC 27599, USA
E-mail: {welch, fuchs}@cs.unc.edu
Telephone: +1 191 962 1700
Fax: +1 919 962 1799

*Author to which all correspondence should be sent.

Exploring the potential of video technologies for collaboration in emergency medical care. Part II: Task performance

Abstract

We conducted an experiment with a posttest, between-.subjects design to evaluate the potential of emerging 3D telepresence technology to support collaboration in emergency healthcare. 3D telepresence technology has the potential to provide richer visual information than current 2D video conferencing techniques. This may be of benefit in diagnosing and treating patients in emergency situations where specialized medical expertise is not locally available. The experimental design and results concerning information behavior are presented in the paper, "Exploring the potential of video technologies for collaboration in emergency medical care. Part I: Information sharing" (Sonnenwald et al). In this paper we explore paramedics' task performance during the experiment as they diagnosed and treated a trauma victim while working alone, or in collaboration with a physician via 2D videoconferencing or via a 3D proxy. Analysis of paramedics' task performance shows that paramedics working with a physician via a 3D proxy performed the fewest harmful interventions and showed the least variation in task performance time. Paramedics in the 3D proxy condition also reported the highest levels of self-efficacy. Interview data confirm these statistical results. Overall the results indicate that 3D telepresence technology has potential to improve paramedics' performance of complex medical tasks and improve emergency trauma health care if designed and implemented appropriately.

1. INTRODUCTION

This paper builds on the research presented in the paper "Exploring the potential of video technologies for collaboration in emergency medical care. Part I: Information sharing" (Sonnenwald et al., submitted) by further examining the potential of video technologies on task performance during collaboration in emergency medical care. As discussed in Part I, trauma is a significant health problem (Meyer, 1998; Coates & Goode, 2001). Rural areas are often disadvantaged when it comes to high quality emergency medical care. Trauma injuries are more common in rural areas, and rural trauma patients face worse outcomes and significantly higher death rates than urban patients (AHCPR, 1996; Wilhide, 2002). Unavailability of health care resources, longer transportation times and lack of transportation, and lack of training opportunities for emergency medical services (EMS) personnel found in rural areas contribute to the lower health care outcomes (AHCPR, 1996; Wilhide, 2002; Rowley, 2001). Using video technologies to provide emergency health care services may help improve trauma patient outcomes in rural areas. However before investing in such technologies we should evaluate their effectiveness (AHCPR, 1996).

An important measure of technology effectiveness is task performance. If medical task performance, i.e., patient care, is not improved with the introduction of a new technology there may be no or little reason to deploy the technology. We explored medical task performance under three conditions: paramedics working alone, paramedics working in consultation with emergency medicine physicians via state-of-the-art 2D videoconferencing, and paramedics working in consultation with emergency medicine physicians via a 3D telepresence proxy. In the 3D proxy condition, the paramedic and physician were collocated but the physician could not touch anyone or anything in the room. In each condition paramedics were asked to manage a difficult airway, including diagnosing the condition and performing a cricothyrotomy, as described in “Part I: Information sharing”. Twenty paramedics participated in each condition and we videorecorded each of their sessions to enable us to measure and compare task performance across conditions.

There are few standard measurements for evaluating complex medical tasks, including managing a difficult airway. We developed three task performance measures based on a review of medical training literature and working in consultation with emergency medical physicians.

We also measured and compared future task performance indicators across conditions using the construct of self-efficacy. Self-efficacy is the perceived capability to perform a specific task (Bandura, 1997). It is a powerful determinant of how well a task will be performed in the future; i.e., it predicts future task performance. We measured self-efficacy through a postquestionnaire that each paramedic completed after participating in a simulated emergency medical scenario. In addition we conducted an interview with each paramedic after completion of the simulated emergency medical scenario. During the postinterviews paramedics reflected on their current and future task performance.

The results indicate that paramedics working with a physician in the 3D proxy condition performed fewer harmful interventions and showed less variation in task performance times compared to the paramedics in the 2D and Alone conditions. Furthermore, the paramedics in the 3D proxy condition reported statistically significant higher levels of self-efficacy. These statistical results are validated and further explained through our analysis of postinterview data. These results in combination with results from “Part I: Information sharing” provide a rich and ultimately consistent understanding of the potential of 3D telepresence technology.

2. PREVIOUS RESEARCH

Previous research on telemedicine, computer-supported cooperative work, and 3D telepresence technology were reviewed in “Part I: Information sharing”. To complement that literature review, we discuss previous research on task performance and self-efficacy.

2.1 Task Performance

Task performance refers to “the detailed examination of observable activity or behavior associated with the execution or completion of a required function or unit of work” (MeSH, 2008). Why measure task performance? The overall goal of emergency healthcare is to provide trauma, or accident, victims and other patients with medical care in order to limit the extent and minimize the effects of possible life-threatening injuries and illnesses. For emergency medical services (EMS) strive to provide this care as quickly as possible at the scene of an accident or in an ambulance while transporting the patient to a hospital or other medical facility. Thus, the key rationale for developing, installing and using a new technology in emergency medicine lies in its potential to effectively improve patient care through improved medical task performance.

The concepts of task and task performance are used to evaluate technology within several fields including human computer interaction (HCI) and computer supported cooperative work (CSCW). For example, Tan et al (2004) explore the effect of display size on the performance of 3D navigation tasks; Fussell and colleagues (2003) examine how multiple views of a remote scene contributes to the performance of a physical task; while Heldal et al (2006) compares collaborative performance of tasks vs. single work in real life settings and distributed virtual environments.

In information science several additional related concepts exist. For example, the concept of system task performance is used to evaluate the performance and effectiveness of IR systems (e.g., Borlund, 2000). The concept, task, is also used to investigate information seeking, needs and use behavior (Vakkari, 2003), primarily to explore *perceived* task difficulty or users’ perceived performance or performance satisfaction (e.g., Choo et al. 2006). Increasingly information behavior, including seeking, sharing and using information, is an integral component of tasks. That is, there are an increasing number of tasks in professional domains, including the emergency healthcare service profession, that can not be successfully completed without dynamically seeking, sharing and using information. For these tasks measuring task performance is an effective way to measure the outcomes of information behavior, complementing measures that rely on self-reports. Yet within information science few efforts, with one exception being Sonnenwald, Whitton and Maglaughlin (2003), have been made to measure results or outcomes of human task performance. Järvelin and Ingwersen (2004) identify this as a critical limitation of current information science research, and they suggest future research should include task performance and outcome measures.

One challenge when developing a task performance measure is that it often needs to be very specific and contextualized as well as valid and reliable. This requires expert input and/or good working knowledge of the task context, and makes the re-use of previously developed measures not feasible or at best limited unless identical tasks are being performed. In addition, measures used in previous research have often either not been validated or details regarding the measures' validity are not provided.

Although management of a difficult airway, the task in our evaluation, is taught to paramedics and medical students throughout the world, a standard task performance measure, or performance assessment protocol, does not exist in the published literature. To develop measures, we researched medical education literature (e.g., Ali et al, 1996; Chappman et al, 1996; Johnson et al, 1999) and previous research on performance assessment in simulated medical scenarios (e.g., Byrne & Greaves, 2001; Devitt et al, 1998; Gaba et al, 2001; Stringer, Bajenov & Yentis, 2002). The most frequently mentioned performance measures include combinations of: technical, cognitive and behavioral skills; number of appropriate/inappropriate procedures; and time for problem solving and decision making. These measures are often based on a recognized treatment algorithm, or in compliance with principles determined by medical researchers as good practice. Our main challenge was to construct an assessment protocol that could measure task performance comprehensively, based on the information which was possible to accurately observe from the video recordings of each session. Our initial grading protocol was based on the ASA Practice Guidelines for Management of the Difficult Airway, developed by American Society for Anesthesiologists (2003). This protocol was then reviewed in-depth and revised by two emergency medical physicians who have decades of experience in managing difficult airways and performing cricrothyrotomies.

2.2 Self-efficacy

The social cognitive theory of self-efficacy refers to a person's judgment of their capability to perform a certain task (Bandura, 1997). Social cognitive theory fundamentally proposes that people are products as well as producers of their own environment and social systems. The concept of self-efficacy takes this idea further, proposing that people have selfbeliefs that let them exercise control over their thoughts, beliefs and feelings which, in turn, effects how they behave. Self-efficacy affects a person's behavior such that people with high self-efficacy try harder when facing a difficult task and are less sensitive to failures and performance obstacles than people with low self-efficacy. Low levels of self-efficacy lead to more stress, slower recovery after failures, and higher likeliness to give up when facing problems. The primary, most influential source of self-efficacy is previous task performance experiences. However, self-efficacy is also influenced by vicarious experience, i.e., observing other people' experiences and the consequences of these experiences, as well as performance feedback and verbal advice.

Self-efficacy has been used in a wide variety of studies. It has been used to explain affective and cognitive aspects of individuals' database searching strategies and interactions with Internet (Nahl, 2005; Nahl & Tenopir, 1996). Kim (2008) uses the concept of self-efficacy to investigate the effects of users' emotional control on their search performance. In a recent study about information management Choo and colleagues (2006) use the concepts of self-efficacy and task performance to measure outcomes of information use.

Self-efficacy is considered a major determinant of future task performance, i.e., task performance creates beliefs in ability which influence future task performance. There have been many studies investigating the validity of self-efficacy as a predictive measure. For example, Downey (2006) compares and validates computer self-efficacy instruments with respect to predictive power on future performance, attitude and usage. Bolman et al (2007) found a strong positive correlation between levels of course completion and self-efficacy beliefs. Phan (2007) found that self-efficacy directly predicted reflective thinking (i.e., the learning stage where previous experiences and behaviors are connected to the solving of complex problems and future action) when investigating learning approaches and academic performance.

Future task performance is affected by self-efficacy partly through learning (Bandura, 1997). Self-efficacy contributes to learning in the sense of skill development and skill retention as it stimulates certain behaviors. High levels of self-efficacy leads to more efforts to solve a challenging problem, and thus additional skills, while low levels of self-efficacy reduce and/or slow down skill development. Because experiences, especially successful ones, are the strongest source of influence on self-efficacy (Bandura, 1997) a paramedic's perceptions of self-efficacy after diagnosing and managing a difficult airway in a simulated medical scenario can help predict the paramedic's future performance in diagnosing and managing a difficult airway. That is, the theory of self-efficacy suggests that paramedics with higher levels of self-efficacy after diagnosing and managing a difficult airway will actually perform the same tasks better in the future.

2.3 Research Hypotheses

We propose that collaboration across geographic distances between a paramedic and physician using 3D telepresence technology may improve a paramedic's ability to diagnose and treat trauma victims now and in the future. Furthermore, because the goal of 3D telepresence technology is to allow remote physicians to see a 3D representation of the accident scene and be able to dynamically change their viewpoint of that scene in real-time (Welch et al, 2005), allowing the physician to quickly view different parts of the victim's body and to see intricate details in depth, we propose that paramedics' task performance now and in the future will be greater when 3D telepresence technology is used than when 2D videoconferencing technology is used. Thus, our hypotheses are:

H1: Paramedics working in consultation with a physician via 3D telepresence technology will provide better medical care to trauma victims than paramedics working in collaboration via 2D videoconferencing or paramedics working alone.

H2: Paramedics working in consultation with a physician via 3D telepresence technology will report higher levels of self-efficacy than paramedics working in collaboration via 2D videoconferencing or paramedics working alone.

3. Research Design

3.1 Research Approach

To investigate task performance and self-efficacy we used the same research design as presented in the paper "Part I: Information sharing". That is, we used an experiment with a posttest, between-subjects design. We simulated an emergency medical situation using a sophisticated computerized mannequin, a METI Human patient simulation. The mannequin was programmed to suffer from a difficult airway and performing a cricothyrotomy was required to in order to save the mannequin. Paramedics were asked to diagnose and treat the mannequin under one of three conditions: working alone, working in consultation with a physician via state-of-the-art 2D videoconferencing, and working in consultation with a physician via a 3D proxy. The protocol for the first condition, working alone, was identical to the other two conditions except that the paramedic was asked to diagnose and treat the mannequin without any assistance from a physician.

There were 20 participants per condition. The 60 participants, 48 males and 12 females, were all certified paramedics working in southeastern U.S. The average years of total EMS work experience of the sixty paramedics was 11 years, with a range of 1 to 26 years. Their average paramedic work experience was 7 years, with a range of 1 to 24 years. The paramedics were randomly assigned across conditions, with equal distribution of gender and years of experience across all three conditions. Of all participants, 14 persons had previously performed a cricothyrotomy on a real patient, 7 in the Alone, 4 in the 2D, and 3 paramedics in the 3D proxy condition. In the 2D and 3D proxy conditions 5 and 6 paramedics respectively had met the consulting physician before, a situation that mirrors daily work of paramedics. Paramedics usually have one or two specific hospitals where they take patients on a regular basis and where they know some of the physicians in the emergency room (ER.) At other times, patients need to be transported to other hospitals or facilities where the paramedic does not know or has not met the physicians before.

Each session was videorecorded using four cameras that captured paramedic's actions on and surrounding the mannequin and the mannequin's medical output status monitor (heart rate, oxygen saturation levels). Furthermore, after each session the paramedics completed a postquestionnaire to report their perceptions of self-efficacy, and participated in an open-ended post-interview that was audiorecorded and transcribed.

3.2 Task performance measures

As discussed earlier task performance measures were developed using the ASA Practice Guidelines for Management of the Difficult Airway developed by American Society for Anesthesiologists (ASA) (2003) and the expertise of two emergency medical physicians who have decades of experience in managing difficult airways and performing cricothyrotomies. The result was three task performance measures, and associated grading protocol, that captured *subtask execution*, *harmful interventions,* and *timing of key events*.

Subtask execution includes recording if certain tasks in the patient treatment process are executed or not. Manual mask ventilation and intubation attempts occur repeatedly throughout the treatment process, and are thus better represented for interpretation in a quantified form rather than being recorded in the timing format. Performing a cricothyrotomy is a binary variable simply representing if the paramedic did this procedure or not. *Harmful interventions* are treatments or procedures that according to standard practice (ASA, 2003) are not indicated to perform by the medical situation. These interventions could be potentially harmful for the patient by **increasing risks** for complications or unwanted side-effects and by **delaying the time** it takes until the victim's breathing is restored. *Timing of events* consists of recording the time (mm:ss) for the most important events and steps in managing and treating a difficult airway. Definitions of each event are provided in Table 1.

Table 1. Summary of task performance measures

Aspect	Item name	Definition	Outcome score
Subtasks	Manual mask ventilation	If the paramedic manually ventilates the patient or not	Yes/No
	Intubation attempts	Each time the paramedic uses the laryngoscope AND puts a tube into the patient's mouth to get it into the airway counts as one time.	Number
	Cricothyrotomy	Performing a cricothyrotomy	Yes/No
Harmful interventions	Nasal intubation	If the paramedic tries to intubate the victim through the nasal passage	Yes/No
	Chest decompression	If the paramedic tries to insert a needle in victim's chest to relieve a suspected accumulation of air in thoracic space	Yes/No
	Locate membrane	If the paramedic feels the throat for locating the cricothyroid membrane before cutting	Yes/No
	Incorrect incision	If incision is done incorrectly, either in the wrong location and/or made too large	Yes/No
	Tube slips out	If the tube slips out of the incision after bag valve mask is attached and has to be reinserted	Yes/No
Timing of events	Breathing stops	When the patient's chest stops moving	Time (mm:ss)
	Decision to perform cricothyrotomy	When the paramedic takes out the cricothyrotomy equipment from the EMS bag	Time (mm:ss)
	Airway access–incision begins	When the paramedic begins to make the incision into the airway either with a needle or scalpel, i.e., when the needle or scalpel touches the skin.	Time (mm:ss)
	Surgical access	Only if needle was used initially: When the paramedic begins to make the surgical incision into the airway, i.e., when the scalpel touches the skin.	Time (mm:ss)
	Tube insertion	When a tube is fully inserted through the incision into the airway (and stays there so attaching the bag valve mask is possible)	Time (mm:ss)
	Patient stable	When the patient's chest movements start again AND the O2 blood saturation reaches 90% or higher.	Time (mm:ss)

After the expert physicians approved the grading protocol, two researchers graded a subset of sessions independently using the protocol. They compared their results, refining the definitions used in the protocol. This procedure was repeated and inter-coder reliability reached 98%, well above standard accepted levels (Robson, 2002). The protocol was then reviewed by the expert physicians to ensure the definitions used were correct. A researcher then graded the remaining sessions. After this grading was done, a second researcher independently graded a random selection of graded sessions to further ensure the grading was accurate.

3.3 Self-efficacy measure

Self-efficacy beliefs vary in *magnitude*, *strength* and *generality*. Magnitude refers to the steps of increasing difficulty in a chain of behaviors a person believes herself capable of performing. Our self-efficacy instrument (a postquestionnaire) was designed to examine different types of skills related to managing a difficult airway, starting with the easiest subtasks and ending with the hardest, most demanding.

Strength of self-efficacy refers to how much a person believes that they can perform a certain task. This dimension is often associated with factors such as persistence in the face of frustration, pain, faulty equipment and other factors affecting performance. On the postquestionnaire, paramedics indicated the strength - how confident they are that they can perform each subtask – on a 7 point Likert scale. Generality concerns what effect success or failure experiences have on self-efficacy in a specific situation. This also implies whether self-efficacy altered by previous experiences is applicable or extended to other similar behaviors or contexts (Maddux, 1995), and raises the question of the appropriate measurement level for self-efficacy beliefs. Bandura (2001) recommends that self-efficacy questions at the most detailed measurement level should be related to a very specific performance under a specific set of conditions. On the next level, self-efficacy measures should be related to a group of performances within the same activity domain under a group of conditions sharing common properties.

Since we did not find a self-efficacy questionnaire for difficult airway management or for performing a cricothyrotomy, we developed questions based on Bandura's recommendations (2001) and in consultation with emergency medical physicians. Following these suggestions the postquestionnaire items refer to participants' self-efficacy to manage a difficult airway, ranging from basic airway management tasks, such as manual mask ventilation (putting a breathing mask on a victim), to more complex tasks, such as securing the opening made in the cricothyrotomy membrane when performing a cricothyrotomy. The questions also referred to both physical and cognitive tasks, such as making an incision in the throat and deciding on alternative treatment strategies. Thus the self-efficacy questionnaire also provides insight into cognitive task performance that cannot be measured through observation. According to Bandura (1997), the most precise relationship between self-efficacy beliefs and task performance is obtained by measuring these as close together in time as possible. Thus, the postquestionnaire was given to the participants immediately after completion of the simulated scenario.

3.4 Paramedics' reflections on their performance

As mentioned in the paper *"*Part I: Information sharing" the paramedics participated in a semi-structured postinterview after their session. To gain insights regarding their perceptions of their task performance we asked paramedics to identify two things about their performance during the session they were most satisfied with, and what two things they were least satisfied with. We also asked if there was anything they wish they had done differently during the session. These questions encouraged paramedics to reflect on their task performance, and share those reflections with us. At no point during the postinterviews did we, the interviewers, criticize a paramedic's performance. As we explained to the paramedics, our role was to learn from them. Some paramedics mentioned more than two things they were most or least satisfied with; others mentioned only one.

The interviews were audiorecorded and subsequently transcribed. Two researchers independently coded a subset of the interviews, initially identifying passages in which paramedics reflected on their task performance. Inter-coder reliability was verified, and one researcher coded all remaining interviews to identify passages where paramedics discussed their performance. A subsequent round of coding was undertaken by two researchers to specifically identify themes emerging from the passages regarding performance. The researchers initially coded the passages independently, and then compared their results checking intercoder reliability. The intercoder reliability was 85% which is above accepted levels (Robson, 2002). However we also discussed and resolved differences achieving full agreement between the two coders.

3.5 Limitations

As discussed in "Part I: Information sharing", our research design has high levels of task, ecological and common ground validity. However one limitation was the use of video-recordings of sessions to evaluate task performance. Although three video cameras were used and the cameras were placed optimally for the simulated scenario, it was occasionally difficult to see details of a paramedic's actions in the video because the paramedic or their medical equipment obscured the view. In fact it is this very problem that prompts the vision of 3D telepresence technology. For example, sometimes it was difficult to see and judge the exact quality of an incision. It was not possible to grade the quality of each incision, with a scale of 1 (very bad) to 7 (very good), due to lack of clarity and details in the videos. Our task performance protocol only evaluated actions that were clearly observable on all videos. An alternative approach would be for an observer to evaluate each session in real time. However this approach also has at least three disadvantages. It would be obvious to paramedics that the observer was grading them and this could be a distraction as well as make a paramedic more nervous than necessary. The observer may get in the way of the physician in the 3D proxy condition because both the observer and physician would be competing for a good view in the limited space around the mannequin's throat. There would be no way to check for inter-coder reliability, i.e., we would have to rely on one observer's observations for all 60 sessions unless there were two observers grading each session in real-time. The latter however would have been very impractical due to space and resource limitations.

One limitation of the self-efficacy measure is also the post-test design. To investigate potential impact of technology on self-efficacy, it would have been optimal to measure self-efficacy both before and after a session. However, as self-efficacy is so task-specific it would not be possible to ask participants to complete a self-efficacy questionnaire before the session without identifying the specific medical tasks they are to perform during the session. This would have affected the paramedics' task performance during the scenario, in the sense of preparing them for what procedure to expect and thus lose the critical aspect of real-time decision-making and problem solving during the session. One way to overcome this problem could have been by administering a self-efficacy questionnaire before the scenario that covered several other medical scenarios as well as our airway scenario. However, this - together with the overall data collection procedures (simulated scenario, introduction, simulator training, questionnaires and interviews) - would impose an additional cognitive and time burden for the participants and hence was not feasible. Another limitation related to task performance and self-efficacy measures is the use of a simulated scenario and manikin. Bandura (1997) discusses performance constraints in terms of people possessing skills needed to manage a certain task but have no incentives to act (possibly because it is not a real patient). This can affect task performance and the reported level of self-efficacy. However, our validity measures did not show any significant differences between the conditions with respect to levels of paramedics' engagement or perceived realism of the scenario.

Another limitation is that we have not analyzed the video-recordings from the 2D and 3D proxy conditions to further investigate paramedic – physician communication. An analysis of the verbal and visual communication could increase our understanding of the collaboration process. We leave this analysis for future work, and focus on task performance outcomes and perceptions in this paper.

4. RESULTS

4.1 Task performance

4.1.1 Subtask performance

Table 2. Subtask performance

Condition	No Manual Mask Ventilation	Intubation Attempts*		No Cricothyrotomy Performed
		Mean	Range	
Alone	1	2.50	0-19	3
2D	0	2.95	0-7	0
3D Proxy	1	1.85	0-6	0

** Number of times each paramedic attempted to intubate the patient*

Table 2 shows frequency, mean and range of subtask performance, i.e., manual mask ventilation, intubation attempts made per session and cricothyrotomy performance, in each condition. One paramedic in the Alone condition and one in the 3D proxy condition did not manually ventilate the patient, and the fewest number of intubation attempts were performed in the 3D proxy condition. However, a Fisher's exact test[1] showed no statistically significant differences across the conditions for manual mask ventilation. An ANOVA with a least significant difference (LSD) post-hoc test comparing the mean number of intubation attempts also showed no statistically significant differences across conditions. The standard accepted protocol for managing a difficult airway includes performing each of these subtasks several times before performing a cricothyrotomy. The lack of statistically significant results appears to indicate that paramedics followed this standard protocol irrespective of the condition.

Three paramedics in the Alone condition did not attempt to perform a cricothyrotomy. This result is not statistically significant. In our simulation, as in many real life situations, the trauma victim dies when a cricothyrotomy is not performed when it is needed. Recall that 7 paramedics in the Alone condition had previously performed a cricothyrotomy

[1] To test for frequency differences in subtask execution and harmful interventions between the three conditions, a chi-square test would be a reasonable choice. However, one limitation of the chi-square test is when there are frequencies lower than 5 in any of the table cells, the test results will be hard to interpret or give inaccurate p-values. We therefore performed Fisher's exact tests to compare subtask and harmful interventions between conditions.

on a real patient, compared to 4 in the 2D condition and 3 in the 3D proxy condition. Thus previous experience does not appear to compensate for the lack of a remote consulting physician.

During postinterviews no paramedic expressed high levels of satisfaction in performing manual mask ventilation, and only one expressed dissatisfaction. Manual mask ventilation is a common task that paramedics perform frequently, and thus in the simulated scenario it most likely was not particularly challenging or satisfying to perform. In comparison, cricothyrotomies are not commonly performed (although all paramedics are trained to perform them), and across all conditions paramedics expressed great satisfaction in completing a cricothyrotomy. However the participants reflected on this performance in different ways. For example, most participants in the Alone and 2D conditions discussed their satisfaction in performing a cricothyrotomy with some hesitation, sometimes framing their performance in terms of avoiding the worst possible outcome, patient death. Participants said:

> *I guess… [I'm most satisfied with] the actual performance with the cric. I think that went very smoothly. I mean, he didn't die.* Participant in the Alone condition
>
> *Well, I thought the cric went ok… I needed to do something.* Participant in the 2D condition
>
> *Getting the cric…I would say that's the thing I'm most proud of…I guess I didn't make him [the mannequin] any worse…the simulation didn't progress to a point where he actually lost a pulse and did all these other things.* Participant in the 2D condition

In comparison, participants in the 3D proxy condition did not appear to hesitate when expressing satisfaction completing the cricothyrotomy. In addition they reported great satisfaction in their collaboration with the physician. Participants explained:

> *Getting the surgical cric [done]. I liked that. I've never done one…on a patient. I have to do one [in training] every year. But if that's the way it's like, I'll be cutting more of them now.*
>
> *The most important thing I'm satisfied with is being able to recognize the difficult airway…The other is probably using the physician's input to work through that difficult airway as a team, and getting it done.*
>
> *I'm most satisfied with actually doing [the cric], and talking with the doctor and understanding what he was talking about, I'm very satisfied with that. I've never dealt with a scalpel or anything like that, and listening to him and knowing what he's saying, it came very easy to me.*

As shown in Table 2, paramedics in the Alone and 2D conditions, on average, made one more unsuccessful intubation attempt than paramedics in the 3D proxy group (prior to considering alternative airway intervention/cricothyrotomy). Similarly, dissatisfaction with their intubation subtask performance was most frequently mentioned by paramedics in the Alone and 2D conditions as the subtask performance they were least satisfied with. Four paramedics in the Alone condition and five in the 2D condition commented they attempted intubation too many times and/or should have recognized intubation was not a solution sooner:

> *I knew I was having difficulty [intubating] and I knew the airway wasn't opening. And after about the third time, I should have gone straight for the cric instead of trying it that fourth time.* Participant in the Alone condition

> *I wasn't getting the patient intubated. He needed an airway. His pulsox was dropping. In a real life scenario that patient would have had brain damage or death from this problem. Unfortunately the doctor had to advise me to do the [cricothyrotomy].* Participant in the 2D condition

In comparison only one paramedic in the 3D condition mentioned dissatisfaction with the intubation subtask:

> *Oh, the intubation! That was a terror…I tried [to intubate] twice but I really shouldn't have.*

4.1.2 Harmful Interventions

As shown in Table 3 two paramedics, one paramedic in the Alone and one in the 3D proxy condition, attempted to perform a nasal intubation. These results are not statistically significant using a Fisher exact test, although performing a nasal intubation in real life can lead to serious medical complications in our scenario where the patient had indications of a head injury. Four paramedics in the Alone condition but none in the 2D or 3D proxy conditions attempted to perform a chest decompression. In our scenario there were no indications for attempting a chest decompression. Furthermore, doing it in real life can lead to medical complications for the patient, and it is unnecessary in the sense that it takes up additional time when the patient is without oxygen. Locating the cricothyroid membrane is important in order to know where to make the incision into the throat. An improper incision can prolong the time until oxygen reaches the lungs or in the worst case, damage vocal cords or an artery. As illustrated in Table 3, three paramedics in the alone condition and one in the 2D condition did not locate the membrane before making an incision. However, all paramedics in the 3D proxy condition located the membrane. These differences while important from a patient's perspective are not statistically significant.

Table 3. Number of Harmful Interventions Performed

Harmful Intervention	**Condition**		
	Alone	**2D**	**3D Proxy**
Nasal intubation	1	0	1
Chest decompression	4	0	0
Not locating the cricothyroid membrane	3	1	0
Improper incision	3	1	0
Airway tube slippage	0	4	1
Totals	*11*	*6*	*2*

Three paramedics working alone and one in the 2D condition performed improper incisions, either cutting in the wrong location and/or making the incision too large. This result is not statistically significant and it mirrors the result regarding locating the cricothyroid membrane. As shown in Table 3, all paramedics who performed a cricothyrotomy in the Alone condition inserted the airway tube correctly without subsequent slippage. In the 2D condition, the airway slipped out in 4 instances, and once in the 3D proxy condition. A Fisher exact test shows the differences approach standard statistically significant levels (p = 0.073) between the Alone and 2D conditions.

In sum the highest number of harmful interventions occurred when paramedics worked alone (Table 3). A total of 11 harmful inventions occurred when paramedics worked alone, and 6 occurred when a paramedic worked in consultation with a physician via 2D videoconferencing technology. In comparison only two harmful interventions occurred in the 3D proxy condition. This result is overall statistically significant (p = 0.021) using a Fisher exact test. However, between conditions the occurrences of harmful interventions are significant between the Alone and 3D proxy condition (p = 0.003) and approach significance between the Alone and 2D condition (p = 0.053). A power analysis showed that approximately 40 additional participants per condition would be required to obtain statistical significance between the 2D and 3D proxy conditions for overall effect size (i.e., each condition's impact on performance of harmful interventions) of 0.7. For overall power of 0.8 we would need approximately 60 additional participants per condition. Effect sizes of 0.7 or 0.8 (and above) are considered large, while 0.5 and 0.2 are considered medium and small respectively (Cohen, 1988).

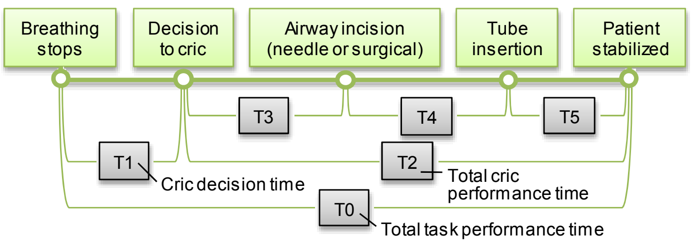
About half of those paramedics in each condition who performed harmful interventions identified their mistakes as what they liked least about their performance or something they wish that had done differently. Their comments mirror the statistical results regarding harmful interventions. For example, a participant in the Alone condition commented he was most dissatisfied with:

recognizing he was not a candidate for oral or nasal intubation. I didn't recognize it quickly enough.

As expected no one mentioned they were satisfied with performing a harmful intervention.

4.1.3 Task execution times

Figure 1. Task execution time calculation



In order to avoid damage to the brain it is crucial to minimize the time when the patient is without oxygen, in addition to perform all difficult airway treatment subtasks and avoiding harmful interventions. Thus, the timing of events (Figure 1) were analyzed by first calculating times for total task performance (T0), time before the paramedic made the decision to perform the cricothyrotomy (T1), and time for the cricothyrotomy procedure (T2) which, if performed correctly, should provide the patients with oxygen again.

Figure 2. Total task performance (T0)

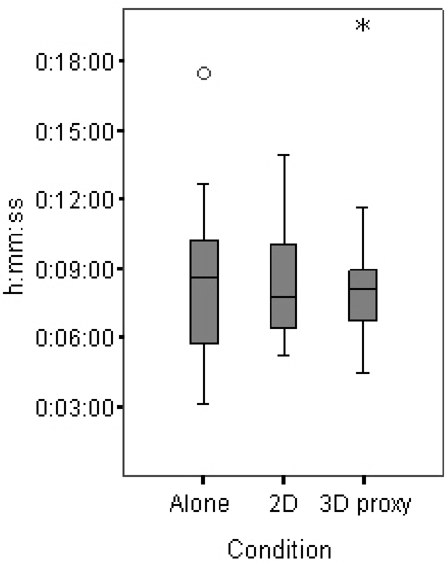


Figure 3. Time to cricothyrotomy decision (T1)

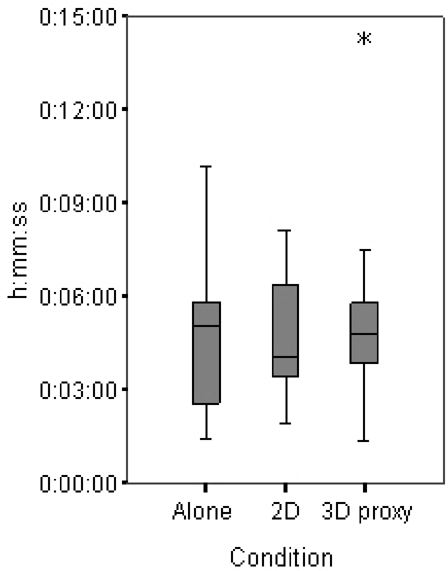


Figure 4. Total cricothyrotomy performance time (T2)

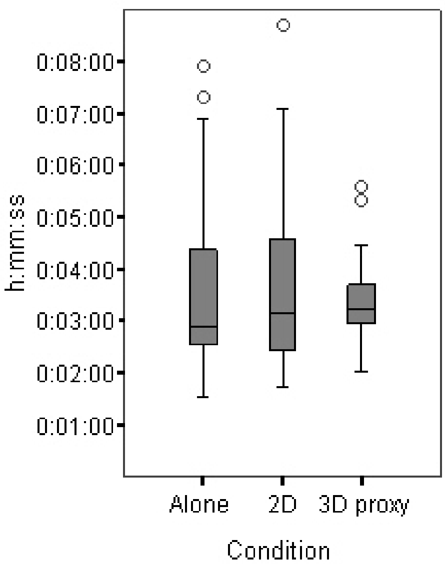
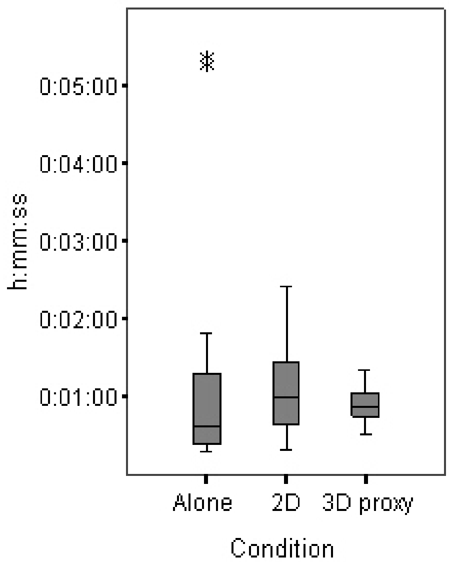


Figure 5. Time from cricothyrotomy decision to airway incision (T3)



The T0, T1 and T2 times also reflects decision making ability – when to take the step to actually perform a cricothyrotomy. The technical skills – such as finding the right place for the incision, inserting the breathing tube through the incision into the airway is reflected in T3 and T4. All these steps have to be done as quickly as possible in order to minimize the time when the patient is without oxygen, although when performed incorrectly could have fatal consequences for the patient.

An ANOVA test did not show any statistical significant differences between average task performance times due to condition. However, we observed that for all tasks (except for T4 "airway incision – tube insertion") the 3D proxy condition seemed to show less variation than the other conditions. A Levene test for equality of variance (Levene, 1960), showed that differences in variance are statistically significant for the total cricothyrotomy performance time (Figure 4, $p = 0.022$) and the time it took from the paramedic made the decision to perform a cricothyrotomy until the airway incision was made (Figure 5, $p = 0.001$). That is, the variation in these times was statistically significantly less for the 3D proxy condition.

We also analyzed the influence of years of paramedic experience and years of total EMS experience. Total EMS experience is defined as years of experience as paramedic plus the years of experience as basic and intermediate emergency medical technician, which require lower levels of education and typically precede paramedic certification.

Table 4. Correlations between Years of Professional Experience (YPE) and Task Execution Times

Task execution times	**YPE**	**Condition**		
		Alone	**2D**	**3D Proxy**
T0 Total task performance	Paramedic	-.633**	—	—
	Total EMS	-.519*	-.590**	—
T1 Breathing stops – cricothyrotomy decision	Paramedic	—	—	—
	Total EMS	—	-.549*	—
T2 Total cricothyrotomy performance	Paramedic	-.777**	—	—
	Total EMS	-.483*	—	—
T3 Cricothyrotomy decision – airway incision	Paramedic	—	—	—
	Total EMS	-.510*	-.534*	—
T4 Airway incision- tube insertion	Paramedic	—	—	—
	Total EMS	—	-.489*	-.453*

*** Correlation is statistically significant at $p \leq 0.01$*
** Correlation is statistically significant at $p \leq 0.05$*

The results from a Spearman's rho correlation analysis (Table 4) indicate that years of paramedic and/or total EMS work experience had a statistically significant impact on three task performance times in the Alone condition, while years of total EMS experience had a statistically significant impact on four task performance times in the 2D condition. In comparison, total EMS experience had a statistically significant impact on only one task performance time in the 3D proxy condition. These correlations are all negative, indicating that the fewer the years of professional experience, the longer it took to perform the task. This suggests that any negative impact on task execution times due to fewer years of professional experience may decrease or disappear with the use of 3D telepresence technology.

The most frequently mentioned timing event in postinterviews was T1. This task performance time was discussed both positively and negatively by participants across all conditions. Positive comments included the following:

> *I think [I'm most satisfied with] my … recognizing that the patient needed to have a surgical cric quickly.* Participant in the Alone condition
>
> *I was like "oh, my God! I can't get past the tongue! Let's move on." … I knew immediately that I needed to cric the patient.* Participant in the Alone condition
>
> *I thought deciding that I had to do a cric was important. Recognizing that I wasn't going to get him intubated…Going ahead because he was without oxygen and I needed to do something.* Participant in the 2D condition
>
> *What I'm most satisfied with… [was] realizing, boom, stop here. Let's move on. Let's try something different. This ain't working; let's go to plan B [to get the airway in.]* Participant in the 3D condition

Negative comments from participants in all conditions regarding T1 task performance time focused on taking too long to recognize the difficult airway, being too slow in deciding to do a cricothyrotomy, and/or hesitating to do it. For example participants reported:

> *I'm disappointed in the amount of time it took me to make the decision 'Okay, this is a difficult airway. This is a clamped patient from the head injury. It just doesn't look like what you have seen before, just move on and make a decision, do it.* Participant in the Alone condition
>
> *Unfortunately the doctor had to advise me to do the [cric] … I didn't make that leap myself and I should have made that leap myself.* Participant in the 2D condition
>
> *probably could have moved on to the needle cric or the cric or surgical cric quicker. That's probably my hesitation because I've never done it before….That's the only [dissatisfying] thing that I can think of is not moving to [the cric] quickly enough.* Participant in 3D proxy condition

Interestingly participants in the Alone condition discussed their dissatisfaction with all of the task performance times. That is, at least one participant and in many cases several participants in the Alone condition were not satisfied with their times for T0, T1, T2, T3, T4 and T5. Only participants working alone expressed such broad dissatisfaction. They reported:

> T0: *I didn't do it fast enough.*
>
> T1: *I guess my transition to the cric was the weakest part of my performance.*
>
> T2: *I was a little disappointed in me as far as being able to place the cric in quicker. I wanted to do it quicker than I did.*
>
> T3: *[I'm dissatisfied with] not getting the landmarks correct first time.*
>
> T4: *And the difficulty in… getting the airway after [starting the] incision, I would have liked to have done that in…less then a minute, but I am sure I was there for a couple of minutes picking.*
>
> T5: *I guess [I'm least satisfied with] the assessment portion afterwards*

The interview data, showing paramedics in the Alone condition had the largest variety of negative assessments related to task performance times mirror our quantitative findings that paramedics in the Alone condition had the largest variation in their actual task execution times.

4.2 Future Task Performance

4.2.1 Self-efficacy

There are two categories of self-efficacy items, referring to: basic airway management tasks preceding the cricothyrotomy (all subtasks occurring in time T1); and, cricothyrotomy tasks (all subtasks occurring in time comprising T2). Basic airway management tasks are tasks performed frequently by paramedics to ensure their patients are getting oxygen into their lungs. Cricothyrotomy tasks are performed less frequently, i.e., when basic airway management is not sufficient.

The mean for each question per condition are shown in Table 5. To determine if there are statistically significant differences between the conditions we performed an ANOVA[2], using least significant difference (LSD) and Games-

[2] To compare self-efficacy across the three groups, we use an ANOVA test instead of a nonparametric test. The nonparametric counterpart of the ANOVA test does not include any *post-hoc* procedures for detecting where or what the differences between the multiple groups are, that is, in variability, means or median. In our case it was important to find where any differences between the conditions are located, i.e. if there were differences between the Alone-2D or 2D-3D proxy conditions. Furthermore, our data meet the most important requirement for doing an ANOVA in that the sample sizes are equal. The ANOVA test is considered robust, even for data that doesn't meet the traditional/standard requirements (e.g. Bryman & Cramer, 2005; Vaughan, 2001).

Howell post-hoc tests (for equal and not equal variances respectively) to reduce the likelihood of Type 1 errors that can emerge when performing tests between multiple groups.

Table 5. Self-efficacy

	Question (response scale 1, strongly disagree, to 7, strongly agree)	**Mean** **Alone**	**2D**	**3D Proxy**
Basic Airway Management	*I can quickly…*			
	- diagnose a difficult airway	*5.95*	*5.75*	*6.15*
	- manually ventilate a patient	***6.80****	***6.35****	*6.55*
	- observe problems with manual mask ventilation	*6.30*	*6.00*	*6.35*
	- decide on alternative strategy when manual mask ventilation is unsuccessful	*6.20*	*5.75*	*6.11*
	- perform initial intubation	*6.30*	*6.00*	*6.11*
	- observe intubation problems	***6.55****	***6.10****	*6.35*
	- decide on alternative strategy when intubation is unsuccessful	***6.30***[a]	***5.75***[a]	*6.10*
Cricothyrotomy	*I can quickly…*			
	- decide when to perform a surgical cricothyrotomy	*5.60*	*5.25*	*5.95*
	- find the location of the cricothyroid membrane (CTM)	*6.05*	***5.55****	***6.35****
	- palpate the CTM	*6.25*	***5.80****	***6.50****
	- secure the opening made in the CTM	*6.05*	***5.63****	***6.35****
	I am confident that I can do a surgical cricothyrotomy	*5.70*	***5.35***[a]	***6.16***[a]

** Difference between groups is statistically significant at $p \leq 0.05$.*
[a] Difference between groups is statistically significant $p \leq 0.06$.

The results are somewhat surprising. For all basic airway management tasks paramedics in the 2D condition reported *lower* levels of self-efficacy than paramedics working alone or in the 3D proxy condition. For one task, manually ventilating patients, the differences between paramedics working alone and in the 2D condition are statistically significant ($p \leq 0.05$). For two other tasks, observing intubation problems and deciding on an alternative strategy when intubation fails, the differences are statistically significant at $p \leq 0.05$ and approach significance level $p \leq 0.06$. For all cricothyrotomy tasks, paramedics in the 2D condition again reported the lowest levels of self-efficacy. Paramedics in the 3D proxy condition reported the highest levels of self-efficacy. The differences in self-efficacy between the 2D and 3D proxy conditions are statistically significant at the $p \leq 0.05$ level for tasks related to the detailed skills of performing a cricothyrotomy. The paramedics in the 3D proxy condition reported higher levels of overall self-efficacy for performing a surgical cricothyrotomy compared to the paramedics in the 2D condition. This difference approach standard statistically significance at $p \leq 0.06$. No significant differences were found between the Alone and 3D Proxy conditions.

Table 6. Correlations between years of professional experience (YPE) and self-efficacy

	Question (response scale 1, strongly disagree, to 7, strongly agree)	**YPE**	**Condition** **Alone**	**2D**	**3D Proxy**
Basic Airway Management	*I can quickly…*				
	- diagnose a difficult airway	Paramedic	.638**	–	–
		Total EMS	.524*	–	–
	- manually ventilate a patient	Paramedic	–	–	–
		Total EMS	–	–	–
	- observe problems with manual mask ventilation	Paramedic	–	.518*	–
		Total EMS	–	.572**	–
	- decide on alternative strategy when manual mask ventilation is unsuccessful	Paramedic	.783*	–	–
		Total EMS	.480*	–	–
	- perform initial intubation	Paramedic	–	.511*	–
		Total EMS	–	.556*	–
	- observe intubation problems	Paramedic	.458*	–	–
		Total EMS	–	–	–
	- decide on alternative strategy when intubation is unsuccessful	Paramedic	.674**	–	–
		Total EMS	.538*	–	–
Cricothyrotomy	*I can quickly…*				
	- decide when to perform a surgical cricothyrotomy	Paramedic	–	.546*	–
		Total EMS	–	.547*	–
	- find the location of the cricothyroid membrane (CTM)	Paramedic	–	.506*	–
		Total EMS	–	.552*	–
	- palpate the CTM	Paramedic	–	.542*	–
		Total EMS	–	.629**	–
	- secure the opening made in the CTM	Paramedic	–	.525*	–
		Total EMS	–	.608**	–
	I am confident that I can do a surgical cricothyrotomy	Paramedic	–	.654**	–
		Total EMS	–	.676**	–

** Correlation is statistically significant at $p \leq 0.05$.*
*** Correlation is statistically significant at $p \leq 0.001$.*

Positive correlations between self-efficacy and years of professional experience were found using Spearman's rho test in the Alone and 2D conditions (Table 6). In the Alone condition, 4 self-efficacy questions were influenced by years of previous work experience. For the 2D condition, perceived ability to perform airway management tasks was influenced by years of professional experience for all cricothyrotomy subtasks. That is, the less work experience the paramedics in these two conditions had, the lower they rated their ability to treat similar patients in the future. All correlations in the Alone condition were related to basic airway management tasks, while the 2D condition had this connection for all of the cricothyrotomy tasks and some basic airway management tasks. There were no correlations between paramedic work experience and self-efficacy in the 3D proxy condition.

4.2.2 Paramedics' reflections regarding future task performance

After the simulation paramedics in all conditions reported increased self-efficacy, i.e., an increased capability to perform a cricothyrotomy.

> *I feel like it's been a learning experience for me.* Participant from the Alone condition
>
> *It's fun, it's a learning experience. It's pretty neat to be involved in something like this.* Participant from the 2D condition
>
> *I learned more today in 20 minutes than I learned in that 36 hour class.* Participant from the 3D proxy condition
>
> *With the doctor walking me through it, it was very simple. I could go in and do it now, no problem.* Participant from the 3D proxy condition

Although paramedics in both the 2D and 3D proxy conditions thought they would have been able to perform a cricothyrotomy if the physician had not been available, the physician helped to make them feel more comfortable and increase their confidence. Confidence and feeling comfortable when performing a task create high levels of self-efficacy.

In both 2D and 3D proxy conditions paramedics talked about feeling more **comfortable** and **confident** after performing the cricothyrotomy while having a physician available. For example, when asked if they felt they would have performed the same tasks the same way without the physician, several replied:

> *Yeah that's part of the question. I feel like, yes, I would have done the same thing, because the choices of treatment were mine. I didn't ask him [the physician] what to do, but I think he gave me more confidence, or would give me more confidence if I was unsure of something to be able to ask questions while doing it, with someone who's done it before. But I don't think it would have changed my treatment, but it may have assured its effectiveness.* Participant in the 2D condition
>
> *It [having the physician to consult with] was very helpful in terms of helping to establish my confidence level during the procedure.* Participant in the 2D condition
>
> *If I was in my truck, I'd say yes. I wouldn't have felt as good about it, and I wouldn't have learnt quite as much about it. But you always wonder if feeling comfortable adds to the care that you're giving your patient? And I think in a lot of cases it does. I think I could have done adequately without him there but I don't think it would have been quite as good because I felt comfortable, and since I felt comfortable I didn't waste time.* Participant in the 3D proxy condition
>
> *Him talking me through that procedure, even though I kinda' know the basic way to do, it was extremely helpful. And I think I can take that and use it…I feel a little bit more comfortable with it now…. So I'm thinking from a training aspect, I think it makes you a more competent provider.* Participant in the 3D proxy condition

Table 7. Phrases used by paramedics reflecting lower levels of self-efficacy

	Condition		
Phrases	**Alone**	**2D**	**3D Proxy**
Feeling clumsy, nervous, scattered, distracted	1	3	5
Frustrated, less confident, insecure about decisions	2	3	0
"not as in my best form", "I had a problem being myself", "not reacting as usual"	2	3	0
Total	*5*	*9*	*5*

If paramedics across all conditions reported increased levels of self-efficacy, why are the self-efficacy scores from the post-questionnaire statistically lower for paramedics participating in the 2D condition? As discussed earlier individuals with low levels of self-efficacy tend to believe that tasks are tougher than they really are, and this creates stress and a narrow vision of how best to go about the task. In the postinterviews, almost twice as many paramedics in the 2D condition as in the other conditions reported feelings of nervousness, frustration, insecurity or feeling "out of place" (Table 7). For example participants from the 2D condition said:

> *The frustrating thing was…establishing an airway. I knew the patient needed an airway…It was frustrating because I was trying to operate a lot of equipment…It was also frustrating because I couldn't get the patient intubated. He needed an airway. His pulsox was dropping. In a real-life scenario that patient would have had brain damage or [would have] died.*
>
> *I don't think I was as confident as I'd like to be…I couldn't remember my protocol or a standard procedure whether you cut vertically first or horizontally instead.*
>
> *I wasn't my…best. Not at all. I wasn't that impressed myself.*
>
> *I didn't feel like I was performing smoothly...not really going with the flow. Not knowing what to do from the beginning.*

The high frequency of negative feelings reported help to verify the postquestionnaire self-efficacy results.

5. DISCUSSION

The results show some support for Hypothesis H1, paramedics working in collaboration with a physician via 3D telepresence technology will provide better medical care to trauma victims. The results illustrate that paramedics working in consultation with a physician via a 3D telepresence technology proxy tend to provide better medical care to trauma victims than paramedics working in consultation via 2D video or paramedics working alone. Fewer subtask errors and harmful interventions were performed in the 3D proxy condition than in the Alone or 2D condition. Three paramedics in the alone condition did not perform a cricothyrotomy, although a cricothyrotomy was required to save the patient. A total of eleven harmful interventions were performed in the Alone condition, and six were performed when paramedics consulted with a physician via 2D video. In comparison only two harmful interventions were performed when the consultation occurred via the 3D proxy.

Although no statistically significant differences with respect to task performance times across conditions emerged from the data analysis, the results from the Levene test for equality of variance indicate 3D telepresence technology may reduce variation overall for the total cricothyrotomy performance time and specifically for the 'cricothyrotomy decision – airway incision' task performance time. Furthermore, only one (out of five) task performance time in the 3D proxy condition was influenced by years of professional experience. In comparison, three and four task performance times were influenced by years of professional experience when paramedics worked alone or in consultation via 2D videoconferencing, respectively. Recall that paramedics were assigned to conditions based on their years of professional experience, such that there was an equal distribution of years of professional experience across all three conditions. It appears that the 3D technology may reduce differences in diagnosis and treatment caused by differences in years of professional experience, with paramedics with fewer years of experience providing care closer to the level or those with more years of experience. From a patient's perspective and managerial perspective this is an important consideration. In emergency situations patients cannot choose which paramedics will treat them, and patients of course want the most experienced and knowledgeable paramedic to treat them. Similarly emergency healthcare service organizations want to provide the highest level of care possible, however, there is always employee turnover with more experienced paramedics retiring and pursuing other career opportunities. The use of 3D telepresence technology could help reduce any negative impact that the lack of experience plays in providing emergency healthcare.

The statistical results were reflected in comments made by paramedics during postinterviews. Paramedics collaborating with a physician via the 3D proxy expressed unequivocal satisfaction in their cricothyrotomy task performance. In comparison paramedics working alone or collaborating with a physician via state-of-the-art 2D video-conferencing technology tended to express their satisfaction hesitantly or tentatively.

The results further show support for Hypothesis H2, paramedics working in collaboration with a physician via a 3D telepresence technology will report higher levels of self-efficacy. As discussed earlier perceptions regarding self-efficacy predict and influence future task performance (Bandura, 1997). Paramedics collaborating via the 3D proxy reported the highest levels of self-efficacy in postquestionnaires. In comparison paramedics collaborating via 2D videoconferencing reported the lowest levels of self-efficacy. Furthermore the less work experience paramedics in the alone and 2D conditions had, the lower they rated their ability to treat similar patients in the future, whereas work experience had no impact at all on feelings of self-efficacy for paramedics in the 3D proxy condition. This suggests that the 3D telepresence technology may have a positive impact on future task performance, irrespective of a paramedic's years of professional experience.

It is interesting to note that the paramedics collaborating via 2D video-conferencing reported the lowest levels of self-efficacy – even lower than the paramedics working alone. This result was mirrored in postinterviews. Almost twice as many paramedics collaborating via 2D videoconferencing than those working alone or collaborating via the 3D proxy expressed feelings using words and phrases that reflect lower levels of self-efficacy. Furthermore, previous work experience influenced their levels of self-efficacy to greater extent in the 2D than in the Alone or 3D proxy conditions, with less experienced paramedics rating their ability lower than paramedics with more experience did. An important area for future research is investigating whether the use of 2D video-conferencing for emergency medical care may actually harm paramedics' future task performance, in particular for less experienced paramedics performing complex medical tasks.

The results regarding information sharing and task performance presented in 'Exploring the potential of video technologies for collaboration in emergency medical care. Part I: Information sharing' and in this paper combine to provide a rich and ultimately consistent evaluation regarding the potential of 3D telepresence in emergency healthcare. The usefulness of the information provided by the physician appears to be greater via the 3D proxy than via 2D video-conferencing, and the results presented in this paper suggest that current and future task performance may be enhanced with 3D telepresence technology.

We saw several features of our 3D proxy frequently utilized during the 3D proxy sessions. For example, physicians used the laser pointer to identify the location and size of the required incision, and to point to specific pieces of medical equipment that the paramedic needed to use. The paramedics paid attention to the physician's pointing. As Clarke (1996) discusses the ability to point to physical objects facilitates mutual understanding and task completion. A second feature used was the ability to dynamically change views. We saw physicians changing their viewpoint – bending down to get a side-angle view, standing up on tiptoe and bending over the victim and paramedic's hands. The physicians did not need to ask paramedics to move so the physician could see the victim better. Similarly the paramedic was free to focus on the medical task at hand, and did not need to worry about the physician's view.

Another requirement discussed by the paramedics is transparency, i.e., the 3D telepresence technology should be as transparent as possible in the sense that it needs to be immediate and require essentially no set-up time. When managing a difficult airway the paramedic has limited time to treat the patient and needs to focus exclusively on the patient. Therefore, any manual operation of the technology, such as changing the physician's views or zooming would not be feasible because the activity could take too much time away from patient care.

However, paramedics' performance outcomes and perceptions of the technology is only part of the story. Design requirements for a complex technology, such as 3D telepresence technology, that has the potential to impact many professionals and individuals, and which requires substantial changes in our technology and social infrastructures, need to take into account as many stakeholders as possible during the design process. Meeting the needs of multiple stakeholders will increase the likelihood that the technology will be successfully adopted and used. For example, while technology transparency and ease of use might important from a paramedic's perspective, a hospital administrator might not be willing to invest in a system that does not meet federal and state regulations for protecting patient privacy. An insurance company may not be willing to reimburse for emergency healthcare services utilizing a new technology if the new costs do not provide quantifiable benefits with respect to patient care. To identify these types of requirements we are conducting interviews with multiple stakeholders in the U.S. healthcare system, including emergency room (ER) physicians and nurses, IT support staff, ER department administrators, and Medicaid administrators. Our goal is to identify multiple stakeholders' needs and constraints from a broad contextual perspective to provide a more comprehensive set of design requirements than is possible from the data reported in this paper.

In summary, today paramedics at the scene of an accident collaborate with physicians via radio or cell phone which provide no visual support. Paramedics are required to verbally 'paint the picture' of the patient and accident scene to the remote physician in these complex and stressful situations where the paramedic's need for information is time-critical and where incorrect decisions based on this information may have fatal consequences. This is particularly crucial in rural areas where the retention of highly skilled and experienced paramedics is problematic.

Our study illustrates that providing physicians with rich, dynamic visual information of the emergency situation may lead to a more effective collaboration between the physician and paramedic and ultimately better patient care. Our results illustrate that state-of-the-art 2D videoconferencing does not appear sufficiently flexible to allow a physician to establish and maintain situation awareness of the dynamic and stressful remote emergency healthcare situation effectively. The physician must still ask the paramedic to provide detailed information regarding the patient and the paramedic's actions, that is, the paramedic must still 'paint the detailed picture' for the physician.

Today's state-of-the-art 2D videoconferencing technology may not provide sufficiently rich and dynamic visual information to provide better emergency healthcare in the long-term. 3D telepresence technology shows potential to improve patient healthcare in complex emergency medical situations.

ACKNOWLEDGMENTS

Our sincere thanks to: the study participants; Jim Mahaney for his expert technical assistance; the team developing the 3D technology; Chris Wiesen and Steve Poltrock for their helpful suggestions regarding statistical analysis; Ann-Sofie Axelsson for comments on an earlier draft; and the anonymous reviewers. Some task performance statistics were reported in an earlier conference paper: Söderholm, H.M., Sonnenwald, D.H., Cairns, B., Manning, J.E., Welch,G., & Fuchs, H. (2007). The potential impact of 3D telepresence technology on task performance in emergency trauma care. *Proceedings of the ACM Group '07 Conference.* NY: ACM Press. This research is supported by the National Library of Medicine; contract N01-LM-3-3514, 3D Telepresence for Medical Consultation: Extending Medical Expertise Throughout, Between and Beyond Hospitals.

REFERENCES

AHCPR. (1996). Improving health care for rural populations: Research in Action Fact Sheet. AHCPR Publication No. 96-P040. Rockville, MD: Agency for Health Care Policy and Research. Retrieved January 18, 2008 from http://www.ahrq.gov/research/rural.htm

Ali, J., Cohen, R., Adam, R., Gana, T., Pierre, I., Bedaysie, H., Ali, E., West, U., & Winn, J. (1996). Teaching effectiveness of the advanced trauma life support program as demonstrated by an objective structured clinical examination for practicing physicians. *World J Surg.*, *20*(8), 1121-1126.

American Society of Anesthesiologists Task Force on Difficult Airway Management. (2003). Practice guidelines for management of the difficult airway. *Anesthesiology*, *98*, 1269-1277.

Bandura, A. (1997). *Self-Efficacy: The Exercise of Control*. NY: H. Freeman.

Bandura, A. (2001). *Guide for constructing self-efficacy scales. (Rev.)* Available from Frank Pajares, Emory University.

Bolman, C. Tattersall, W. Waterink, J. Janssen, B. van den Berg, R. van Es, R. Koper (2007) Learners' evaluation of a navigation support tool in distance education. *Journal of Computer Assisted Learning, 23*(5), 384–392.

Borlund, P. (2000). Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation, 56*, 71-90.

Byrne, A., & Greaves, J. (2001). Assessment instruments used during anaesthetic simulation: Review of published studies. *Br. J. Anaesth,*, *86*, 445-450.

Chapman, D., Rhee, K., Marx, J., Honigman, B., Panacek, E., Martinez, D., Brofeldt, B., & Cavanaugh, S. (1996). Open thoracotomy procedural competency. *Ann. Emerg. Med, 28,*(6), 641-647.

Choo, C.W., Furness, C., Paquette, S., van den Berg, H., Detlor, B., Bergeron, P., & Heaton, L. (2006). Working with information: information management and culture in a professional services organization**.** *Journal of Information Science, 32*, 491- 510.

Clark, H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

Coates, T., & Goode, A. (2001). Towards improving prehospital trauma care. *Lancet, 357*(9274), 2070.

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

Devitt, J-H., Kurrek, M-M., Cohen, M-M., Fish, K., Fish, P., Noel, A-G., & Szalai, J-P. (1998). Testing internal consistency and construct validity during evaluation of performance in a patient simulator. *Anaesth-Analg.*, *86*, 1160-4.

Downey, J. (2006). Measuring general computer self-efficacy: The surprising comparison of three instruments in predicting performance, attitudes, and usage. In *Proceedings of the 39th Annual Hawaii international Conference on System Sciences (HICSS), Volume 08.* IEEE Computer Society, Washington, DC

Fussell, S. R., Setlock, L. D., & Kraut, R. E. (2003). Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In *Proc. CHI '03 Conf*erence (pp. 513-520). NY: ACM Press.

Gaba, D., Howard, S., Fish, K., Smith, B., & Sowb, Y. (2001). Simulation-based training in anesthesia crisis resource management. S*imulation & Gaming*, *32*(2), 175-193.

Heldal, I., Spante, M., & Connell, M. (2006). Are two heads better than one? Object-focused work in physical and in virtual environments. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology* (pp. 287-296). NY: ACM Press.

Järvelin, K. & Ingwersen, P. (2004). Information seeking research needs extension towards tasks and technology. *Information Research, 10*(1), October 2004. Retrieved January 12, 2008 from http://informationr.net/ir/10-1/paper212.html

Johnson, D., Macias, D., Dunlap, A, Hauswald, M., & Doezema, D. (1999). A new approach to teaching prehospital trauma care to paramedic students. *Ann Emerg Med.*, *33*(1), 51-5.

Kim, K. (2008). Effects of emotion control and task on Web searching behavior. *Inf. Process. Manage. 44*(1), 373-385.

Levene, H. (1960). Robust tests for equality of variances. In: I. Olkin, S.G. Ghurye, W. Hoffding, W.G. Meadows, & H.B. Mann (Eds.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling* (pp. 278-292), Stanford: Stanford University Press.

MeSH. (2008). Medical subject headings: Task performance and analysis. National Library of Medicine. Retrieved January 15, 2008 from http://mesh.kib.ki.se/swemesh/show.swemeshtree.cfm?Mesh_No=F02.808.600&tool=karo

Meyer, A. (1998). Death and disability from injury: A global challenge. *J Trauma*, *44*(1), 1-12.

Nahl, D., & Tenopir. C. (1996). Affective and cognitive searching behavior of novice end-users of a full-text satabase. *Journal of the American Society for Information Science, 47*, 276-286.

Nahl, D. (2005). Affective and cognitive information behavior: Interaction effects in Internet use. In A. Grove (Ed.), *Proceedings 68th Annual Meeting of the American Society for Information Science and Technology*, *Vol. 42.* Washington, DC: American Society of Information Science & Technology.

Phan, H.P. (2007). An examination of reflective thinking, learning approaches, and self-efficacy beliefs at the University of the South Pacific: A path analysis approach. *Educational Psychology, 27* (6), 789-806.

Robson, C. (2002). *Real World Research*. MA: Blackwell.

Rowley, T. (2001). Solving the paramedic paradox. *Rural Health News, 8*(3), 1-6, Rural Information Center Health Service, National Agriculture Library, Beltsville, MD. Retrieved January 18, 2008 from http://www.raconline.org/newsletter/web/fall01_vol8-3.html

Sonnenwald, D.H., Whitton, M.C., & Maglaughlin, K.L. (2003). Evaluating a scientific collaboratory: Results of a controlled experiment. *ACM Transactions on Computer Human Interaction, (10)*2, 150-176.

Stringer, K.R., Bajenov, S., & Yentis, M. (2002). Training in airway management. *Anaesthesia*, *57*(10), 967–983.

Tan, D. S., Gergle, D., Scupelli, P. G., & Pausch, R. (2004.) Physically large displays improve path integration in 3D virtual navigation tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 439-446). NY: ACM Press.

Vakkari, P. (2003). Task-based information searching. In B.Cronin (Ed.), *Annual Review of Information Science and Technology, Vol. 37* (pp. 413-464). Medford, NJ: Information Today.

Welch, G., Sonnenwald, D.H., Mayer-Patel, K., Yang, R., State, A., Towles, H., Cairns, B., & Fuchs, H. (2005). Remote 3D medical consultation. In *Proceedings of the BROADMED Conference* (pp. 103-110). Boston: Omnipress.

Wilhide, S. (2002). Rural health disparities and access to care. National Rural Health Association. Retrieved January 18, 2008 from http://www.iom.edu/Object.File/Master/11/955/Disp-wilhide.pdf

