Collaboration at a distance: Remote expert support during a troubleshooting task

Rosan Sengler

University of Twente

First supervisor: Prof. Dr. Jan Maarten Schraagen

Second supervisor: Simone Borsci

Third supervisor: Tom Hueting

2 April 2021

UK English – APA Style

Word Count: 9,476

Abstract

Introduction: Several different work fields could benefit from remote expert support to novices on location. A possible problem that could arise from this approach is that it would create a new generation of employees that are not truly experts at troubleshooting.

Objective: To investigate the effect of being supported by a remote expert or working alone with a manual during a troubleshooting task on the ability of a novice to complete tasks on their own at a later time.

Methods: The experiment, with 36 participants, simulated a troubleshooting task. During the unsupported (baseline) condition the participant solved several puzzles on their own using a manual that explained how to solve the puzzles. In the supported condition the participant solved the puzzles while being supported by an expert at a distance. The supported condition consisted of three phases, throughout which the participant gradually worked more independently: video-audio phase, audio-only phase, and the individual phase. After this, both conditions performed the transfer task which measured whether the participants were able to complete a more difficult task with similar solving strategies on their own.

Results: During the individual phase, the supported condition showed a significantly faster time on task, but the unsupported condition was significantly more correct. During the transfer task only one of the five trials showed a significant result for time on task in which the supported condition was faster, and no significant difference in number of correct answers.

Conclusion/Discussion: The supported condition was faster, but the unsupported condition was more correct, suggesting there may have been a speed-accuracy trade-off that could have been influenced by initially having expert support or not. The transfer task and

training component may not have been optimally designed, causing the lack of significant results in the transfer task.

Keywords: Collaboration at a distance; Expert support; Remote; Deskilling; Troubleshooting.

Table of Contents

Collaboration at a distance: Remote expert during a troubleshooting task

When two or more parties have to collaborate whilst each party is based in a different location it can be referred to as remote collaboration or collaboration at a distance. It is becoming increasingly common for companies to require their employees to collaborate at a distance when completing a certain project or task. Distant collaboration can be required due to various constraints (e.g., global pandemic, urgency of the issue, geographic location) or benefits (e.g., cost effectiveness, time efficiency, decrease in travel time). However, the distance can most definitely influence the collaboration, either in a positive or negative way.

The Influence of Distance

Distance can make the collaboration more laborious and the coordination of the collaborators more challenging, because it requires considerable verbal explicitness (Stone & Posey, 2008). Patterson, Bernal and Stephens (2012) found that distributed teams more often had to clearly rule out options when trying to find a solution to a problem than teams working in the same room. Thus, requiring indisputable and effortful communication between parties. Additionally, there is a lack of 'mutual knowledge' in dispersed teams (Cramton, 2001), which can lead to fragmentation of information within a group of collaborators. Mutual knowledge is the knowledge that the interacting parties share, and of which they all know that they share it. People work best when they are able to see and anticipate what their colleague is doing or will be doing, as this helps build common ground (Fletcher & Major, 2006; Salimian, Brooks, & Reilly, 2019). Thus, when they have situational awareness of their co-workers, it allows others to give feedback and provides for easier coordination (Fletcher & Major, 2006).

Remote collaboration can, however, be beneficial to the focus, effectiveness and productivity of the collaborators because it allows for fewer non-task-related 'side conversations'

(Hatem, Kwan, & Miles, 2012). As the means for communication are restricted people are less likely to engage in informal conversations, thus, leading to less distractions (Bos, Shami, Olson, Cheshin, & Nan, 2004; Deshpande, Sharp, Barroca, & Gregory, 2016; González-Ibáñez, Haseki, & Shah, 2013; Hatem et al., 2012; Tang, Isaacs, & Rua, 1994). Therefore, any interaction that does occur becomes more task-focused and contains very little small talk. It's preferable that these teams have already created an interpersonal bond through several face-to-face meetings at the beginning of the collaboration because it helps bring the teammates closer together and allows for the social aspect of working together (Aritz, Walker, & Cardon, 2018).

Communication Channel

Several studies that have compared collocated (face-to-face) collaboration with remote (connection through video-audio) collaboration, found that collocated teams generally performed better or equal to the remote teams (Cohen & Thias, 2009; Hammond, Harvey, Koubek, Compton, & Darisipudi, 2005; Pazos & Beruvides, 2011; Wheeler, Valacich, & Alavi, 1995). Additionally, the majority of studies found that remote teams did not surpass the collocated teams in terms of performance.

In the cases where the use of a communication channel is necessary because remote collaboration is required, the video-audio combination generally facilitates the best performance during collaboration (Hammond et al., 2005; Tang & Isaacs, 1993; Veinott, Olson, Olson, & Fu, 1999). Video allows for natural processes of communication, such as turn-taking, gestural language, determining each other's understanding (Tang & Isaacs, 1993; Veinott et al., 1999), which are all elements normally found in face-to-face interaction. A video connection complements a remote collaboration quite a bit, but the audio channel plays a more crucial role

in collaborations (Goebbels & Lalioti, 2001; Goebbels, Lalioti, & Göbel, 2003; Tang & Isaacs, 1993; Tang et al., 1994), without audio the collaboration cannot be performed as well.

Remote Expertise

As stated above, in some cases collaboration at a distance can be beneficial to the performance in terms of productivity, efficiency and effectiveness (Bos et al., 2004; Deshpande et al., 2016; González-Ibáñez et al., 2013; Hatem et al., 2012; Tang et al., 1994). To further explore the concept of remote collaboration and its possible advantages, it is interesting to look at remote expert support for troubleshooting (finding and fixing problems).

Remote expert support can be explained as an expert on a specific subject matter, such as machine repair on naval ships, being readily available to support a capable novice (for instance, a maintenance worker with general knowledge of repairs) in troubleshooting when one of the machines on the ship needs a repair. Although the capable maintenance worker is a novice on this specific subject matter, he is already capable of repairs and knows how to use tools. The reason this situation would require an expert is that there can be very specific issues that present themselves, that a general maintenance worker would not know the particular fix to while at the same time the expert would not physically be present to assist.

The expert could work from a type of control centre on the shore that could support multiple ships with different issues simultaneously. Given these specific issues do not constantly present themselves, distributing expertise across platforms could be an efficient option. Therefore, troubleshooting from a control centre could provide the solution. By utilizing a remote expert, it can be ensured that their time and knowledge is used more efficiently (Bos et al., 2006), because they will not be asked to help out with common tasks that novices could have completed on their own.

There are many different organisations that could benefit from an expert that provides remote support to novices on location, a few examples being the navy, maritime companies, or a shared resource lab during the pandemic. The navy only has a limited number of highly specialized maintenance staff available that can be deployed to their ships, and it was shown that novices could perform the job well and in a safe manner when being supported by an expert (Post, Van Den Boogaard, & Rasker, 2004). Additionally, from a private sector perspective, companies could provide remote customer support instead of having an expert travel to customer that requires troubleshooting (Zoric et al., 2016). Furthermore, with a reduction of on-site staff due to the pandemic, shared resource labs were forced to use remote support in order to assist the end-users in troubleshooting (Gravano, Chakraborty, Pesce, & Thomson, 2021). Thus, both the private and public sector could employ such techniques to make the best use of their experts.

Remote expert support is hypothetically able to solve a lot of problems dealing with distances the experts would otherwise have to travel and allows for more rapid troubleshooting solutions to facilitate repairs in remote locations. Therefore, this research focusses on whether expert support helps or hinders the ability of the individual on location to later complete more difficult troubleshooting tasks on their own.

However, the question this research raises is whether remote expert support of general maintenance workers is a lack of investment in human capital similar to the kind that leads to deskilling.

Deskilling is a consequence of automation (Bainbridge, 1983). The process of deskilling can be explained as skilled labour being replaced through implementing technologies that merely have to be operated by semi-skilled or unskilled employees. Although the process can initially be

cost effective as less skilled employees are required for the same jobs, it eventually also reduces the investment in human capital.

An even more pressing form of deskilling is latent deskilling, as long as the technology is in place the deskilling happens without the workers' knowledge. Thus, only once the system fails or is discontinued (such as automation of certain complex tasks) the effects of deskilling are brought to the surface (Rinta-kahila, Pentinnen, Salovaara, & Soliman, 2018). It can be detrimental in terms of level of resilience, if, once a system fails, none of its workers know how to perform the tasks that were previously being carried out without their input. It can be quite a dilemma to decide whether one could better use little skill, more technology reliance and fewer costs or more skill, more costs, and more self-reliance (Bhardwaj, 2013).

Deskilling is for instance found in cockpit automation, which causes the pilot to gradually lose some of their skills. Extensive use of technology can lead to over-reliance on assistance, which can cause a lack of psychomotor and cognitive skills that would be necessary when going into manual flight (Ferris, Sarter, & Wickens, 2010). Hence, in the rare situation that the pilot is required to take over and manually control the airplane, they might not be able to intervene in an adequate manner anymore. Problems with return-to-manual performance are not only found in cockpit automation, but can occur to any operator of partially or fully automated systems (Endsley & Kiris, 1995). Since remote expert support in essence allows the more complicated processes to be taken over by the expert, it might create a similar situation where the general maintenance worker will not be able to solve the issues on their own once the connection with the expert fails.

Research Goal

Therefore, the question this research will aim to answer is: What is the effect on a novice's performance when they complete troubleshooting tasks either supported by a remote expert or on their own with a troubleshooting manual?

To simulate a troubleshooting task, participants will have to solve several tasks either whilst being supported by an expert that guides them through the troubleshooting process or individually using a manual that explains how to solve each task. The unsupported condition provides the baseline to which the supported conditions will be compared. The performance measures that this research will focus on are: (1) Time to complete the task; (2) Number of correct answers; and (3) Number of trials needed to reach peak performance (three correct trials in a row). As the participants will be 'novices' to the task, the gradual progression of time on task over time will also be visualised to illustrate the learning curve of the participants.

The participants in the supported condition will gradually move from the most optimal manner of communication at a distance (audio and video connection; Hammond et al., 2005; Tang & Isaacs, 1993; Veinott et al., 1999), to a less optimal condition (audio-only), and finally, the least optimal condition (no contact with the expert at all) to simulate a real-life scenario of losing the connection when collaborating. The aim is to test whether the participants learned enough from the collaboration with the expert to then be able to correctly complete the tasks on their own as well, or to see if the person that has been working individually from the start will perform better. Thus, the combination of the three phases (video-audio, audio-only, and the individual phase) can be seen as training.

Less investment in human capital, in terms of training each individual to become a well-rounded expert (Bhardwaj, 2013), or an over-reliance on the usual assistance (Endsley & Kiris,

1995; Ferris et al., 2010), can cause poorer performance once the usual assistance or the system's automation fails. When the expert is no longer available to the participant, the expectation will be that the participant will not be able to perform as well individually. They have been able to rely on assistance up until that point, but may, subconsciously, not be as capable of solving the troubleshooting tasks on their own.

Therefore, the expectation within this research is that the unsupported participants, that work individually from the start, will be able to achieve the best performance (measured in time on task, number of correct answers, and trials required to reach peak performance). The expectation is that the supported condition will perform worse on time on task, because when assistance fails it can negatively impact decision time (Endsley & Kiris, 1995):

> *H_1 = The participants in the supported condition will perform slower, during the individual phase and the transfer task, than the participants in the unsupported condition.*

That the unsupported condition will be able to achieve the best performance in number of correct answers, because the dependence on assistance can lead to a lack of cognitive skills (Ferris et al., 2010):

> *H_2 = The participants in the supported condition will have fewer correct answers, during the individual phase and the transfer task, than the participants in the unsupported condition.*

That the unsupported condition will be able to achieve the best performance in number of trials required to reach peak performance, because return-to-manual performance is problematic once someone has relied on assistance (Endsley & Kiris, 1995):

> *H_3 = The participants in the supported condition will require more trials, during the individual phase, to reach peak performance.*

The participants in the unsupported condition will get to work with the troubleshooting instructions from the start and through the process they will fully learn how the task must be performed. The supported participants are expected to not perform as well, due to the fact that they might become too reliant on the expert, so they do not become capable of actually solving the task on their own very well.

Methods

The experiment simulated a troubleshooting task in which the participant (the novice) was either supported by a remote expert or had to work on their own using a manual with instructions on how to solve the tasks. The independent variable, 'remote support', was a between-subjects variable. In order to ensure that the expert's skill or method of work did not vary over the different trials with the different participants, the researcher took the role of expert.

The experiment consisted of a training component and a transfer component. During the training component the participants performed three types of problem-solving tasks and during the transfer component they completed a more difficult task five times.

Participants

Forty-four people signed up to participate in the research. The participants were recruited through convenience sampling, via the participant pool of the University of Twente and social media platforms such as Facebook. The inclusion criteria were as follows:

- **Age** – The participants had to be aged between 18 and 65 years of age. This range represents people of a working age, which reflects the group of people that remote expert support could possibly apply to if it becomes regular practice.
- **Language** – The participant had to be able to speak English or Dutch. The experiment could be performed in both languages, and the expert (the researcher) was able to speak both fluently.
- **Colour Blindness** – As the colours of items in the puzzles were key to reaching the solution, it was important that the participants were able to correctly recognize the colours. Thus, individuals with colour blindness could not take part in this study.

- **Keep Talking and Nobody Explodes** – The puzzles from this research were based on the puzzles from the game 'Keep Talking and Nobody Explodes'. If participants had played this game before the results might not show a realistic learning curve, therefore they were excluded. This was, however, not a criterion that could be checked during the pre-screen, because it could have prompted participants to look up the game and familiarize themselves with it. Thus, this criterion was checked once the participant was present for the experiment.

The final sample consisted of thirty six participants for several reasons: (1) Five people had played the game 'Keep Talking and Nobody Explodes before; (2) One individual did not show up for the experiment; (3) Two people clearly did not understand the puzzles and indicated that they were guessing the answer. Additionally, one participant's information was excluded from the awareness check analysis (which will be explained further in the design section), because the video file was corrupted, and all the awareness data was lost. The lost recording only influenced the awareness check, and the participant was still included in the other analyses.

The average age of the participants was 27 years (SD = 11.89). Furthermore, exactly half of the participants were male, and the other half female. The majority of the sample performed the questionnaire in Dutch (61.1%), and the rest performed it in English (38.9%). Lastly, all participants indicated they were not colour blind.

Materials

The following materials were used:

- Researcher computer – HP EliteBook
- **Computer remotely located participant** – Unfortunately, the make and model of the computer was different depending on which computer the participant used.

- **Qualtrics** – Survey tool used to present the puzzles and record participants' answers.
- **The manual** – The manual with the instructions on how to solve the separate puzzles. Either an English or Dutch version of the manual was used.
- **Skype** – The communication channel used (both audio and video) between the participant and the researcher.
- **Experiment packages** – An experiment package was a collection of all the relevant papers to conduct an experiment. For each experiment, the researcher used one experiment package. The package included:
    - An answer key to all the puzzles, so that the researcher could keep track of the number of correct answers given during the experiment. Each puzzle required one answer, thus, one correct answer meant one correctly solved puzzle.
    - Space to make notes during the experiment.
    - The awareness check, assessing how aware the participants were of how to solve the puzzles.

Design

The experiment consisted of two conditions. The unsupported condition created the baseline measurement and the supported condition simulated collaboration between the expert and the novice. An overview of the progression of each condition can be seen in Figure 1, both conditions will be further detailed below. All the blocks before the transfer task are the training component, and the block transfer task represents the transfer component.

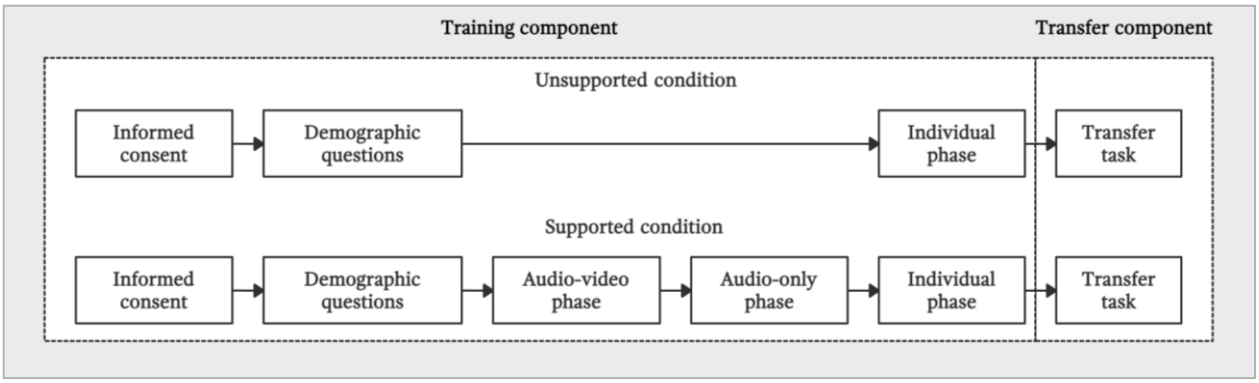


Figure 1. Timeline of the progression of the two conditions

During the training component of the experiment the participants had to perform three types of problem-solving tasks, in order to simulate troubleshooting. Each task was a small puzzle that the participant had to solve. The various puzzles differed slightly from each other and therefore had a different solution each time. The physical features of the puzzle (e.g., the colours of its parts, the orientation of the puzzle, or what words were written on it), determined what its solution should be. Thus, the solution strategy for each puzzle was the same. A manual explained how to solve each puzzle depending on those features.

Examples of the three puzzles (called wires, keypads, and the button) and the corresponding sections from the manual can be found in Appendix I, II, and III. The three puzzles were presented in the same order every time (wires, keypads, and the button, respectively), and the three puzzles together made up one trial.

Unsupported condition. During this condition, the participant had to work independently and had access to the manual with instructions on how to solve the three different puzzles. In the questionnaire, the manual was presented at the top of the page, and the puzzle could be found underneath it. The participants had to perform trials until they reached peak performance, which was defined as completing three trials correctly in a row. Each puzzle required one answer, thus, a correctly solved puzzle counted as one correct answer.

The participants did not know they had to complete three trials correctly in a row, in order to be allowed to move on to the transfer task. They were simply instructed by the researcher to continue solving the puzzles or to move on to the next component.

Once they had completed the 'individual phase', they would continue to the transfer task which measured whether they were able to complete a more difficult task with similar solving strategies on their own. During the transfer task the participant had to complete five puzzles of the same type on their own with the manual. They could see the time they were spending on each puzzle at the top of the page. Once they had completed the five puzzles the experiment was concluded. Each of the five puzzles required six independent answers, thus, over the entire transfer task the participant gave a total thirty answers which were either correct or incorrect.

Further explanation of the transfer task and an example of how it looked can be found in Appendix IV.

Supported condition. During this condition, the participant and the expert worked together to solve the puzzles. For the supported condition in the training component the decision was made to gradually move from the most optimal medium at a distance (audio and video connection), to a less optimal medium (audio-only), and finally, the least optimal medium (no contact with the expert at all and access to the manual; the individual phase) to simulate a real-life scenario of losing the connection when collaborating. Thus, this condition consisted of three separate, subsequent phases:

1) **Video-audio phase** – There was video and audio contact between the participant and the expert via video-conferencing software. This was achieved through sharing the screen of the participant, so the expert could see the puzzles that the participant had on their screen, as well as being connected via audio. Additionally, the expert and the

participant were able to see each other through video. Only the expert had access to the manual with instructions on how to solve the puzzles. In this phase, the participant mainly followed instructions without knowing the exact rationale for the action, because the expert was able to see all the physical features of the puzzle and figure out what the right solution should be for each. The participant gradually got to know the puzzles and how they can be solved during this phase, thus, building up a slight understanding of how the puzzles work. This was the optimal communication medium for working at a distance.

2) **Audio-only phase** – During this phase, the expert and participant were only connected via audio. Thus, the expert was not able to see the puzzle anymore, which simulated the video connection failing during the collaboration. Again, only the expert had access to the manual. The lack of visual access by the expert to the puzzle required the participant to become verbally explicit about the puzzle and its physical features so that the expert understood what the puzzle looked like. The additional descriptions could simultaneously give the participant a better sense of what information might be useful or essential when solving the different puzzles. The participant had to be more engaged during this phase and could build up a moderate understanding of the puzzles and how to solve them. During this phase, the expert took on a reserved role. Initially, the expert waited for the participant to volunteer all the information that the participant deemed important to solve the puzzle. Once they had finished, the expert then asked about the information that was still missing to produce a solution. Data from the audio-only phase was analysed to perform the awareness check, which measured how aware the participants were of what

information was required to solve a puzzle. The awareness check will be further explained later on.

3) **Individual phase** – During this phase, the participant had to work independently without expert support using the manual which instructed them on how to solve the different puzzles. Thus, the participant truly worked alone, to simulate the connection failing during the collaboration. It tested whether the participant was able to independently solve the puzzles.

In this condition, the pair (the participant and the expert) kept performing the puzzle-solving tasks within a specific phase until they reached peak performance (three trials completed correctly in a row). Once peak performance was reached, they moved on to the next phase. When they had completed the first two phases the participant went on to the individual phase, which the participant performed on their own. Once they had three correct trials in a row they were allowed to move on to the transfer task, which tested their ability to complete a more difficult task with similar solving strategies individually. The participant performed the transfer task, which was constructed the same as for the unsupported condition (five puzzles on their own using the manual).

Awareness check. The awareness check was a form of a manipulation check to ensure that the participants, supported by an expert, were learning somewhat throughout the trials. During the video-audio and audio-only phase of the supported condition, the participants worked in collaboration with the expert (the researcher) to find a solution to each puzzle. During the audio-only phase, the participant and researcher were only allowed to use audio to find the solution together. Thus, the participants had to be verbally explicit, so that the researcher could understand what the puzzle looked like and find how it should be solved. During this phase, the

awareness that the participant had of the information that was essential to solving the transfer task was measured. The awareness check itself can be found in Appendix VII.

In each trial there were seven essential information pieces necessary to solve the puzzles (two for wires, one for keypads, and four for the button), meaning that for each trial the participant could score a maximum of seven points. The number of information pieces that each participant provided on their own initiative, which helped towards solving the puzzle, was individually counted. The participant's score was transformed into a percentage of the pieces of essential information that they provided on their own initiative, thus, quantifying how many bits of information the participant correctly identified as being crucial to the completion of each trial.

Procedure

After the experiment was thoroughly explained to the participants, they signed an informed consent that was integrated within the Qualtrics questionnaire. An example of the questionnaire can be found in Appendix V, and informed consent can be found in Appendix VI. The participant was then informed that the fastest participant in each condition would receive a bol.com voucher worth €15,-. This was to motivate the participants to work through the problems correctly but also quickly. After this, the participant went through a different procedure depending on the condition:

Unsupported condition. The participant started with the individual phase. They had the manual and worked through the puzzles on their own. Once they reached peak performance (three correct trials in a row), the researcher instructed them to move on to the transfer task, in which they had to solve five puzzles of the same type on their own with the manual. There was a timer at the top of every page that showed how long the participant was working on that specific puzzle. When they had worked through the transfer task the experiment was finished. The

participants had the option to leave their email address if they wanted to win one of the two bol.com vouchers.

Apart from technical or practical questions, the participant was not allowed to ask anything about how to actually complete the trials. If participants did have questions about how the puzzles should be solved, the researcher merely restated the instructions that were already shown to the participant. No additional help was provided.

Supported condition. The participant was connected through audio and video with the researcher. During the video-audio and audio-only phase of the experiment the researcher was the only one who had access to the manual, and the participant would follow the instructions provided by the researcher in order to solve the different puzzles that together constituted one trial.

During the video-audio phase, the participant shared their screen on Skype, to make sure the expert was able to see the puzzles. Additionally, they were able to communicate with each other through the audio channel that Skype provides. During this phase, the participant mostly just listened to the instructions from the expert. As the expert was able to see the puzzles directly, the participant did not have to describe or explain anything.

During the audio-only phase, the participant was still able to communicate using the audio channel, but the expert no longer could see what the participant had in front of them. Thus, the participant had to verbally describe what the puzzle looked like and gained more insight in which information would be essential to be able to solve the puzzles.

Finally, in the individual phase, the participant had to solve the puzzles on their own entirely. Within the questionnaire, they were presented with the applicable sections of the

manual. Thus, they were presented with all the information necessary to be able to solve the puzzles.

In each of the phases described above the participant was only allowed to move on to the next phase once they had completed three correct trials in a row. Once the three phases were complete they moved to the transfer task (identical to the end of the unsupported condition), in which they had to solve five puzzles of the same type on their own with the manual. There was a timer at the top of every page that showed how long the participant was working on that specific puzzle. Once the participant had completed the five transfer task puzzles, the experiment was concluded. The participants had the option to leave their email address if they wanted to win one of the two bol.com vouchers.

Data Analysis

The data analysis was divided into four parts. Firstly, the results from the awareness check will be presented through a graph to indicate the general progression of awareness of the participants throughout the trials. Their awareness was the level of knowledge they had of what pieces of information were essential to solving the puzzle.

Secondly, the individual phase in the unsupported and supported condition were compared on time to complete each task, the number of correct answers given, and the number of trials required to reach peak performance. These phases are comparable because in both of them the participants had to work without collaboration with the researcher and were only allowed to use the manual to solve the puzzles. These were compared between the conditions using the independent samples t-test and the Mann-Whitney U-tests, depending on whether the dependent variables were normally distributed or not.

Thirdly, the transfer task of each condition was compared on time to complete the task and the number of correct answers given. This was also done by using the independent samples t-tests and the Mann-Whitney U-test. to compare whether the unsupported and supported condition significantly differs in their outcomes.

Lastly, an exploration of the data was included to investigate whether the results found in the previous analyses held up when the participants were matched and when only the participants that had between zero and three mistakes in the individual phase were analysed.

The normality of the relevant data samples for each analysis was tested to determine whether it was appropriate to use the parametric independent samples t-test or the non-parametric Mann-Whitney U-test. In the results section, only the outcomes of the Shapiro-Wilk's test will be discussed, but for each sample, the following numerical and visual outputs were considered to determine whether the data was normal: (1) Whether the skewness and kurtosis z-values were between -1.96 and +1.96; (2) Whether the data plotted in a histogram looked approximately normally distributed, approximately following a bell-curve shape; (3) Whether the data points in the Q-Q plots were on or very close to the line; (4) Whether a box plot of the data was approximately symmetrical; and finally (5) Whether the Shapiro-Wilk's test showed a non-significant result.

Results

Awareness Check

The average awareness, measured as the number of items volunteered out of seven possible, for each trial is shown in Figure 2. The audio-only phase consisted of three trials for each participant because there were no mistakes made as the participants collaborated with the expert.

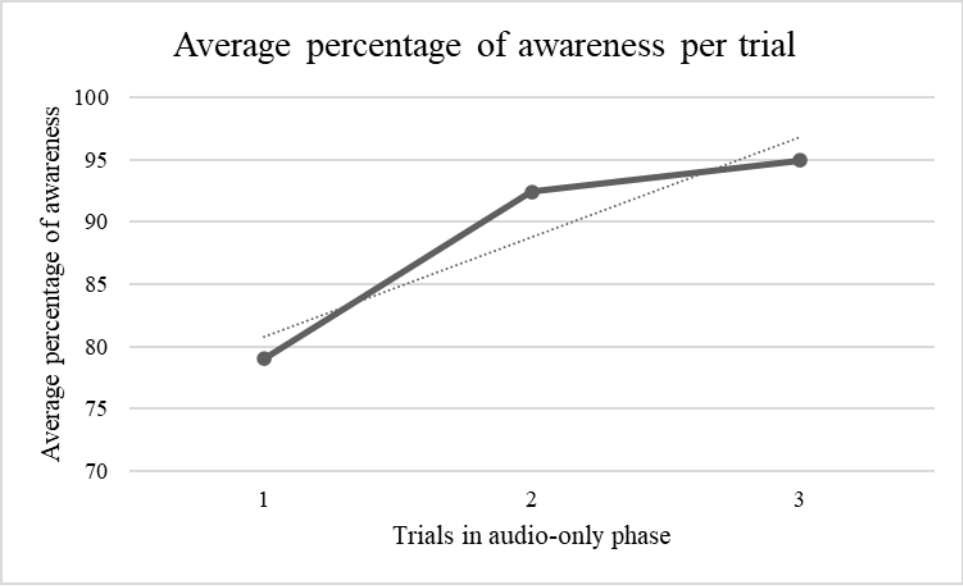


Figure 2. Average percentage of awareness per trial

As can be seen from Figure 2, there was a general trend upwards in the information that participants provided on their own, without any prompt from the researcher. The trend line (R^2 = .87) shows a positive correlation of R = .93, indicating that as the trials progress the average awareness increases. Thus, participants became more aware of what information was key to solving the puzzles as they went through the trials.

Individual Phase

Total trials. The means for the unsupported condition (M = 5.06, SD = 3.39) and the supported condition (M = 4.78, SD = 1.76) were compared using a Mann-Whitney U-test. A non-

parametric test was chosen because the Shapiro-Wilk's test showed the data from both conditions was not normally distributed (unsupported condition $p = .00$; supported condition $p = .02$). The Mann-Whitney U-test showed a non-significant difference between the unsupported condition ($Mdn = 3$) and the supported condition ($Mdn = 4$), $U = 131.00$, $p = .17$. Thus, the null hypothesis cannot be rejected, because the unsupported condition did not perform better than the supported condition.

Time on task. When visualizing the time on task per condition (Figure 3 and Figure 4) it can be seen that the initial time it took the participants in the supported condition to complete the first puzzles looks lower than the time it took the participant from the unsupported condition to complete the first puzzles. It can furthermore be seen that all of the participants' progression lines flatten out after they have completed more trials.

A graphical presentation on time on task which includes both conditions in one figure is included in Appendix VIII.

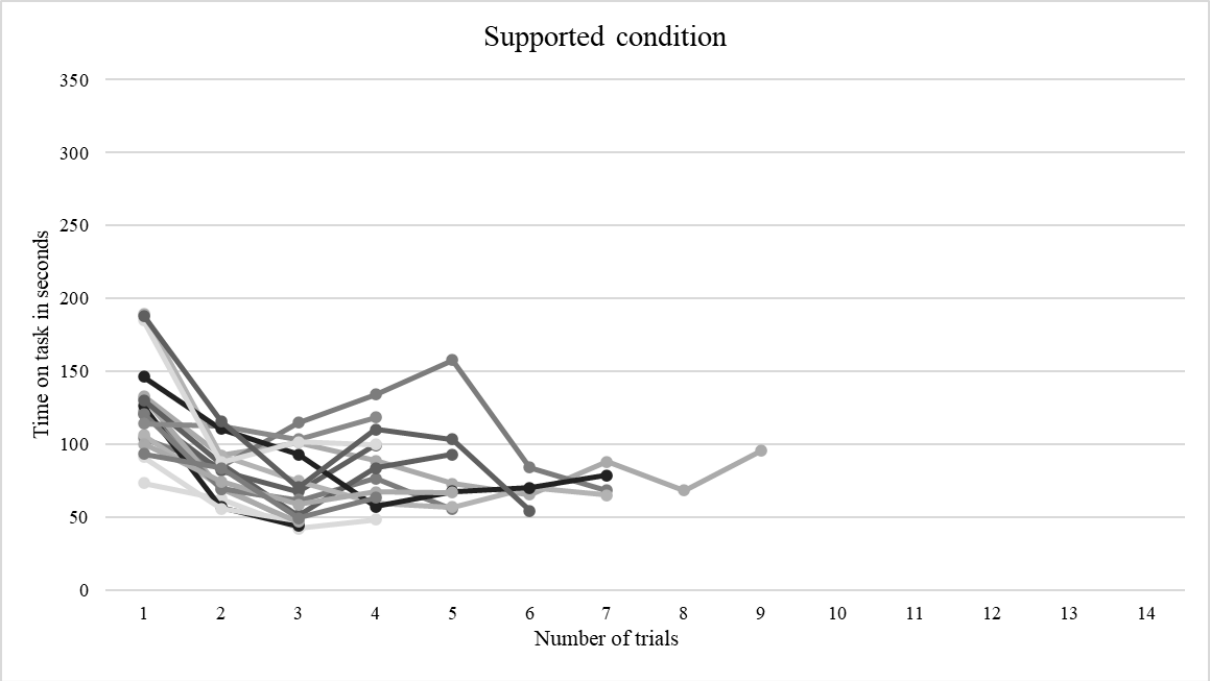


Figure 3. Time on task per trial for the supported condition

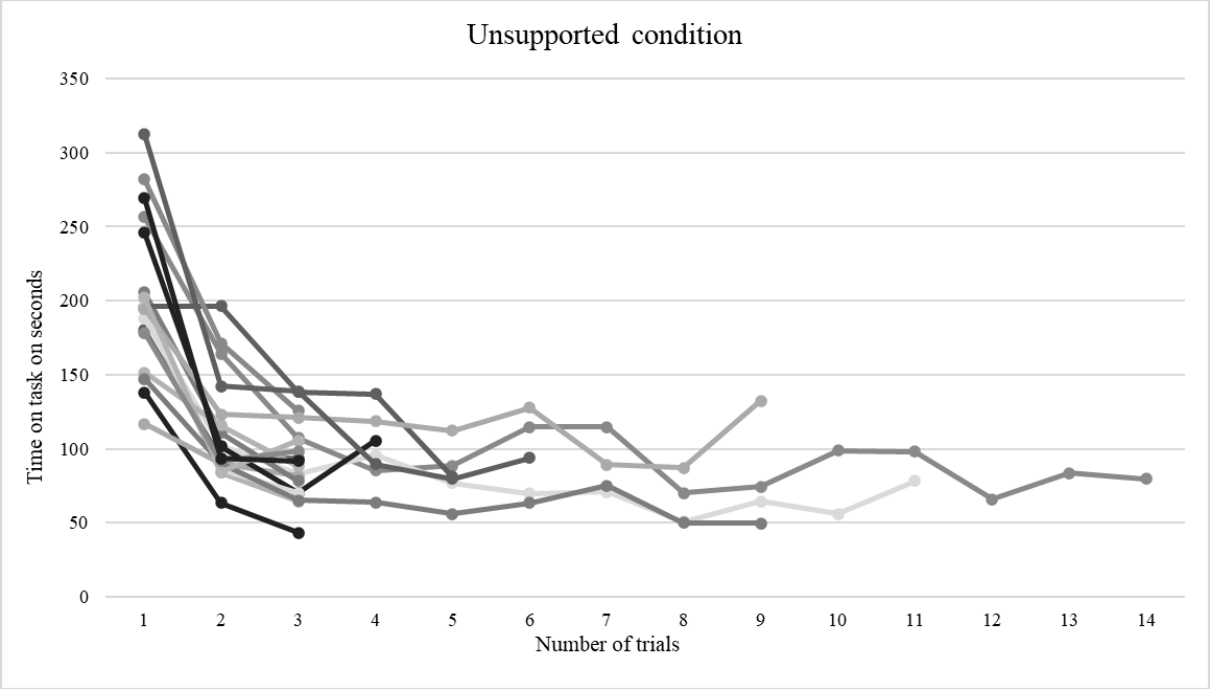


Figure 4. Time on task per trial for the unsupported condition

To further explore the time on task between the conditions a statistical test was performed on the first three trials of the individual phase which all participants conducted. After the first three trials the first opportunity arose at which participants could have reached peak performance. With each extra trial the number of participants that reached peak performance and moved on to the transfer task increased, and the number of participants that had to perform further trials steadily declined. Less participants completing a trial means fewer data points to compare, which affects the statistical power. It was chosen to only look at the first three trials to ensure large groups and many data points to ensure statistical power.

The normality test results of each trial per condition can be found in Table 1. Trial 1 was tested using an independent samples t-test, because both conditions were normally distributed. Trial 2 and 3 were analysed using a Mann-Whitney U-test, because in both trials one of the conditions was not normally distributed, thus, violating the assumption of normality.

Table 1. Shapiro-Wilk's test *p*-values on time on task for the first three trials in the individual phase per condition

Trial	Unsupported condition	Supported condition
1	.43	.09
2	.02	.29
3	.73	.02

In each comparison the participants of the supported condition on average completed the individual trials more quickly than the participants from the unsupported condition. The independent samples t-test showed a significant difference between the unsupported condition (M = 203.05, SD = 52.14) and the supported condition (M = 126.68, SD = 33.11) for trial 1 ($t(34)$= 5.25, p = .00, Hedge's g = 1.75). Additionally, the Mann-Whitney U-tests showed a significant difference between the unsupported condition (Mdn = 97.50) and the supported condition (Mdn = 83.10) for trial 2 (U = 69.00, p = .00, η^2 = .25) and a significant difference between the unsupported (Mdn = 89.86) and the supported condition (Mdn = 61.43) for trial 3 (U = 82.00, p = .01, η^2 = .18).

For all the trials the null hypothesis is rejected because the supported condition performed faster than the unsupported condition. This result was not in line with the expectation.

Number of correct answers. The number of correct answers provided in the unsupported condition and the supported condition was compared by taking the total number of correct answers given per participant divided by the number of trials they had completed. By this calculation, the average number of correct answers given was calculated. This was done to control for the number of trials completed. A participant could have a maximum of three correct answers on each trial. The averages had to be compared using a Mann-Whitney U-test, because the number of correct answers is count data and, therefore, is not normally distributed.

The test results showed that there was a significant difference between the unsupported condition (M = 2.85, SD = .22, Mdn = 3) and the supported condition (M = 2.65, SD = .28, Mdn = 2.66) when it comes to number of correct answers given (U = 94.00, p = .02, η^2 = .15). The null hypothesis is rejected because the average from the unsupported condition was higher than the average from the supported condition. The results are in line with the expectations that were set, the unsupported condition performed better in that they generally made fewer mistakes.

Transfer Task

The transfer task was used to test whether the participants had mastered solving the puzzles with similar techniques as they had employed during the previous phase(s). As they had a fixed number of puzzles that they had to solve in this section, the variable of 'number of trials' was not relevant for this task.

Time on task. The time on task per trial for the transfer task of both conditions was visualized in Figure 5. There was a participant in the supported condition during trial 1 who took more than 700 seconds to complete the task. In the experiment footage, it becomes clear that this participant spent quite some time familiarizing themselves with the task. However, after this, they swiftly worked through the next trials making few mistakes (five in total).

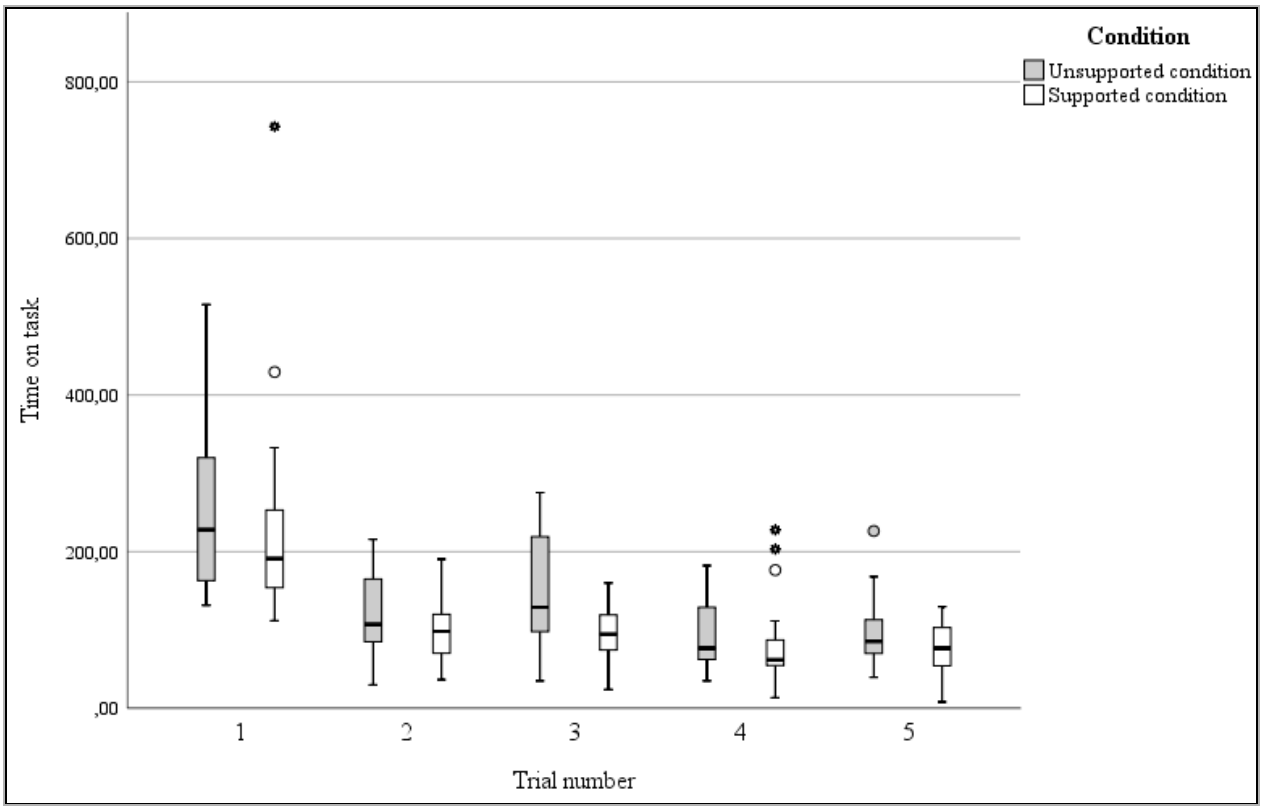


Figure 5. Time on task per condition and trial

A Shapiro-Wilk's test was conducted on each trial per condition to test the normally of each distribution. The results are summarized in Table 2. Based on the results trials 1, 4 and 5 were analysed using a Mann-Whitney U-test, because one of the conditions in each of those trials was not normally distributed. Furthermore, trials 2 and 3 were analysed using an independent samples t-test due to their data being normally distributed.

Table 2. Shapiro-Wilk's test *p*-values on time on task for the five trials in the transfer task per condition

Trial	Unsupported condition	Supported condition
1	.08	.00
2	.64	.55
3	.27	.53
4	.10	.00
5	.01	.78

It was decided to analyse each trial separately to ensure that the improvement or deterioration of the participants' performance over the trials was not lost in an average (Table 4). The independent samples t-test on trial 2 showed no significant difference between the unsupported condition (M = 121.40, SD = 56.42) and the supported condition (M = 96.23, SD = 41.34), $t(34) = 1.53$, $p = .07$; and on trial 3 a significant difference between the unsupported condition (M = 148.06, SD = 70.66) and the supported condition (M = 94.23, SD = 41.43), $t(27.45) = 2.79$, $p = .01$, Hedge's $g = .93$. Thus, in trial 3 the unsupported condition performed significantly slower than the supported condition.

Furthermore, the results of the statistical tests on trials 1, 4, and 5 are shown in Table 3 and Table 4. As can be seen from Table 4, none of the trials were significantly different between the unsupported condition and the supported condition. Thus, the results from these trials (1, 4, 5) are not in line with the expectation that the unsupported condition would perform better than the supported condition, and the null hypothesis cannot be rejected.

Table 3. Overview of medians of time on task in transfer task per condition

Trial	Condition	Median
1	Unsupported	227.93
	Supported	191.26
4	Unsupported	76.77
	Supported	61.75
5	Unsupported	85.45
	Supported	76.62

Table 4. Mann-Whitney U-test of time on task in transfer task per trial

Trial	*U*	Sig. (1-tailed)
1	138.00	.23
4	113.00	.06
5	112.00	.06

Number of correct answers. The number of correct answers given during each trial was compared between the conditions with a Mann-Whitney U-test, because a Shapiro-Wilk's test showed that the data for each separate trial was not normally distributed (Table 5).

Table 5. Shapiro-Wilk's test *p*-values on correct answers for the five trials in the transfer task per condition

Trial	Unsupported condition	Supported condition
1	.00	.01
2	.00	.00
3	.00	.04
4	.00	.00
5	.00	.00

Per puzzle a participant could have a maximum of six correct answers. The results are summarized below in Table 6 and Table 7.

Table 6. Means and standard deviations of correct answers in the transfer task per trial and condition

Trial	Condition	Mean	Std. Deviation	Median
1	Unsupported	4.89	1.08	5
	Supported	4.89	1.08	5
2	Unsupported	5.61	.78	6
	Supported	5.17	1.10	6
3	Unsupported	4.94	1.39	5.5
	Supported	4.56	1.10	5
4	Unsupported	5.44	.92	6
	Supported	5.17	1.04	5.5
5	Unsupported	5.06	1.30	5.5
	Supported	4.28	2.22	5.5

Table 7. Mann-Whitney U-test of correct answers in transfer task per trial

Trial	*U*	Sig. (1-tailed)
1	162.00	.50
2	125.00	.13
3	118.50	.09
4	135.50	.20
5	144.00	.29

As can be seen from the tables above, there were no significant differences found between the two conditions. Thus, the unsupported condition did not have more correct answers than the unsupported condition, and the null hypothesis cannot be rejected.

Influence of Accuracy

As an additional exploration of the data two different methods were used to look at more specific groups within the data set, to see if the results would differ from the ones found above.

Matched pairs. Pairs were matched between the unsupported and supported condition that had made an identical total number of mistakes in the individual phase. Then time on task in the individual phase and transfer task, and the number of mistakes made in the transfer task were compared between the conditions. The goal of matching pairs was to account for the natural variation of the participants in either condition. The sample was reduced to 18 participants, with 9 in each condition.

The full data analysis can be found in Appendix IX, but the outcomes were similar to the findings of the statistical tests that were performed over the entire sample. In the individual phase the supported condition consistently showed a lower mean or median time on task, meaning that the supported condition required less time to complete the tasks. Time on task during the transfer task was only significantly different between conditions in trial three, where the supported condition was faster than the unsupported one. There were no significant differences between the two conditions for the number of correct answers in the transfer task.

Correct participants. The participants were selected that had made between zero and three mistakes in the individual phase of the experiment, to then see if there was a significant difference in time on task between the unsupported and supported condition in the individual

phase and the transfer task. Furthermore, the correctness of the participants in the transfer task was checked. The sample was reduced to 30 participants with 15 in each condition.

This exploration was done in order to evaluate whether there would be different results for the individuals that had only made a small number of mistakes in the individual phase. In this manner the individuals with a large number of mistakes, which may have influenced the statistical tests, were filtered out.

The reason the inclusion criteria for this exploration was between zero and three mistakes is because any lower number of mistakes made the two conditions quite unequal in size, and the sizes of the groups were quite small. Thus, it was impossible to conduct a statistical test on such a low number of data points.

The full data analysis can be found in Appendix X, but the outcomes were similar to the findings of the statistical tests that were performed over the entire sample. In the individual phase the supported condition consistently showed a lower mean or median time on task, meaning that the supported condition required less time to complete the tasks. Time on task during the transfer task was only significantly different between conditions in trial three, where the supported condition was faster than the unsupported one. There were no significant differences between the two conditions for the number of correct answers in the transfer task.

Discussion

The focus of this research was to investigate whether expert support helps or hinders the ability of the individual on location to later complete more difficult troubleshooting tasks on their own. The expectation was that people who were not supported by an expert from the start would perform better on the more difficult tasks than people who were supported by an expert through the initial troubleshooting problems.

Individual Phase

The individual phase consisted of three different measures that assessed the difference in performance between the conditions: the number of trials required to reach peak performance, the time on task required to solve the puzzle, and the number of correct answers given.

The number of trials required to reach peak performance did not significantly differ between the two conditions. As the participants in the supported condition initially were able to rely on assistance from an expert, but then suddenly had to solve the problems individually, it was expected to cause problems similar to return-to-manual performance issues (Endsley & Kiris, 1995). However, the results show no worse performance in the supported condition. It is notable that there is no difference between the conditions' performances, as the participants in the supported condition had completed six trials prior to the individual phase (three in video-audio and three in audio-only). Yet apparently they were not able to perform better than the unsupported condition for whom the individual phase was the first phase they completed. This null result was not due to either a floor or a ceiling effect. As there was no ceiling effect present, it is unlikely that the puzzles were not difficult enough to show any differences. It could be that the participants in the supported condition did not grasp the puzzles enough while being supported by an expert, to then be able to perform better than participants that had not seen the

puzzle before. Alternatively, the participants from the supported condition could have become overconfident, as they had dealt with the puzzles in the previous phases, and possibly rushed through the trials heedlessly. Thus, resulting in a lack of a significant difference between the two conditions.

The time on task did differ significantly, with the supported condition completing the individual phase trials faster than the unsupported condition. These results, the supported condition being faster than the unsupported condition, were against expectations. The expectation was that the supported condition would be slower, because when assistance is withdrawn it can affect decision time negatively (Endsley & Kiris, 1995). Furthermore, the over-reliance on assistance can lead to a lack of cognitive skills (Ferris et al., 2010), which could reasonably impact time on task. This result could be explained by the fact that the supported condition did familiarise themselves with the puzzles during the video-audio and the audio-only phases. Not enough to reach peak performance faster, but enough to complete the individual trials faster.

The unsupported condition tended to have more correct answers than the supported condition, however, the effect size was relatively small. The results are in line with the expectations that the unsupported condition would make fewer mistakes, because an over-reliance on assistance can lead to a lack of cognitive skills (Ferris et al., 2010) which surfaces once the assistance is suddenly removed (Rinta-kahila et al., 2018). The small effect size could partially explain why there was no significant difference in the number of trials required to reach peak performance. As there was not a large difference between the mistakes made between the conditions, it probably did not cause a large difference in trials it took to reach peak performance.

The literature suggests that an over-reliance on assistance can be detrimental to individual performance once this assistance fails (Endsley & Kiris, 1995; Ferris et al., 2010). In the current research bad performance was represented by mistakes made and considerable time spent per task. As the unsupported condition performed slower, but the supported condition made more mistakes, bad results were obtained by both conditions. However, it can be argued that making more mistakes weighs heavier than taking more time to fix the issue.

A possible explanation for the results that indicate that the supported condition was faster yet less correct could have been a speed-accuracy trade-off. It could be that participants in the supported condition valued speed over accuracy and rather did not check twice if they understood the task correctly or had filled out the right answer due to overconfidence or a self-set standard of pace. As this is not found in the unsupported condition, it is likely that the supported condition was influenced by their remote expert support.

As was discussed earlier, the supported participants could have felt overconfident in their ability to solve the troubleshooting tasks and did not think they would need to double check or read carefully. Alternatively, the pace at which the participant was able to work through the problems while being supported could have created a subconscious standard of speed, as the supported trials were quite quickly without mistakes.

Although we know that effect does not last in any long-term capacity, it could be detrimental if a novice is convinced the speed of the repair is more important than the accuracy of the repair. The fact that the troubleshooting process was fast when supported by an expert, and a solution was found more quickly should not make the individual impatient or overconfident in their own skills. Thus, this should be considered and, if possible, remedied when companies want to make use of remote expert support.

Transfer Task

The transfer task had two different performance measures: the time on task required to solve the puzzle and the number of correct answers given. When comparing the time on task between the two conditions, none of the five trials showed a significant difference except for trial 3. During trial 3 the supported condition performed significantly faster than the unsupported condition. The results for trial 3 are against the expectations that the unsupported condition would perform faster, due to the negative effect on decision time if assistance fails (Endsley & Kiris, 1995). It is unclear why this effect could only be measured on trial 3.

The number of correct answers given during the transfer task did not show any significant results when being compared between the conditions. The expectation was that the supported condition would make more mistakes, due to a possible lack in cognitive skills brought on by an over-reliance on assistance (Ferris et al., 2010). There could be different reasons why the transfer task did not overall show a significant difference in time on task and number of correct answers: (1) The transfer task did not fit the skills learned in the training component of the experiment; or (2) After the individual phase, during the training component, the difference between conditions diminished due to more training with the puzzles.

The significant differences between time on task and number of correct answers in the individual phase did not show up in the transfer task. Thus, the shortage in investment in human capital in terms of training each individual to become a well-rounded expert (Bhardwaj, 2013) did not hinder the supported condition in their long-term performance. Furthermore, the sudden lack of assistance (Ferris et al., 2010), after having been supported before, did not cause a different outcome between the conditions on the long-term. This could suggest that the effect of the difference between being supported by a remote expert or not only lasts shortly after the

individual is forced to complete the tasks on their own. Thus, as the effect is not that large, remote expert support does not have any long-term negative influences on the performance of the novice. The initially supported novices were able to pick up the manual and eventually perform similarly to individuals that worked alone from the start.

Influence of Accuracy

Additionally, to further explore the data, statistical tests were conducted on a sample with only matched pairs of participants between the two conditions and on a sample of the most correct participants (that had between zero and three mistakes in the individual phase). The results from these tests showed exactly the same results as the statistical tests over the entire sample did. Thus, it can be said that the lack of significant results or the presence of significant results were not influenced by the level of accuracy of the participants. The matched pairs group and the participants with the least mistakes performed similarly to the entire sample when they were statistically tested individually.

Limitations

Firstly, as has been mentioned above, the transfer task may not have been optimally designed in this research. The results that were found in the training component were not carried through to the transfer component. Thus, the transfer task either did not fit the skills learned through the training component of the experiment, or it was not difficult or different enough to present a true challenge.

Secondly, the training component in the supported condition went through three different phases (video-audio, audio-only and individual). During the video-audio and audio-only phase the participant collaborated with the expert. The assumption was made that not everyone would be able to reach peak performance within the first three trials, however, the results showed that

no mistakes were made during those phases. It could be that three trials were enough for each of the participants to understand the puzzle adequately for that phase, but it is also likely that the training period did not challenge the participant enough.

Lastly, as the experiment was set up the supported condition participants would lose the communication channels as they were still in the learning aspect of the experiment. Each participant did have to reach their 'peak performance' in order to move on to the next phase, but as they collaborated with the expert in the first two phases (video-audio and audio-only) there were no mistakes made. Thus, 'peak performance' was not a pure measure of the participant's performance or skills learned for those phases, and it could be that not everyone's learning process was completed before they were moved to the next phase with less support.

Future Research

Firstly, it would be interesting to be able to perform an experiment that is much longer and actually allows for an entire learning process. Something that is more spaced out over time, as in this experiment all the phases were completed in a short time frame. When the experiment allows for the entire learning process it would also ensure that the communication modalities are not failing whilst the participant is still in the process of learning the tasks. Additionally, the experiment would more accurately simulate the contact that an individual on location and a remote expert might have over time, to then be able to test if the individual on location is capable of solving similar and more difficult troubleshooting problems individually.

Secondly, the combination of the tasks during the training component and the transfer component has to be examined. During this experiment, there seemed to be a mismatch between the two which may have caused the results found in the training component not being replicated in the transfer component. The transfer task is important because it tests whether the supported

troubleshooting enables the individual on location to solve similar and more complicated troubleshooting problems on their own. It is essential to know find out if the speed-accuracy trade-off is a problem might carry on into the long-term troubleshooting activities of the individual on location.

Thirdly, due to certain constraints, the tasks the participants completed were provided in the form of a questionnaire and based on puzzles with clear indisputable features (e.g., one red and one blue wire) that led to a clear-cut solution. It would be interesting to allow participants to work actual physical problems with more complicated and less straightforward solutions. On an experiment level it allows for a more immersive experience for the participants, and the entry of the answers would be a lot more intuitive than it was through the questionnaire. From an implementation perspective, a more authentic scenario could add a lot of nuance to the straightforwardness of the solutions found in the manuals.

Lastly, if particular companies or work fields are interested in the idea of a remotely located expert (e.g., the navy), they would be able to provide actual tasks or problems that they might require the remote experts for. It would allow for a more specified research with immediate and direct implications because it would give a more accurate representation of how effective the collaboration would be in reality. The tasks or problems could be taken from more real-life examples, the experts would actually be experts on specific topics and the general employees would already have general knowledge of the field.

References

Aritz, J., Walker, R., & Cardon, P. W. (2018). Media use in virtual teams of varying levels of coordination. *Business and Professional Communication Quarterly*, *81*(2), 222–243. https://doi.org/10.1177/2329490617723114

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Bhardwaj, S. (2013). Technology, and the up-skilling or deskilling conundrum. *WMU Journal of Maritime Affairs*, *12*, 245–253. https://doi.org/10.1007/s13437-013-0045-6

Bos, N., Olson, J. S., Nan, N., Shami, N. S., Hoch, S., & Johnston, E. (2006). Collocation blindness in partially distributed groups: is there a downside to being collocated? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1313–1321. https://doi.org/10.1145/1124772.1124969

Bos, N., Shami, N. S., Olson, J. S., Cheshin, A., & Nan, N. (2004). In-group/out-group effects in distributed teams: an experimental simulation. *Proceedings of the 2004 ACM Conference on Computer Supporter Cooperative Work*, 429–436. https://doi.org/10.1145/1031607.1031679

Cohen, B., & Thias, M. (2009). The failure of the off-shore experiment : A case for collocated agile teams. *2009 Agile Conference*, 251–256. https://doi.org/10.1109/AGILE.2009.8

Cramton, C. D. (2001). The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science*, *12*(3), 346–371.

Deshpande, A., Sharp, H., Barroca, L., & Gregory, P. (2016). Remote working and collaboration in agile teams. *Thirty Seventh International Conference on Information Systems, Dublin 2016*.

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of

Control in Automation. *Human Factors*, *37*(2), 381–394.

Ferris, T., Sarter, N., & Wickens, C. D. (2010). Cockpit automation: Still struggling to catch up... In E. Salas & D. Maurino (Eds.), *Human factors in aviation* (2nd ed., pp. 479–503).

Fletcher, T. D., & Major, D. A. (2006). The effects of communication modality on performance and self-ratings of teamwork components. *Journal of Computer-Mediated Communication*, *11*(2), 557–576. https://doi.org/10.1111/j.1083-6101.2006.00027.x

Goebbels, G., & Lalioti, V. (2001). Co-presence and co-working in distributed collaborative virtual environments. *Proceedings of the 1st International Conference on Computer Graphics, Virtual Reality and Visualisation*, 109–114. https://doi.org/10.1145/513867.513891

Goebbels, G., Lalioti, V., & Göbel, M. (2003). Design and Evaluation of Team Work in Distributed Collaborative Virtual Environments. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, (October), 231–238.

González-Ibáñez, R., Haseki, M., & Shah, C. (2013). Let's search together, but not too close! An analysis of communication and performance in collaborative information seeking. *Information Processing & Management*, *49*(5), 1165–1179. https://doi.org/10.1016/j.ipm.2012.12.008

Gravano, D. M., Chakraborty, U., Pesce, I., & Thomson, M. (2021). Solutions for shared resource lab remote quality control and instrument troubleshooting during a pandemic. *Journal of Quantative Cell Science*, *99*(1), 51–59. https://doi.org/10.1002/cyto.a.24266

Hammond, J. M., Harvey, C. M., Koubek, R. J., Compton, W. D., & Darisipudi, A. (2005). Distributed collaborative design teams: media effects on design processes. *International Journal of Human-Computer Interaction*, *18*(2), 145–165.

https://doi.org/10.1207/s15327590ijhc1802

Hatem, W. A., Kwan, A., & Miles, J. (2012). Comparing the effectiveness of face to face and computer mediated collaboration. *Advanced Engineering Informatics*, *26*(2), 383–395. https://doi.org/10.1016/j.aei.2012.01.001

Patterson, E. S., Bernal, F., & Stephens, R. (2012). Differences in macrocognition strategies with face to face and distributed teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 282–286. https://doi.org/10.1177/1071181312561066

Pazos, P., & Beruvides, M. G. (2011). Performance patterns in face-to-face and computer-supported teams. *Team Performance Management*, *17*, 83–101. https://doi.org/10.1108/13527591111114729

Post, W. M., Van Den Boogaard, S. A. A., & Rasker, P. C. (2004). Distributed trouble-shooting. In F. Darses, R. Dieng, C. Simone, & M. Zacklad (Eds.), *Cooperative Systems Design: Scenario-based Design of Collaborative Systems* (pp. 55–68). Amsterdam: IOS Press.

Rinta-kahila, T., Pentinnen, E., Salovaara, A., & Soliman, W. (2018). Consequences of discontinuing knowledge work automation – Surfacing of deskilling effects and methods of recovery. *Proceedings of the 51st Hawaii International Conference on System Sciences*, 5244–5253.

Salimian, H., Brooks, S., & Reilly, D. (2019). Immersed and Integral : Relaxing View Congruence for Mixed Presence Collaboration With 3D Content. *Proceedings of ACM Human-Computer Interaction*, *3*, 1–22.

Stone, N. J., & Posey, M. (2008). Understanding coordination in computer-mediated versus face-to-face groups. *Computers in Human Behavior*, *24*(3), 827–851. https://doi.org/10.1016/j.chb.2007.02.014

Tang, J. C., & Isaacs, E. (1993). Why do users like video ? Studies of multimedia-supported collaboration. *Computer Supported Cooperative Work (CSCW)*, *1*, 163–196.

Tang, J. C., Isaacs, E. A., & Rua, M. (1994). Supporting distributed groups with a montage of lightweight interactions. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 23–34.

Veinott, E., Olson, J., Olson, G. M., & Fu, X. (1999). Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (May), 302–309. https://doi.org/10.1145/302979.303067

Wheeler, B. C., Valacich, J. S., & Alavi, M. (1995). The emergence of desktop videoconferencing for collaborative work. *Proceedings of the Twenty-Eighth Annual Hawaii International Conference on System Sciences*, 808–817. IEEE.

Zoric, G., Domova, V., Ralph, M., Vartiainen, E., Björndal, P., & Muñoz, A. A. (2016). Supporting maritime remote experts working over distance. *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, *October*, 1–6. https://doi.org/10.1145/2971485.2995347

Appendix I – Puzzle ‘Wires’ with manual section

An example of the puzzle itself can be seen in Figure 6. The puzzle can consist of three to six wires that can be coloured either red, white, yellow, black, or blue. The participant has to cut one of the wires. The manual section can be found in Figure 7, where the participant had to read in the right text box depending on the number of wires and then had to act on the first description that applied. The serial number was provided to the participant above the puzzle. The correct answer to this specific puzzle would have been to cut the second wire.

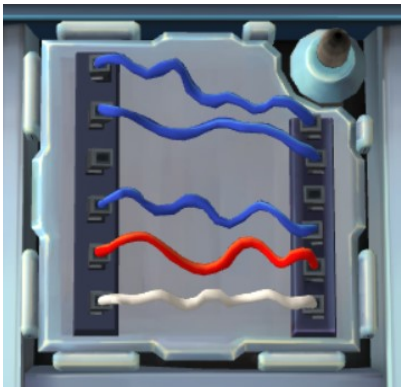
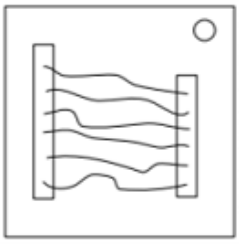


Figure 6. Example of puzzle 'Wires'

On the Subject of Wires



- A wire module can have 3-6 wires on it.
- Only the one correct wire needs to be cut to disarm the module.
- Wire ordering begins with the first on the top.

3 wires:
If there are no red wires, cut the second wire.
Otherwise, if the last wire is white, cut the last wire.
Otherwise, if there is more than one blue wire, cut the last blue wire.
Otherwise, cut the last wire.

4 wires:
If there is more than one red wire and the last digit of the serial number is odd, cut the last red wire.
Otherwise, if the last wire is yellow and there are no red wires, cut the first wire.
Otherwise, if there is exactly one blue wire, cut the first wire.
Otherwise, if there is more than one yellow wire, cut the last wire.
Otherwise, cut the second wire.

5 wires:
If the last wire is black and the last digit of the serial number is odd, cut the fourth wire.
Otherwise, if there is exactly one red wire and there is more than one yellow wire, cut the first wire.
Otherwise, if there are no black wires, cut the second wire.
Otherwise, cut the first wire.

6 wires:
If there are no yellow wires and the last digit of the serial number is odd, cut the third wire.
Otherwise, if there is exactly one yellow wire and there is more than one white wire, cut the fourth wire.
Otherwise, if there are no red wires, cut the last wire.
Otherwise, cut the fourth wire.

Figure 7. Manual section for 'Wires'

Appendix II – Puzzle ‘Keypads’ with manual section

An example of the puzzle itself can be seen in Figure 8. The puzzle consists of four keypads that has a symbol on it. The participant has to figure out in which order they need to press the keys in order to solve the puzzle. The manual section can be found in Figure 9, where it is explained that all four symbols should be found in one specific row. Their order of occurrence is simultaneously the order in which the keys must be pressed. The correct answer to this puzzle would have been: 

Figure 8. Example of puzzle 'Keypads'

Figure 9. Manual section for 'Keypads'

Appendix III – Puzzle ‘The Button’ with manual section

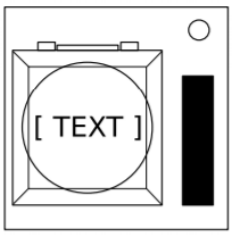
An example of the puzzle itself can be seen in Figure 10. The puzzle consists a button of which the colour and text can vary. When reading the manual, the participant would have to follow the first rule that applied, which would lead them to either press and immediately release the button or hold it and refer to an additional section in the manual. The additional section of the manual then explained when the participant was supposed to release the button depending on which colour the strip next to it turned. The main manual section can be found in Figure 11 and the additional manual section can be found in Figure 12.

The number of batteries, whether the two possible indicators FRK and CAR are lit or not, and the colour the strip turns when holding the button are important to finding the solution. The additional information was presented to the participant above the puzzle.

For the purpose of this example the number of batteries will be two and there are no lit indicators. The correct answer to this puzzle is that it should be held. The strip turns white and when referring to the additional section it shows that it should then be released when the countdown timer has a 1 in any of the positions.

Figure 10. Example of puzzle 'The Button'

On the Subject of The Button



Follow these rules in the order they are listed. Perform the first action that applies:

1. If the button is blue and the button says "Abort", hold the button and refer to "Releasing a Held Button".
2. If there is more than 1 battery on the bomb and the button says "Detonate", press and immediately release the button.
3. If the button is white and there is a lit indicator with label CAR, hold the button and refer to "Releasing a Held Button".
4. If there are more than 2 batteries on the bomb and there is a lit indicator with label FRK, press and immediately release the button.
5. If the button is yellow, hold the button and refer to "Releasing a Held Button".
6. If the button is red and the button says "Hold", press and immediately release the button.
7. If none of the above apply, hold the button and refer to "Releasing a Held Button".

Figure 11. Main manual section for 'The Button'

Releasing a Held Button

If you start holding the button down, a colored strip will light up on the right side of the module. Based on its color, you must release the button at a specific point in time:

- Blue strip: release when the countdown timer has a 4 in any position.
- White strip: release when the countdown timer has a 1 in any position.
- Yellow strip: release when the countdown timer has a 5 in any position.
- Any other color strip: release when the countdown timer has a 1 in any position.

Figure 12. Additional manual section for 'The Button'

Appendix IV – Transfer Task

The manual for the transfer task can be seen in Figure 13, and an example of the transfer task itself can be seen in Figure 14. The manual was presented above the task itself on the page that the participants saw. The participants had to decide for each separate wire whether they would cut it or not by reading off the Venn diagram in the manual. When the participant found the right position in the Venn diagram where all the wire's conditions (red or not, blue or not, star or not, and led or not) overlapped, the diagram showed a letter which corresponds with which action should be taken.

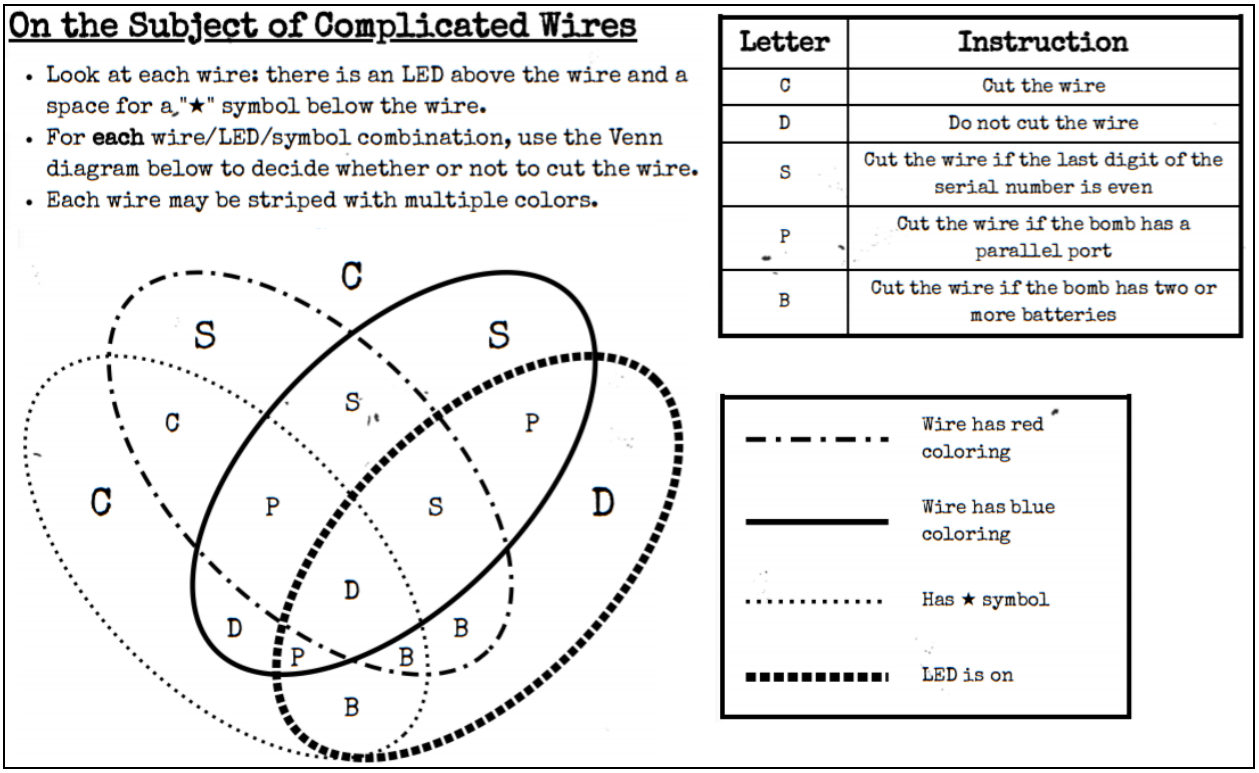
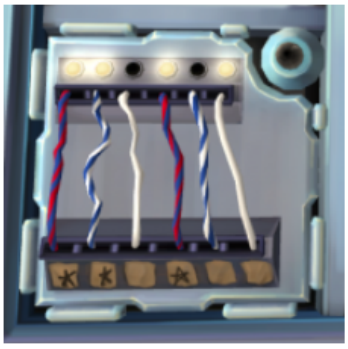


Figure 13. Manual section for the transfer task

For this particular puzzle (Figure 14) the solution would be to cut the second and third wire and leave the rest of the wires uncut.

Serial number: TH2JB9
Ports: Parallel and DVI-D
Batteries: 1



Which wire(s) do you cut? (from left to right; if the wire shouldn't be cut do not check the box)

- ☐ Cut the first wire
- ☐ Cut the second wire
- ☐ Cut the third wire
- ☐ Cut the fourth wire
- ☐ Cut the fifth wire
- ☐ Cut the sixth wire

Figure 14. Example of the transfer task in the questionnaire

Appendix V – Example of the Questionnaire

Examples of the different puzzles are shown below. In the conditions where the participant was working on their own with the manual, the correct section of the manual was also given above the puzzle.

Wires

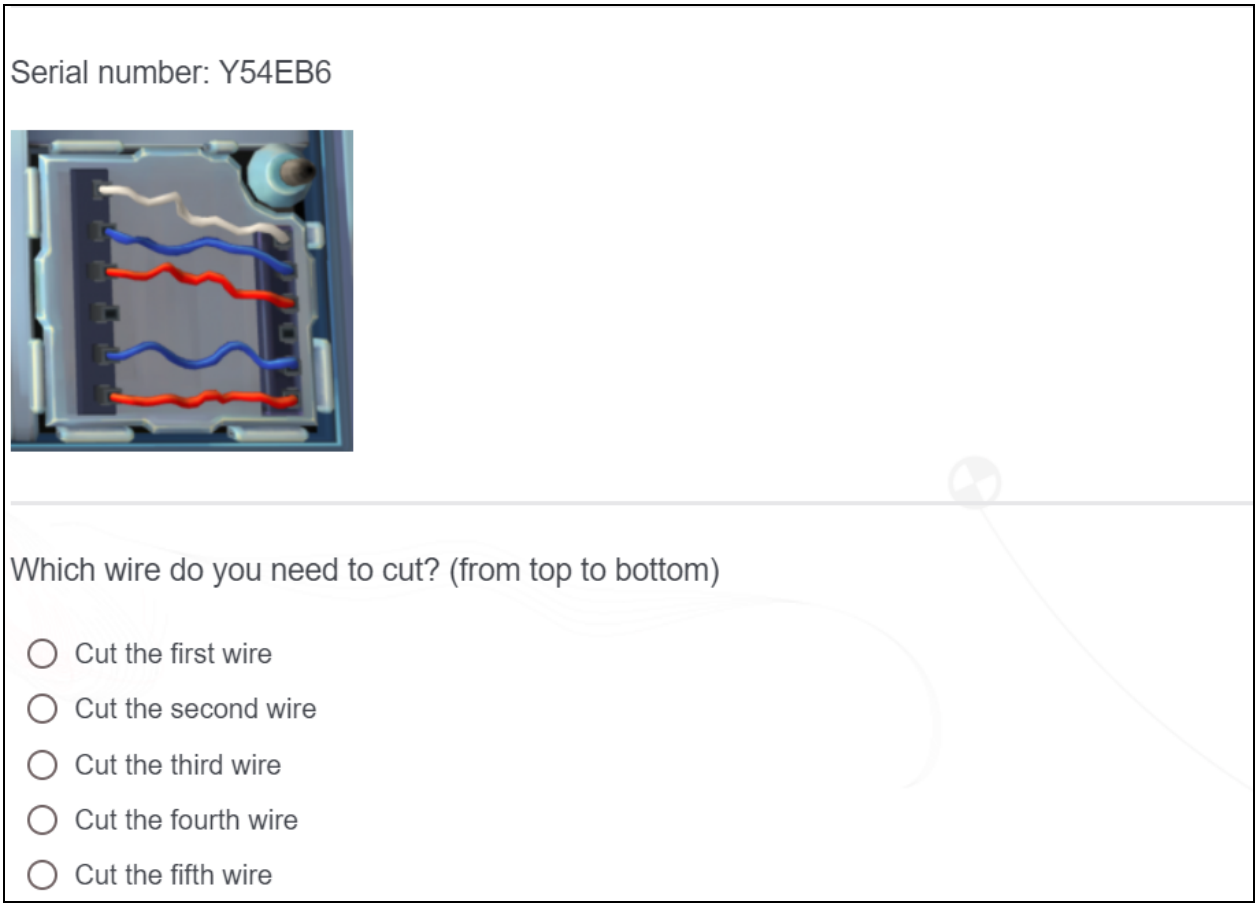


Figure 15. Example of 'Wires' in the questionnaire

Keypads

In what order will you click the symbols? Drag and drop the symbols into the right order (with the first being the first you'd click, and so on).

ɷ

¿

ɔ

ɜ̈

Figure 16. Example of 'Keypads' in the questionnaire

The Button

The bomb has 5 batteries, and two lit indicators labelled CAR and FRK

What action will you take with the button?

- ○ Press and immediately release
- ○ Hold button and refer to "Releasing a Held Button" (next step is explained on next page of this questionnaire)

Figure 17. Example of the main part of 'The Button' in the questionnaire

If the participant selected ‘Press and immediately release’ as the answer, they would immediately go on to the next puzzle. If they selected ‘Hold button and refer to “Releasing a Held Button” (next step is explained on the next page of this questionnaire)’ as an answer, they would be led to the question below:

The strip turns yellow, what will you do?

- ○ Release when the countdown timer has a 1 in any position
- ○ Release when the countdown timer has a 4 in any position
- ○ Release when the countdown timer has a 5 in any position

Figure 18. Example of the additional part of 'The Button' in the questionnaire

This question last part of the question (the strip part) was not included in the final time per trial and the awareness check that was conducted, because whether participants answered this question was dependent on what they had answered on the first part of the question (the button). Thus, if people had incorrectly answered the button it seemed unfair to get a correct score for the second part of the question. For the awareness check this did not really check awareness because it was literally one sentence that already indicated all the information.

Appendix VI – Informed Consent

Below the consent form in both English and Dutch are shown. At the end of each question (yes/no) or (ja/nee) is indicated to show that they all had to be individually answered.

English Version

Dear participant,

Thank you for participating in my study into remote expert support.

Important! You can only participate in this study if you:
- Are aged between 18 and 65 years old.

If this is not the case, I would like to ask you not to participate in this study.

This research is part of my thesis project in the master "Human Factors & Engineering" at the University of Twente. The purpose of this research is to investigate whether remote expert support helps or hinders the completion of certain tasks.

In this study, different demographic aspects will be asked to provide an overview of the participants included in the research. Then you will have to complete different tasks in collaboration with an expert or alone. Video and audio recordings will be made during the experiment. This is done to accurately keep track of how the experiments proceed. No video or audio clips will be used in the report of this research. Only the researcher has access to these audio and video recordings, and they will be destroyed once the research is completed.

You can withdraw from the study at any time without indicating your reason. Withdrawing from the study will not have any consequences. The data that is collected will be processed anonymously. This means that the results cannot be traced back to you after the completion of the experiment. If you have any comments and/or questions after the research, please contact me at r.sengler@student.utwente.nl.

There are no known risks associated with this research study; however, as with any online related activity, the risk of a breach is always possible. To the best of our ability, your answers in this study will remain confidential. Any risks will be minimized by safely storing the data and confidentiality will be secured by anonymizing the data before storage.

If you want to participate in the study, you can indicate this on the next page by agreeing to the informed consent. If you decide not to participate in this study I would, regardless, like to thank you for your time.

Thank you very much for your participation!

Rosan Sengler
Master student "Human Factors & Engineering"
University of Twente

I have read and understood the study information, or it has been read to me. I have been able to ask questions about the study and my questions have been answered to my satisfaction. (yes/no)

I consent voluntarily to be a participant in this study and understand that I can refuse to answer questions and I can withdraw from the study at any time, without having to give a reason. (yes/no)

I understand that taking part in the study involves an experiment, audio-and-video recording, and the completion of this questionnaire. (yes/no)

I understand that information I provide will be used for the purpose of this thesis, and this thesis only. (yes/no)

Dutch Version

Beste deelnemer,

Bedankt dat u wil deelnemen aan mijn onderzoek naar expert ondersteuning op afstand.

Belangrijk: u kunt alleen deelnemen aan dit onderzoek als u:
- Tussen de 18 en 65 jaar oud bent.

Mocht dit niet het geval zijn, wil ik u vragen uw deelname aan dit onderzoek te staken.

Dit onderzoek is onderdeel van het masterproject van de studie 'Human Factors & Engineering' aan de Universiteit Twente. Het doel van dit onderzoek is te onderzoeken of ondersteuning van een expert op afstand het voltooien van een specifieke taak kan helpen of hinderen.

In dit experiment zullen verschillende demografische aspecten uitgevraagd worden om een beeld te scheppen van de deelnemers in dit onderzoek. Vervolgens zult u een paar taken moeten voltooien in samenwerking met de expert of alleen. Tijdens het experiment zullen er video en audio opnames gemaakt worden. Dit wordt gedaan om accuraat bij te kunnen houden hoe de experimenten verlopen. Er zullen geen beeld- of geluidsfragmenten gebruikt worden in de rapportage van dit onderzoek. Alleen de onderzoeker heeft inzage tot de betreffende audio en video opnames, en deze zullen vernietigd worden zodra het onderzoek afgerond is.

Tijdens het onderzoek kunt u op elk moment stoppen met de deelname, zonder hiervoor een reden te geven. Stoppen tijdens het onderzoek heeft geen enkele consequentie. De verzamelde gegevens worden anoniem verwerkt. Dat betekent dat de resultaten na het onderzoek niet meer terug te leiden zijn naar u. Als u na afloop van het onderzoek opmerkingen en/of vragen heeft, kunt u contact met mij opnemen via r.sengler@student.utwente.nl.

Er zijn geen risico's bekend bij dit onderzoek; echter, zoals bij elke activiteit online, is het risico van inbreuk altijd mogelijk. Voor zover mogelijk, zullen de antwoorden in deze studie vertrouwelijk blijven. Eventuele risico's worden geminimaliseerd door gegevens veilig op te slaan en de vertrouwelijkheid wordt gewaarborgd door de gegevens vóór opslag te anonimiseren.

Als u mee wil doen aan het onderzoek, kunt u dit op de volgende pagina aangeven door akkoord te gaan met de toestemmingsverklaring. Indien u besluit niet deel te nemen bedank ik u voor uw tijd.

Bij voorbaat hartelijk dank voor uw medewerking!

Rosan Sengler

Master student "Human Factors & Engineering"
Universiteit Twente

Ik heb studie-informatie gelezen, of deze is aan me voorgelezen, en ik begrijp het. Ik heb al mijn vragen kunnen stellen over de studie en mijn vragen zijn naar tevredenheid beantwoord. (Ja/nee)

Ik geef vrijwillig toestemming om deel te nemen aan dit onderzoek en begrijp dat ik kan weigeren vragen te beantwoorden en dat ik mij op elk moment kan terugtrekken uit het onderzoek, zonder uitleg te hoeven geven. (Ja/nee)

Ik begrijp dat deelname aan het onderzoek een experiment, audio- en video-opname en het invullen van een vragenlijst inhoudt. (Ja/nee)

Ik begrijp dat de informatie die ik verstrek zal worden gebruikt voor het doel van deze scriptie, en alleen voor deze scriptie. (Ja/nee)

Appendix VII – Awareness Check

The table below (Table 8) shows the awareness check that was completed on each puzzle that the participant completed in phase video-audio phase and audio-only phase of the supported condition. The elements stated here are the minimum requirement for solving the puzzle. For the Wires puzzle the participants could also read out the serial number, but this was only necessary in a couple of situations. Participants oftentimes noticed soon enough that it was only necessary to tell the serial number if they were asked for it, therefore, it is not considered for this check.

Table 8. Awareness check per puzzle

Item	Question	Check?
Wires	Number of wires	
	Colour of wires	
Symbols	Describe the symbols	
Button	Colour of button	
	Text on button	
	Number of batteries	
	Lit indicators	

Appendix VIII – Time on Task Throughout the Trials

The graph (Figure 19) shows the combination of the unsupported condition and the supported condition when it comes to time it took to complete each trial as they progressed through the trials in the individual phase. It shows that the more trials the participants had completed they generally took less time to complete the following trial.

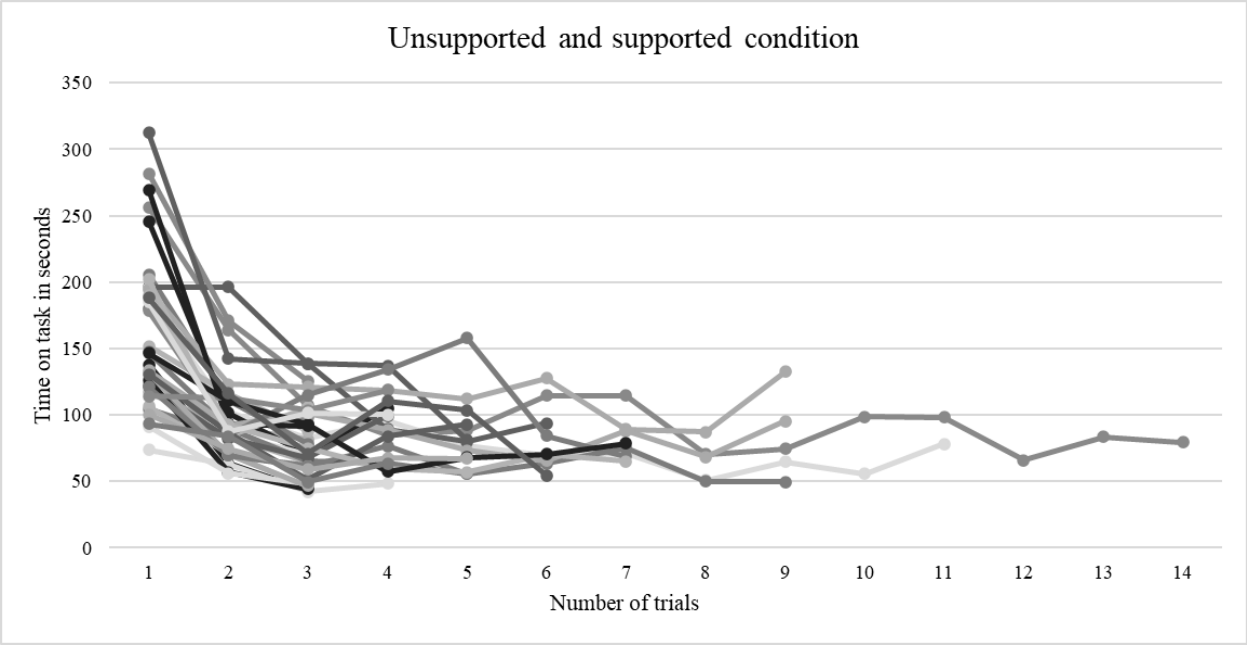


Figure 19. Time on task per trial in the unsupported and supported condition

Appendix IX – Matching Pairs Statistical Analysis

Time on Task in Individual Phase

A statistical test was performed on time on task during the first three trials of the individual phase. After the first three trials the number of participants that had done any further trials steadily declines, because each participant that had reached peak performance moved on to the next phase. Fewer data points to compare affects the statistical power, thus, it was chosen to only look at the first three trials.

The normality of the time on task during the individual phase was tested per trial and condition using the Shapiro-Wilk's test, and the results can be found in Table 9. As trials 1 and 2 had normally distributed data in both of the conditions, the data was analysed using an independent samples t-test. Trial 3 only had normally distributed data in one of the conditions, so it was analysed using a Mann-Whitney U-test.

Table 9. Shapiro-Wilk's test *p*-values on time on task in the individual phase with matched pairs

Trial	Unsupported condition	Supported condition
1	.47	.08
2	.30	.06
3	.43	.03

The independent samples t-test showed a significant difference between the unsupported condition (M = 198.23, SD = 56.90) and the supported condition (M = 126.21, SD = 26.69) for trial 1 ($t(16)$= 3.44, p = .00, Hedge's g = 1.62); and a significant difference between the unsupported condition (M = 111.71, SD = 39.10) and the supported condition (M = 79.43, SD = 22.46) for trial 2 ($t(16)$= 2.15, p = .02, Hedge's g = 1.01). Furthermore, the Mann-Whitney U-test showed a significant difference between the unsupported condition (Mdn = 81.47) and the supported condition (Mdn = 60.95) for trial 2 (U = 21.00, p = .05, η^2 = .17). Thus, as the

supported condition consistently showed a lower mean or median time on task, they required less time to complete the tasks.

Time on Task in Transfer Task

The normality of the time on task during the transfer task was tested per condition using the Shapiro-Wilk's test, and the results can be found in Table 10. All but trial 4 are normally distributed and will be analysed with an independent samples t-test. Trial 4 will be analysed using a Mann-Whitney U-test.

Table 10. Shapiro-Wilk's test *p*-values on time on task in the transfer task with matched pairs

Trial	Unsupported condition	Supported condition
1	.22	.14
2	.56	.09
3	.30	.87
4	.36	.00
5	.26	.74

Table 11 and Table 12 show the results from the independent samples t-tests on trial 1, 2, 3, and 5. As can be seen, the only trial that showed a significant difference between the unsupported condition (M = 156.93, SD = 84.78) and the supported condition (M = 93.02, SD = 44.88) was trial 3 (Hedge's g = .94). The trial showed that the supported condition was faster than the unsupported condition. This outcome is similar to the findings of the statistical test that was performed over the entire sample for time on task in the transfer task.

Table 11. Means and standard deviations of time on task in transfer task per trial and condition with matched pairs

Trial	Condition	Mean	Std. Deviation
1	Unsupported	230.96	91.28
	Supported	182.77	71.27
2	Unsupported	103.02	62.54
	Supported	96.08	40.93
3	Unsupported	156.93	84.78
	Supported	93.02	44.88
4	Unsupported	98.18	46.71

	Supported	77.38	59.39
5	Unsupported	87.71	36.63
	Supported	73.26	29.77

Table 12. Independent samples t-tests of time on task in transfer task per trial with matched pairs

Trial	df	t	Sig. (1-tailed)
1	16	1.25	.12
2	16	.28	.39
3	12.16	2.00	.03
5	16	.92	.19

The Mann-Whitney U-test showed a non-significant difference between the unsupported condition (*Mdn* = 78.94) and the supported condition (*Mdn* = 56.93) for trial 4 ($U = 25.00$, $p = .10$).

Number of Correct Answers in Transfer Task

The normality of the number of correct answers during the transfer task was tested per condition using the Shapiro-Wilk's test, and the results can be found in Table 13. As can be seen none of the trials had data where both conditions were normally distributed, thus, all the data was analysed using a Mann-Whitney U-test.

Table 13. Shapiro-Wilk's test *p*-values on number of correct answers in the transfer task with matched pairs

Trial	Unsupported condition	Supported condition
1	.27	.04
2	.00	.03
3	.00	.11
4	.00	.00
5	.01	.02

The results from the statistical tests can be found in Table 14 and Table 15 below. As can be seen, none of the results showed a significant difference between the two conditions. This

finding is similar to the statistical test that was conducted on the entire sample for number of correct answers in the transfer task.

Table 14. Medians of correct answers in the transfer task per trial and condition with matched pairs

Trial	Condition	Median
1	Unsupported	5
	Supported	5
2	Unsupported	6
	Supported	5
3	Unsupported	6
	Supported	5
4	Unsupported	6
	Supported	6
5	Unsupported	5
	Supported	5

Table 15. Mann-Whitney U-tests of correct answers in transfer task per trial with matched pairs

Trial	*U*	Sig. (1-tailed)
1	37.50	.40
2	28.50	.15
3	32.50	.24
4	36.50	.37
5	32.00	.24

Appendix X – Correct Participants Statistical Analysis

Time on Task in Individual Phase

A Shapiro-Wilk's test was conducted to test the normality of the sample. Its results (Table 16) show that only trial 1 had normally distributed data in both conditions and can be analysed using an independent samples t-test. Trial 2 and 3 will therefore be analysed using a Mann-Whitney U-test.

Table 16. Shapiro-Wilk's test *p*-values on time on task in the individual phase with most correct participants

Trial	Unsupported condition	Supported condition
1	.43	.21
2	.03	.28
3	.76	.01

The independent samples t-test found a significant difference between the unsupported condition (M = 204.20, SD = 53.46) and the supported condition (M = 123.63, SD = 31.83) for trial 1 ($t(28)$= 5.02, p = .00, Hedge's g = 1.83). Furthermore, the Mann-Whitney U-test showed a significant difference between the unsupported condition (Mdn = 93.34) and the supported condition (Mdn = 74.43) for trial 2 (U = 41.00, p = .00, η^2 = .30); and a significant difference between the unsupported condition (Mdn = 91.56) and the supported condition (Mdn = 58.96) for trial 3 (U = 43.00, p = .00, η^2 = .29). These results are similar the statistical test that was conducted on the entire sample for the time on task in the individual phase.

Time on Task in Transfer Task

A Shapiro-Wilk's test was conducted to test the normality of the sample. Its results (Table 17) show that only trial 2 and 3 had normally distributed data in both conditions and can be analysed using an independent samples t-test. Trial 1, 4 and 5 will therefore be analysed using a Mann-Whitney U-test.

Table 17. Shapiro-Wilk's test *p*-values on time on task in the transfer task with most correct participants

Trial	Unsupported condition	Supported condition
1	.04	.00
2	.60	.72
3	.24	.59
4	.08	.01
5	.00	.92

The independent samples t-test on trial 2 showed a non-significant difference between the unsupported condition (M = 115.10, SD = 55.74) and the supported condition (M = 105.25, SD = 38.84), $t(28) = .56$, $p = .29$; and a significant difference between the unsupported condition (M = 150.74, SD = 76.33) and the supported condition (M = 100.43, SD = 41.01), $t(21.46) = 2.25$, $p = .02$., Hedge's $g = .82$.

Additionally, the Mann-Whitney U-tests (Table 18 and Table 19) on trial 1, 4 and 5 showed no significant differences between the conditions. These results, that only trial 3 shows a significant difference where the supported condition was generally faster than the unsupported condition, are similar the statistical test that was conducted on the entire sample for the time on task in the transfer task.

Table 18. Medians of time on task in transfer task per trial and condition with most correct participants

Trial	Condition	Median
1	Unsupported	219.26
	Supported	196.15
4	Unsupported	74.60
	Supported	62.73
5	Unsupported	84.80
	Supported	79.36

Table 19. Mann-Whitney U-tests of time on task in transfer task per trial with most correct participants

Trial	*U*	Sig. (1-tailed)
1	110.00	.47
4	79.00	.09
5	85.00	.13

Number of Correct Answers in Transfer Task

The normality of the number of correct answers the transfer task was tested per condition using the Shapiro-Wilk's test, and the results can be found in Table 20. The normality tests showed that none of the data was normally distributed, thus, all the trials will be analysed using a Mann-Whitney U-test.

Table 20. Shapiro-Wilk's test *p*-values on number of correct answers in transfer task with most correct participants

Trial	Unsupported condition	Supported condition
1	.01	.01
2	.00	.00
3	.00	.01
4	.00	.00
5	.00	.00

The Mann-Whitney U-tests (Table 21 and Table 22) showed no significant differences between the condition in each trial. These results are similar the statistical test that was conducted on the entire sample for the number of correct answers in the transfer task.

Table 21. Medians of number of correct answers in transfer task with most correct participants

Trial	Condition	Median
1	Unsupported	5
	Supported	5
2	Unsupported	6
	Supported	6
3	Unsupported	6
	Supported	5
4	Unsupported	6
	Supported	6
5	Unsupported	6
	Supported	6

Table 22. Mann-Whitney U-test on number of correct answers in transfer task with most correct participants

Trial	*U*	Sig. (1-tailed)
1	103.50	.36
2	99.00	.30
3	77.00	.07
4	97.00	.27
5	110.00	.47