LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Department of Psychological and Behavioural Science

COURSEWORK SUBMISSION FORM AND PLAGIARISM/GENERATIVE AI/ACADEMIC HONESTY DECLARATION

Please ensure that a completed copy of this form is uploaded as part of your coursework submission.

Candidate Number: 34479

Course code: PB4D3

Word-count: 9965

Date: 17/08/2024

The Department wishes to draw your attention to the School Calendar Regulations on Assessment Offences and Plagiarism:

https://info.lse.ac.uk/Staff/Divisions/Academic-Registrars-Division/Teaching-Quality-Assurance-and-Review-Office/Assets/Documents/Calendar/RegulationsAssessmentOffences-Plagiarism.pdf

All work submitted as part of the requirements for any assessment of the School (e.g., examinations, essays, dissertations, and any other work, including computer programs), whether submitted for formative or summative assessment, must be expressed in your own words and incorporate your own ideas and judgments. Plagiarism must be avoided in all such work.

Plagiarism can involve the presentation of another person's thoughts or words as if they were your own. However, please note that plagiarism also includes self-plagiarism, which is where you as the author re-use your own previously submitted work or data in a "new" written piece of work without letting the reader know that this material has appeared elsewhere. The definition of "your own work" also includes work produced by collaboration or group-work expressly permitted by the Department.

Furthermore, it is prohibited to use any form of generative artificial intelligence in an unauthorised way when working on a summative assessment.

Please also note that plagiarism as defined by the School above, need not be deliberate or intentional for it to constitute an assessment offence.

Declaration (without signature, to preserve anonymity): Having read and understood LSE's guidelines on plagiarism/academic honesty, I hereby confirm by completing and attaching this form that the work submitted is my own. By submitting this form I hereby confirm I understand the Department's policy on summative assessment, and that I have read the relevant parts of the MSc programme handbook.

Shall we meet in VR:
Could virtual reality be a better alternative to videoconferencing?

DISSERTATION

Submitted in partial fulfilment of the requirements for

MSc Behavioural Science

Student ID: 34479

Supervised by: Dr. George Melios

The London School of Economics and Political Science

Department of Psychological and Behavioural Sciences

August 2024

Abstract

Since the coronavirus pandemic, working from home has become increasingly prevalent, leading to challenges such as reduced social interactions, diminished sense of belonging, reduced collaboration, and 'Zoom fatigue' among remote workers. Conducting some of the group activities in virtual reality could potentially address these issues. This study contributes to the very limited body of literature, investigating the effects of meetings in virtual reality (VR) on 1) group dynamics, 2) group performance and 3) levels of tiredness caused by the technology. The study involved a lab, between-subjects, randomised control experiment with 54 LSE students and took place at the LSE Behavioural Lab. Participants completed the Desert survival task, a classic exercise in the study of group performance and dynamics, both individually and in small groups (of up to 3), and answered questionnaires before and after the task. Participants were split across two conditions based on the technology medium used for the group interactions - virtual reality and videoconferencing.

Due to its small sample size, this study provides inconclusive results regarding the effect of VR meetings on group dynamics and users' levels of tiredness. However, in contrast to the existing literature, it finds that groups interacting in VR perform worse than those using videoconferencing. The consistency of these results was confirmed through robustness checks. The inferior performance of groups in the VR condition could be attributed to participants' unfamiliarity with the novel technology and the technical issues faced by groups, likely disrupting their discussions. These findings highlight the importance of training if VR technology is to be widely implemented across organisations. Still further research on the topic is required and this study outlines the exciting research path ahead.

Keywords: virtual reality, videoconferencing, remote work, remote meetings, group dynamics, group performance, tiredness

Acknowledgements

"It takes a village to raise a child…"

… and so does academic research!

When I initially decided to explore the impact of VR on group interactions, I was warned by many about how difficult group experiments are, especially without any funding. Having previously conducted an experimental study for my undergraduate dissertation, I wondered, *'How difficult could it be?'*

There were challenges, but I am beyond grateful for the incredible support provided by the LSE.

I would like to extend my warmest appreciation to my dissertation supervisor, Dr. George Melios, for this guidance and critical feedback; to Prof. Liam Delaney for his valuable insights and thought-provoking questions; to Dr. Dario Krpan for his insightful lectures and his above-and-beyond helpfulness; to Dr. Laura Guirge for her constructive feedback on my experimental design; and to Dr. Nicole Abi-Esber for sharing her experience and tips on running group studies.

Moreover, the execution could not have been possible without the support of Dr. Dina Rabie, managing the LSE Behavioural Lab, who not only helped setting up the rooms, but offered ideas and insights from her own research, and, with her care and enthusiasm, was the biggest cheerleader for my study. Nedelin Velikov from the LSE Digital Skills Lab was also instrumental for the smooth running of the experimental sessions, with his help troubleshooting technical issues at the lab before the commencement of the study and his support in resolving them.

I would also like to extend my gratitude to Jeremy Dalton, author of *Reality Check: How Immersive Technologies Can Transform Your Business* and Head of Immersive Technologies at PwC UK, for generously sharing his opinions on the future of VR as a meeting environment, which inspired the research question of this study.

Last but not least, I would like to extend my heartfelt gratitude to my mother for her unwavering support and passionate encouragement throughout my postgraduate journey; to my brother, who trusted me with his VR headset and provided round-the-clock technical support during the development of this study; and to my dear friend, PG, whose MSc in Applied Mathematics from the University of Cambridge finally paid off as he was able to answer all my statistics questions and reassure me that my analysis was on the right track.

Shall we meet in VR:
Could virtual reality be a better alternative to videoconferencing?

Table of Contents

1. Introduction

In recent years, following the coronavirus pandemic, remote work has become more common (Gould et al., 2023), with some organisations now operating fully remote workforces (Jasgur, 2023), while others are pursuing hybrid options, combining remote work with time in the office (Alexander et al., 2020). According to Statista (2024), as of last year, 28% of employees worldwide work from home all or most of the time, and that number is increasing. While remote and hybrid work has benefited employees by reducing commute costs and times, and providing flexible working hours (Park et al., 2023), it has also had some adverse effects such as decreased social interactions, sense of belonging, collaboration, work ethics and knowledge sharing (Kirchner et al., 2021; Yang et al., 2022). Moreover, many remote and hybrid workers have been reporting feelings of exhaustion and impaired well-being attributed to the prolonged use of videoconferencing (the so-called 'Zoom fatigue') (Elbogen et al., 2022; Nesher Shoshan & Wehrt, 2022).

Could there be a way to combine the benefits of remote working and the collaborative atmosphere of physical office environments, while also reducing exhaustion?

Advances in immersive technologies, like virtual reality (VR), may hold the answer. Those are immersive virtual environments that users can interact with using specialised headsets (Yenduri et al., 2024). While there has been extensive research on the applications of VR to education and medical training for more than a decade, (Kyaw et al., 2019; Merchant et al., 2014; Yu, 2023), studies targeting knowledge workers, and particularly those working remotely, have been limited. This is not surprising, given that remote work has only recently become prevalent (Gould et al., 2023) and that VR headsets have only recently achieved the quality needed for truly immersive experiences (Yenduri et al., 2024). As such, the application of VR and the Metaverse - a virtual environment in which users can interact through digital avatars (Lee et al., 2021) - to remote working represents a promising and developing area of research.

Through a lab experiment, this paper aims to examine the effect of meetings conducted in VR on the performance and dynamics of small groups, as well as the effect on individuals' levels of tiredness following those interactions.

The rest of this paper is structured as follows: Section 2 provides a review of the existing literature and forms the basis of research hypotheses. Section 3 outlines the methodology undertaken, including the study design, the procedure, and the framework used for the statistical analysis. Section 4 presents the results of the statistical analysis. Section 5 discusses the findings,

their limitations, and areas of future work. Section 6 summarises the contributions of this study and concludes the paper.

2. Literature Review

Despite the novelty of using VR as a meeting medium in the context of remote working, some trends have emerged in the available literature on its benefits and challenges.

Most studies find that groups using VR technologies as a meeting medium feel a better sense of 'togetherness' (social presence) and immersion compared to peers using videoconferencing (VC) (Campbell et al., 2020; Abramczuk et al., 2023; Aliman et al., 2023; Macchi & De Pisapia, 2024). This could be explained by the accurate body language tracking (Kurzweg et al., 2021), the spatial sound (Kobayashi et al., 2015), and the customizable avatars (Panda et al., 2022), all contributing to patterns of behaviour, such as gaze, longer dialogue overlap, and non-verbal communication, closer mimicking face-to-face interactions (Abdullah et al., 2021). Despite the enhanced social presence offered by VR technologies, there has been no established connection to improved team spirit, bonding, or connectedness (Abramczuk et al., 2023). Some studies have, however, found that groups interacting in VR tend to have more fun and look forward to future meetings (Aliman et al., 2023), as well as experience a more collaborative and peaceful environment (Macchi & De Pisapia, 2024). The effects of VR as a meeting medium on group dynamics, such as team belonging, trust, cooperation, shared goals, conflict resolution and team relationships, is therefore ambiguous and requires further research.

There is strong evidence, however, that the meeting environment does not significantly affect team performance, given the meeting medium is appropriate for the task (i.e. has the required materials or functionalities available). Studies across different contexts have found similar group performance levels, in terms of quality of group outputs, across interaction environments such as face-to-face, virtual reality and videoconferencing. Furumo & Pearson (2006) found no significant difference between students' graded outputs after interacting only over videoconferencing or face-to-face for two weeks. Abdullah et al. (2021) did not observe any performance differences between VR and videoconferencing on four consecutive tasks of different types on the McGrath's circumplex (McGrath, 1984). Macchi & De Pisapia (2024) did not discover significant differences in team performance and decision making efficacy in idea generation and prioritisation tasks across three conditions - face-to-face, VR and videoconferencing.

The case for remote teams moving some interactions to VR is, however, impeded by an established finding in the literature that VR headsets are uncomfortable to wear and can cause

fatigue and sickness symptoms for many users (Chang et al., 2020; Saredakis et al., 2020). Even though the 'Zoom fatigue' phenomenon is also well-established (Li & Yee, 2023; Nesher Shoshan & Wehrt, 2022), most studies do find higher levels of exhaustion for individuals using VR headsets compared to those interacting over VC (Abramczuk et al., 2023; Aliman et al., 2023; Hennig-Thurau et al., 2023), with some more severe negative effects, such as nausea, headache, disorientation and anxiety, appearing for a small number of participants (Weech et al., 2019; Biener et al., 2022; Abramczuk et al., 2023). Continued use though has the potential to ease those adverse effects (Biener et al., 2022; Hennig-Thurau et al., 2023; Macchi & De Pisapia, 2024). Kourtesis et al. (2019) also found that deeper immersion, better quality of graphics and sound, and shorter interactions (of maximum duration ~1h) reduce VR induced symptoms and effects. Still, wider adoption within teams should be carefully considered as the technology has the potential to be exclusionary for team members who are not comfortable with the VR headsets or experience severe symptoms (Abramczuk et al., 2023).

2.1. Gaps in literature

Not only is the literature on the effects of VR as a meeting environment limited, but the studies investigating the interactions of small groups are even more sparse. This is an important gap in the literature as small groups are usually the decision-making units in organisational settings (Arrow et al., 2000) and their interactions might differ from those of dyads or larger groups (Cooney et al., 2020; McGrath, 1984).

Moreover, most available studies of small group interactions in VR come from the field (Abramczuk et al., 2023; Hennig-Thurau et al., 2023; Macchi & De Pisapia, 2024), which offers higher ecological validity, but provides less control over confounding factors, such as pre-existing relationships among participants and team dynamics (Wilson et al., 2010). Most studies on the topic also follow a within-subject design (Abramczuk et al., 2023; Macchi & De Pisapia, 2024), which offers control over individual differences, but may lead to biases due to carry-over, order and demand effects (Charness et al., 2012).

Lab experiments with between-subject design provide a controlled and simple setup, in which the effects of VR as a meeting medium can be isolated and studied (List et al., 2011). Further research of this kind is required as the only such study (Abdullah et al., 2021) is focused on the difference in communication patterns rather than group performance or dynamics per se.

A question the wider social sciences literature does not clearly answer is how group processes and performance should be measured (De Jong et al., 2016; Feitosa et al., 2020; Grossman et al., 2022), however this is outside the scope of this paper.

2.2. Research question and hypotheses

Based on the existing literature, the following research question still remains open:

RQ: ***Could virtual reality be a better alternative to videoconferencing, by fostering more positive group dynamics, maintaining equivalent group performance, and avoiding increased levels of fatigue for individuals?***

The research question will be addressed in this paper by testing the following three hypotheses:

H_1: Individuals interacting over **virtual reality (VR) medium** will experience **more positive group dynamics**, relative to individuals interacting over videoconferencing (VC) medium.

H_2: Individuals interacting over **virtual reality (VR) medium** will achieve **better group performance** than individuals interacting over videoconferencing (VC) medium.

H_3: Individuals interacting over **virtual reality (VR) medium** will feel **less tired** at the end of the group interactions than individuals interacting over videoconferencing (VC) medium.

This study expects to find evidence for H_1 and evidence against H_2 and H_3.

3. Methodology

3.1. Experimental Design

To test the above hypotheses, a randomised controlled experiment at the LSE Behavioural Lab was conducted, after receiving ethical approval in line with the research ethics policy of the London School of Economics and Political Science (Grove, 2023). Using a between-subjects design, participants were randomly assigned to one of two conditions based on the technology media used - virtual reality (VR, treatment) or videoconferencing (VC, control). Both conditions included the same pre-task survey, individual task and post-task survey, and only differed by the hardware and meeting software used to complete the same task as a group (see Figure 1 for the experimental flow).

The study was conducted in a lab setting (rather than a field one) as such setting offers greater control over confounding factors (i.e. familiarity with other participants, tasks performed and their duration, meeting environment and distractions) and has higher internal validity (provides causal evidence) (Wilson et al., 2010).

Between-subject (opposed to within-subject) design was pursued as it avoids carry-over and order effects (interactions between treatments and effects resulting from the order of exposure) and

demand effects (participants changing their behaviour to act in accordance with their perception of the experimenter's expectations) (Charness et al., 2012).

Given that most of the limited, due to its novelty, prior research on the topic involves field studies with within-subject designs, this study offers clearer insights into the exact role of meeting technology media in group decision-making and dynamics.

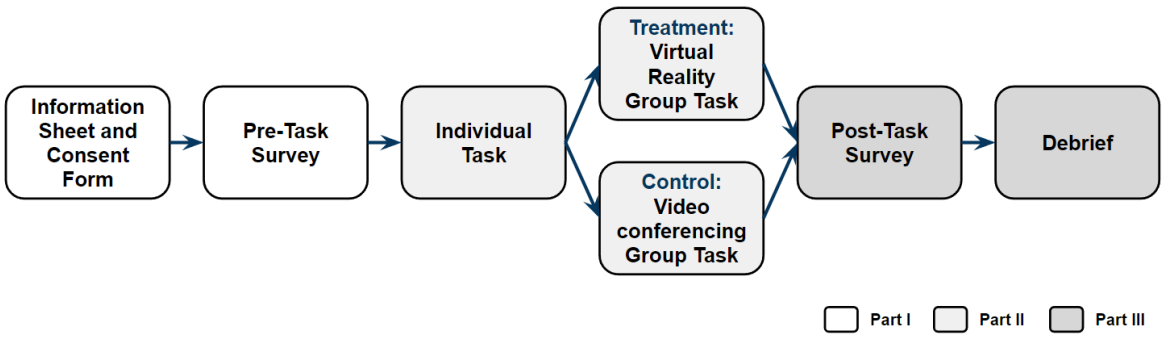


Figure 1. Experimental flow diagram

3.2. Sample requirements

Due to the logistical challenges of requiring ideally 3 (and a minimum of 2) participants attending each experimental session in person at the LSE Behavioural Lab, and the budget constraints of no external funding, only LSE students and staff were targeted for the study and no additional restrictions were applied to the participant pool.

A priori power analysis was conducted in G*Power 3.1.9.7 (Faul et al., 2007) to determine the minimum sample size of participants required for the study. The conventional values of significance level of 95% and a power requirement of 0.8 were used in the analysis (Perugini et al., 2018). Due to the novelty of the topic researched, there are limited comparable studies and data on effect size benchmarks. Data from a recent paper (Macchi & De Pisapia, 2024) points towards large effect sizes (0.9-1.0, Appendix A. Table A1), however due to the lack of an established benchmark a more conservative assumption of a medium effect size of 0.5 (Cohen, 1988) was used.

The results of the power analysis showed that with such a setup a total sample size of 128 participants was required with 64 individuals in each condition (Appendix A. Figure A1). The experiment design required groups of triads, so the sample size pursued was at least 132 participants (66 individuals across 22 groups for each condition), which was very optimistic given the study was not funded.

3.3. Recruitment

Recruitment for the study took place from the 20th of May to the 31st of May (which was the last day of the experiment). Due to budget constraints, participation was unpaid and recruitment was limited to LSE staff and students. It was conducted through various media:

- LSE WhatsApps Groups
- PBS Department Communications emails and Instagram posts
- Posters on LSE campus

Participants could sign up for the study via Calendly (a meeting scheduling website) and select a 30-minute time slot between the 23rd of May and the 31st of May (10am-4pm). In total, there were 20 time slots available and each had a maximum limit of three bookings. To avoid potential selection bias, conditions were assigned to time slots randomly before the study was advertised and participants recruited (Stigler, 1969).

The specific objective of the study was deliberately concealed during participant recruitment to reduce demand effects. The marketing of the study focused on the general topic of "group decision-making in technology mediated meetings" without mentioning any specific technologies (i.e. virtual reality). It was clarified that the experiment was conducted as part of a Masters dissertation and is taking place at the LSE Behavioural Lab, as affiliation with academic institutions provides credibility and helps recruitment efforts (Joseph et al., 2016). For transparency, the marketing materials emphasised that the experiment is not paid, but snacks will be provided to participants.

Given the group nature of the study, participants' punctuality and reliability was important for the successful running of each session. To avoid late-arrivals, last-minute cancellations and no-shows, an automated calendar invite was sent upon sign-up and reminders were shared with each participant 24 hours and 2 hours in advance of their time slot, asking them to arrive 5 minutes early for a prompt start (Mclean et al., 2014).

3.4. Experimental procedure

The study took place at the LSE Behavioural Lab between the 23rd and the 31st of May. Upon arrival of all expected participants, a session would begin with a short introduction from the researcher (myself) covering the format of the study, the equipment provided and brief instructions on how to proceed if any technical issues arise. Then, each participant was taken to a separate room which included a desk, two chairs and a laptop. For the VR Treatment condition, there would also be a Meta Quest 2 VR headset and two controllers in the room (Figure 2).

Figure 2. Treatment condition (VR) room setup

Each experimental session had three parts (Figure 1):

1. Providing information about the study, obtaining informed consent from participants, in accordance with the research ethics policy of the LSE (Grove, 2023), and collecting information on demographics, baseline measures of trust and tiredness, and use of prescription glasses;

2. Completion of Desert survival task (Lafferty & Pond, 1974) individually, and as a group in technology mediated meeting (Control - in Google Meet, Treatment - in MeetinVR);

3. Completion of post-task questionnaire, including information on perception of group dynamics, distribution of time across group task activities, prior experience with VR and a measure of tiredness; participants were also debriefed about the aims of the study.

Three attention checks were included throughout the experiment to ensure the quality of collected data (Abbey & Meloy, 2017).

3.5. Materials and Measures

3.5.1. Pre-task survey

Upon providing informed consent, participants were automatically directed to a short survey, collecting demographic information and other baseline data (Appendix B). The demographic questions referred to age, gender, ethnicity and level of education and were included to confirm sample balance across conditions. Additionally, ***age*** was used as a covariate in analysis as different age groups may have varying levels of comfort and familiarity with VR technology (Macchi & De Pisapia, 2024; Staddon, 2020). Other baseline information collected was:

- ***Use of prescription glasses***

Participants wearing prescription glasses might experience discomfort or blurrier vision when wearing a VR headset (Güzel et al., 2023).

- ***Level of tiredness***

Tiredness was measured on a 5-point Likert response format from 0 (Not tired at all) to 4 (Extremely tired). The measure was included as it might impact performance and engagement with the group task, and the experiment as a whole (Hockey, 2013).

- ***Baseline measures of Trust***

To capture individual differences among participants related to their propensity to trust, three statements based on modified items from the Trust scale (Costa & Anderson, 2011) and the German Socio-Economic Panel Study (Naef & Schupp, 2009) were included. The statements were purposefully vague to capture baseline trust attitudes toward strangers (Naef & Schupp, 2009), which might impact participants' behaviour in the group task (Dirks & Ferrin, 2001). Answers were measured on a 5-point Likert response format from -2 (Completely disagree) to 2 (Completely agree) for more intuitive interpretation of results, with answers of 0 indicating neutrality (Neither agree nor disagree). Only three statements were included to strike a balance between collecting baseline data and minimising the risk of revealing the study's aims and inducing demand effects, and survey fatigue (Stantcheva, 2023). The statements captured participants' self-assessments about *'relying on others'*, *'trusting others'*, and *'believing that others have good intentions'*.

One attention check was included in the pre-task survey to confirm participant's focus and to prompt them to pay attention as further checks might be included later in the study.

3.5.2. Desert survival task

Upon completion of the pre-task survey, participants were automatically directed to individually complete a modified version of the Desert Survival Problem (Lafferty & Pond, 1974), and then complete the same task as a group. The task was carefully selected as it had to be meaningfully executable individually and at the same time simulate an activity often performed by small groups in a work context. The Desert survival task is an intellective task (Type 3 on McGrath's circumplex (McGrath, 1984)) such that an objectively correct solution exists but is difficult to verify - a scenario very common in business and personal life (Littlepage et al., 1995) which made it an appropriate choice for this experimental study.

The task itself places participants in a desert plane crash scenario, where they have to decide on a course of action that will maximise their chances of survival, and accordingly rank 15 items (e.g. a mirror, a box of matches, a compass, etc.). The task has an optimal answer, ranking performed by survival experts (Appendix C), but individual expertise, though important, is not enough for the group to arrive (close) to that answer - group processes and dynamics are also key (Littlepage et al., 1995).

Based on the participants' individual rankings, an ***expertise score*** was calculated, measuring individual task performance. The score reflected the similarity of the individual rankings with those provided by survival experts (Lafferty & Pond, 1974):

$$(1)\ expertise\ score\ =\ 112\ -\ (\sum_{i=1}^{15} \left|individual\ ranking_i\ -\ expert\ ranking_i\right|),$$

where i indexes the items.

Given that the difference between individual and expert rankings represent errors (how distant is an answer from the optimal), the sum of the discrepancies was subtracted from the constant of 112 (the maximum discrepancy possible if all individual answers are furthest away from the expert answers), so higher scores reflect higher expertise (Littlepage et al., 1995).

Similarly, a ***group performance score (dependent variable)*** was calculated, measuring the output of group-decision making by capturing the difference between group and expert rankings:

$$(2)\ group\ performance\ score\ =\ 112\ -\ (\sum_{i=1}^{15} \left|group\ ranking_i\ -\ expert\ ranking_i\right|),$$

where i indexes the items.

Based on the expertise score, two additional variables were calculated - ***average individual score in group*** and ***best individual score in group***, to provide a baseline for group resources, existing knowledge within each group (Innami, 1994).

- **Videoconferencing (VC) - Control condition**

In the control condition, participants were asked to navigate to another open tab and complete the same Desert survival task, but as a group over Google Meet (Zoom-like video conferencing cloud-based software). They were instructed to keep their cameras and microphones on for the full duration of the meeting, and to spend up to 15 minutes to complete the task (see Figure 3 for screenshot of the setup).

- **Virtual reality (VR) - Treatment condition**

In the treatment condition, participants were first asked to watch tutorials to familiarise themselves with how to use the VR headset and controllers, and on how key MeetinVR (the VR meeting software used) functionalities work. Then, they were asked to adjust the VR headset to fit comfortably and read the further instructions provided on screens in the VR meeting environment (see Figure 4 for screenshot of the setup).

A hidden timer was included on the Qualtrics group task instruction page for both conditions to serve as an approximation of how long participants took to complete the task. An average from all individuals in a group was taken to derive the ***group task duration*** for each group. The group task instruction page also advised participants to return to it after completing the task to proceed with the next part of the study.

Technical issues observed during each session, such as participants needing help with a VR headset or software, were recorded and documented in a ***tech issues*** variable, as disruptions might impact task performance (Sell et al., 2013).

Due to limitations in the MeetinVR software, it was not feasible for participants to customise their avatars. Instead, the researcher (myself), created those while participants were completing Part I of the study, aiming to approximately match their looks.

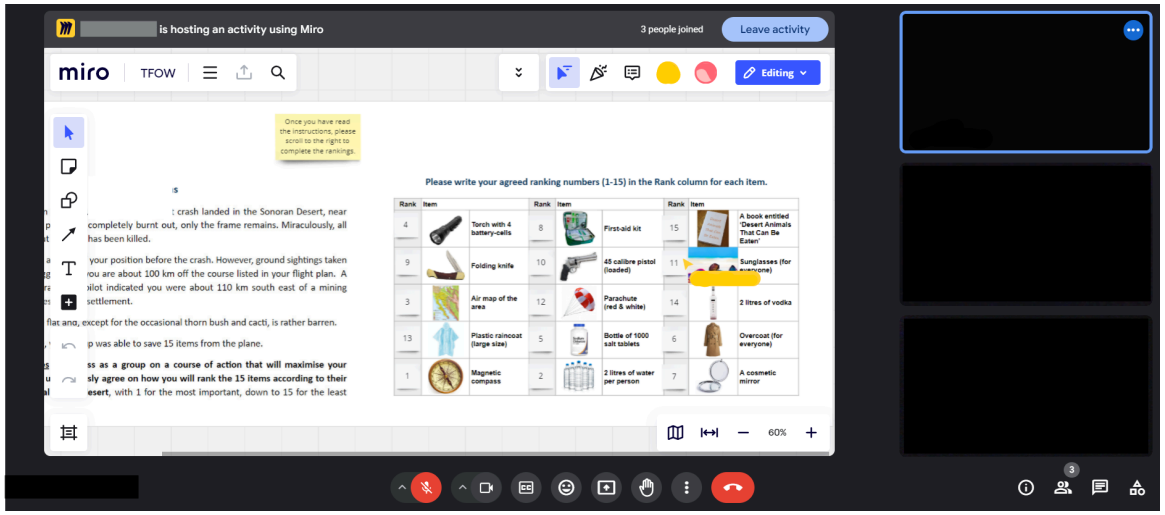


Figure 3. VC Group task setup (Control)

Figure 4. VR Group task setup (Treatment)

3.5.3. Post-task survey

Once participants had completed the group task and had navigated back to the Qualtrics open tab, they were asked to complete a post-task survey (Appendix B). The survey collected information on:

- ***Familiarity with other participants in the same group***

Group dynamics can vary between strangers and individuals who already know each other, as well as among mixed groups (Moreland & Levine, 2002). Therefore, data was collected on whether participants knew anyone else in their group and, if so, how many acquaintances they had.

- ***Level of tiredness***

The same question on tiredness from the pre-task survey was included to determine the levels of participants' tiredness at the end of the study and to be able to calculate the ***change in tiredness (dependent variable)***. Since most VR studies find that participants in VR conditions tend to experience higher levels of fatigue (Abramczuk et al., 2023; Aliman et al., 2023; Hennig-Thurau et al., 2023; Macchi & De Pisapia, 2024), it is important to validate the finding and factor it in the analysis.

- ***Finding group discussion enjoyable***

How enjoyable participants found the group discussion was measured on a 5-point Likert response format scale from 0 (Not at all enjoyable) to 4 (Extremely enjoyable). The question was included as having a positive experience of the group discussion might impact engagement and performance (Geue, 2018)..

- ***Feeling present during group discussion***

How present participants felt during the group discussion was measured on a 5-point Likert response format from 0 (Not at all present) to 4 (Extremely present). The question was included as feeling present might impact engagement and performance (Witmer & Singer, 1998).

- ***Time allocation during group discussion***

Participants were asked to report what percentage of the group discussion they spent on various activities (i.e. figuring out how the technology works, productive/unproductive conversation, etc., see full list in Table 1). The individual estimated percentage breakdowns were then averaged across each group, and a variable for ***time spent on each activity (in minutes)*** was calculated based on the total group task duration (in minutes) for that group.

Most previous studies on group dynamics in technology mediated environments have used observational methods to gather such insights (Abdullah et al., 2021; Macchi & De Pisapia, 2024). However, due to time limitations, budget constraints and additional ethical considerations related to obtaining session recordings, a time allocation self-report method (Gross, 1984) was pursued instead in this study. Nevertheless, this method offered insights into participants' perceptions of how the group discussions went, and the collected data was used to identify patterns and differences across the two conditions.

Table 1. List of suggested group discussion activities

#	Activity
1.	Trying to figure out how to make the tech work
2.	Introductions/Getting to know group members
3.	Productive conversation contributing towards the task completion
4.	Unproductive conversation (i.e. interrupting, arguing, repeating the same points, etc.)
5.	Other (Please specify)

- ***Prior VR experience***

A question was included to record familiarity with VR technology, as it might impact confidence using the technology and performance (Sagnier et al., 2020).

- ***Measurement of group dynamics***

To gain insights into the interactions during the group task and participants' perceptions of the group dynamics, 12 statements covering aspects such as team belonging, trust, cooperation, shared goals, conflict resolution and team relationships were included (see Table 2 for full list of statements). Those statements included versions of the trust measures from the pre-task survey and additional items from the Trust scale (Costa & Anderson, 2011), the German Group Development Questionnaire (Leuteritz et al., 2022) and the Global Satisfaction Scale (Hamlyn-Harris et al., 2006). Answers were measured on a 5-point Likert response format from -2 (Completely disagree) to 2 (Completely agree) for more intuitive interpretation of results, with answers of 0 indicating neutrality (Neither agree nor disagree). Only 12 statements were included due to time constraints related to the unpaid nature of this study, and to minimise the risk of survey fatigue (Stantcheva, 2023). Moreover, two attention checks were incorporated to ensure the quality of collected data, as participants may experience fatigue and reduced focus following the group task (Abbey & Meloy, 2017).

A group dynamics scale was created to measure participants' perceptions of how positive (or negative) interactions in their group were. The ***group dynamics score (dependent variable)*** was calculated as an average score from all 12 statements, where scores close to 2 indicate extremely positive group interactions, -2 - extremely negative, around 0 - neutral.

$$(3)\ group\ dynamics\ score\ = \frac{\sum_{i=1}^{12} statement_i\ score}{12},$$

where i indexes the statements.

Table 2. Statements on group dynamics included in the post-task survey

#	Statement	Source	Modified?	In baseline?
1.	I can rely on other group members	(Costa & Anderson, 2011)	✅	✅
2.	I trust other group members	(Naef & Schupp, 2009)	✅	✅
3.	I believe other group members have good intentions		✅	✅
4.	While making a decision we take each other's opinion into consideration	(Costa & Anderson, 2011)	❌	❌
5.	In this group we work in a climate of cooperation		❌	❌
6.	In this group we discuss and deal with issues or problems openly		❌	❌
7.	I feel I am an important part of this group	(Leuteritz et al., 2022)	✅	❌
8.	Members are committed to the achievement of the group objectives		✅	❌
9.	I feel good about our group activity	(Hamlyn-Harris et al., 2006)	❌	❌
10.	My group develops good and useful ideas		❌	❌
11.	I feel comfortable in my group		❌	❌
12.	As a group, we like one another		❌	❌

3.5.4. Pilot study

A pilot study was conducted on the 22nd of May with 6 participants (3 in each condition) who completed all parts of the study and provided detailed feedback on the clarity of the questions and instructions. While the feedback was mostly positive, the individuals in the treatment (VR) condition felt that more instructions on how to use the VR headset, controllers and the MeetinVR software were needed. As a result, new tutorial videos were created and placed before the group task instruction page, and extra 'How-to' screens were added within the VR environment. Pilot participants also validated the clarity and selection of the Group dynamics scale statements, and suggested minor changes which were reflected in the final version of the materials.

3.6. Hardware and Software

For the treatment (VR) condition, three identical Meta Quest 2 headsets were utilised. Although more advanced models like the Meta Quest 3 or Apple Vision Pro have been released since the Meta Quest 2's launch in 2020 (Yenduri et al., 2024), the Meta Quest 2 provides a good balance between functionality and affordability. Consequently, it has been widely used in similar recent studies (Abramczuk et al., 2023; Hennig-Thurau et al., 2023).

The software used for the group task in the control (VC) condition was Google Meet with a Miro add-on. This setup was selected as it allowed participants to collaborate on the same browser tab where the video call was taking place. All that was required from participants was to input the agreed rankings on that screen (Figure 3).

The software used for the group task in the treatment (VR) condition was MeetinVR. It was selected as it is the only virtual reality meeting environment which offers a free trial version with the required functionalities - customisable avatars and meeting rooms that can be setup in advance, high image resolution, and collaboration capabilities. In order to complete the group task, participants had to move the agreed ranking numbers on a screen (Figure 4). How-to instructions were provided before the start of the task and on screens in the virtual environment.

Identical HP ZBook laptops were available in all experiment rooms and all data collection (i.e. consent, surveys, individual task) was conducted via Qualtrics.

3.7. Analytic Strategy

Microsoft Excel was used to clean the raw survey data, and to prepare a clean dataset including all variables required for the analysis. Then, all quantitative analyses were conducted in R version 4.3.3 (Appendix D).

All responses from questions/statements with a Likert response format were assumed to be interval-scaled. Some academics criticise the appropriateness of such assumptions as response options are ordered and the intervals between values cannot be assumed to be equal (Jamieson, 2004). However, others support it if a 5 to 7-point Likert response format is used, especially when a meaningful scale with at least 8 items is being analysed (Carifio & Perla, 2007).

3.7.1. Group dynamics scale validation

Given the breadth of aspects covered by the group dynamics scale (i.e. trust, cooperation, communication, etc.), Exploratory Factor Analysis (EFA) was conducted to examine the underlying structure of the 12 scale items. First, tests were completed to determine if the data is appropriate for

factor analysis - the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.84 and Bartlett's test of sphericity was significant ($p < 0.001$), suggesting that it is (Tabachnick et al., 2019). Both Cattell's scree test (Cattell, 1978) and parallel analysis (Horn, 1965) indicated an optimal solution of only one factor though, so analysis was done at the 12-item scale. To assess the scale reliability, Cronbach's alpha was calculated and showed high internal consistency ($\alpha = 0$ .92).

3.7.2. Main Analysis

Each of the three hypotheses regarding the effect of VR technology as a meeting media (independent variable) on group dynamics (H_1), group performance (H_2), and tiredness (H_3) (dependent variables) was evaluated using separate simple linear regression analyses. To assess the robustness of the findings, multiple regression analyses were conducted, incorporating selected covariates relevant to each hypothesis being tested (see Appendix E1). To further refine the models, highly non-significant predictors ($p > 0.5$) were excluded. Simple linear regression models were used instead of t-tests for ease of results comparison. All models and their coefficients were evaluated for two-sided significance at the 95% confidence level ($p < 0.05$), and assumption checks were conducted for each model.

3.7.3. Additional analysis

To determine if there were differences in group discussion flows between the two conditions, t-tests were conducted on the duration (in minutes) of each activity (Table 1) and the overall task duration. To mitigate the risk of multiple comparisons problem (Chen et al., 2017), the significance levels were adjusted using the Holm-Bonferroni method (Holland & Copenhaver, 1988).

4. Results

In total, 56 participants took part in the study, of which 2 were excluded from the analysis as they failed more than one (out of three) attention checks, resulting in a final sample size of 54. Of these, 27 were assigned to each of the two conditions (VC and VR).

Based on the a-priori G*Power analysis, the study had fewer participants than required to be sufficiently powered. Post-hoc analysis, based on average effect sizes from this study, revealed that the achieved power was only 0.24 (Appendix A). This indicates a very high likelihood of a Type II error—failing to reject the null hypothesis when it is actually false (Schroeder et al., 2017).

4.1. Sample characteristics

All participants in the study were LSE students aged between 19 and 38 (with a mean age of 24), most of which (94%) pursuing a postgraduate degree (MSc or PhD). Less than half of individuals (33%) knew someone in their group and most (69%) had never used a VR headset before. The sample was balanced across conditions with respect to all demographic characteristics (age, gender, level of education, ethnicity), and other relevant ones, including whether participants wore prescription glasses, their baseline levels of trust, prior experience with VR technology, and whether they were acquainted with any other group members before the study (Table E2.1 in Appendix E2).

4.2. Main results

4.2.1. H_1: Effect of virtual reality (VR) medium on group dynamics

A simple linear regression analysis (R1) was performed to determine whether participation in the VR treatment significantly affected group dynamics and to evaluate the direction and magnitude of the predicted relationship. The data met all key assumptions, apart from normality, which was not required for the validity of the regression (Osborne & Waters, 2002). No influential outliers were identified (see Appendix E4.1).

The R1 results were non-significant, $F(52) = 0.441$, $p = 0.510$, Adj. $R2 = -0.0107$, indicating that VR Treatment does not explain a significant proportion of the variation in Group dynamics, and therefore providing no initial support for H_1 (Table 3).

Table 3. R1: Regression analysis results of VR Treatment impact on Group dynamics

Variables	β	SE	95% CI	t	*p*
(Intercept)	1.366	0.1	[1.165, 1.567]	13.632	<0.001 ***
VR Treatment	0.094	0.142	[-0.19, 0.378]	0.664	0.510

Notes: Results: $F(52) = 0.441$, $p = 0.510$, Adj. $R^2 = -0.0107$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To check the robustness of the above finding, a multiple linear regression (R2) was conducted, controlling for all relevant covariates (see Appendix E1 for inclusion reasoning). The data met all key assumptions, and no influential outliers were identified (see Appendix E4.2).

The R2 results were non-significant, $F(38) = 1.606$, $p = 0.118$, $R^2 = 0.3881$, Adj. $R^2 = 0.1465$, and VR Treatment continued to remain a non-significant predictor, $\beta = -0.091$, $p = 0.717$ (see Appendix E3, Table E3.1). As a further robustness check, highly non-significant predictors ($p>0.50$)

were dropped for a better fitting model. The resulting model (R3) was significant overall, $F(44) = 2.861$, $p = 0.010$, $R^2 = 0.3692$, Adj. $R^2 = 0.2401$, but confirmed that VR treatment does not have a significant effect on Group dynamics, $\beta = -0.09$, $p = 0.652$. However, feeling tired at the end of the task had a significant negative impact on Group dynamics, $\beta = -0.16$, $p = 0.020$ (see Appendix E3, Table E3.2). Given that R3 tests multiple (9) hypotheses, an adjusted significance level using the Holm-Bonferroni method should be used ($\alpha^* = 0.006$) (Holland & Copenhaver, 1988). With the adjustment, feeling tired at the end of the task is no longer a significant predictor.

Given that VR Treatment did not have a significant effect on Group dynamics within any regression model (R1-R3), there is not enough evidence to accept H_1: positive effect of VR treatment on Group dynamics and the null hypothesis cannot be rejected (Table 4). Given the low power of this study, this finding is inconclusive.

Table 4. VR Treatment coefficients, p-values, and sensitivity analyses for all regressions of Group dynamics scale

Regression model	β for VR Treatment	*p* for VR Treatment	Tested predictors	Model p	Adj. R^2
R1: Simple linear regression	0.094	0.510	1	0.510	-0.011
R2: Multiple linear regression, with selected covariates [1]	-0.091	0.717	15	0.118	0.1465
R3: Multiple linear regression, highly non-significant predictors (p>0.5) removed	-0.090	0.652	9	0.010	0.2401

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
[1] Reasoning for covariates selection available in Appendix E1

4.2.2. H_2: Effect of virtual reality (VR) medium on group performance

A simple linear regression analysis (R4) was performed to determine whether participation in the VR treatment significantly affected group performance and to evaluate the direction and magnitude of the predicted relationship. The data met the assumption of linearity and no influential outliers were identified, but did not meet the other key assumptions of independence, normality and homoscedasticity, and results should be interpreted with caution (Osborne & Waters, 2002) (see Appendix E4.4).

The R4 results were significant, $F(52) = 4.138$, $p = 0.047$, Adj. $R^2 = 0.0559$, indicating that approximately 6% of the variation in Group performance could be attributed to the VR Treatment. The VR Treatment significantly predicted the Group performance score, $\beta = -5.41$, $p = 0.047$, decreasing it by an average of 5.41 units, and thus providing initial evidence against H_2 (Table 5).

Table 5. R4: Regression analysis results of VR Treatment impact on Group performance

Variables	β	SE	95% CI	t	*p*
(Intercept)	44.815	1.880	[41.043, 48.587]	23.842	<0.001 ***
VR Treatment	-5.407	2.658	[-10.742, -0.073]	-2.034	0.047 *

Notes: Results: $F(52) = 4.138$, $p = 0.047$, Adj. $R^2 = 0.0559$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

To check the robustness of the above findings, a multiple linear regression (R5) was conducted, including all relevant covariates (see Appendix E1 for inclusion reasoning). The data met all key assumptions except for independence, which is fulfilled by the design of the study regardless (Osborne & Waters, 2002). No influential outliers were identified (see Appendix E4.5).

The R5 results were significant, $F(37) = 6.303$, $p < 0.001$, $R^2 = 0.7255$, Adj. $R^2 = 0.6069$, indicating that approximately 61% of the variation in Group performance could be attributed to the VR Treatment and the selected predictors (better model fit than R4). The Treatment continued to predict a significant decrease in the Group performance score, $\beta = -11.47$, $p < 0.001$, and therefore providing further evidence against H_2. Additionally, believing in good intentions also predicted a significant decrease in Group performance score, $\beta = -3.11$, $p = 0.025$, while a number of covariates predicted a significant increase - wearing prescription glasses, $\beta = 4.75$, $p = 0.030$; average individual score in group, $\beta = 1.10$, $p = 0.001$; feeling present during group discussion, $\beta = 3.37$, $p = 0.042$; and duration of group task, $\beta = 0.58$, $p = 0.040$ (see Appendix E3, Table E3.3).

The difference between R^2 and the Adjusted R^2 for R5 indicated that there may be variables in the model that do not contribute to its explanatory power (Schroeder et al., 2017). To improve the model fit, highly non-significant predictors ($p > 0.50$) were removed. The resulting model (R6) confirmed the significant impact of VR treatment on the Group performance score, $\beta = -11.57$, $p < 0.001$, and further highlighted the significant predictive power of believing in good intentions, $\beta = -3.07$, $p = 0.014$; wearing prescription glasses, $\beta = 4.48$, $p = 0.028$; average individual score in group, $\beta = 1.06$, $p < 0.001$; feeling present during group discussion, $\beta = 3.41$, $p = 0.030$; and duration of group task, $\beta = 0.57$, $p = 0.028$ (see Appendix E3, Table E3.4). Given that R6 tests multiple (12) hypotheses, adjusted significance levels using the Holm-Bonferroni method should be used (Holland & Copenhaver, 1988). With the adjustments, only average individual score in group remains a significant predictor.

Across all regression models (R4-R6) VR Treatment reliably predicted a significant decrease in Group performance score of between 5.4 to 11.6 points, therefore the hypothesis being tested, H_2: positive effect of VR treatment on group performance, can be rejected (Table 6).

Table 6. VR Treatment coefficients, p-values, and sensitivity analyses for all regressions of Group performance score

Regression model	β for VR Treatment	*p* for VR Treatment	Tested predictors	Model p	Adj. R^2
R4: Simple linear regression	-5.407	0.047 *	1	0.047	0.0559
R5: Multiple linear regression, with selected covariates [1]	-11.471	<0.001 ***	16	<0.001 ***	0.6069
R6: Multiple linear regression, highly non-significant predictors (p>0.5) removed	-11.572	<0.001 ***	12	<0.001 ***	0.6390

Notes: * p < 0.05, ** p < 0.01, *** p < 0.001
[1] Reasoning for covariates selection available in Appendix E1

4.2.3. H_3: Effect of virtual reality (VR) medium on tiredness

A simple linear regression analysis (R7) was performed to determine whether participation in the VR treatment significantly affected change in tiredness and to evaluate the direction and magnitude of the predicted relationship. The data met all key assumptions, apart from normality and homoscedasticity, so results should be interpreted with caution (Osborne & Waters, 2002). No influential outliers were identified (see Appendix E4.5).

The R7 results were non-significant, $F(52) = 0.948$, $p = 0.256$, Adj. $R^2 = 0.0060$, indicating that VR Treatment does not explain a significant proportion of the variation in change in tiredness, and therefore providing no initial support for H_3 (Table 7).

Table 7. R7: Regression analysis results of VR Treatment impact on Change in tiredness

Variables	β	SE	95% CI	t	*p*
(Intercept)	-0.111	0.182	[-0.477, 0.255]	-0.609	0.545
VR Treatment	0.296	0.258	[-0.221, 0.814]	1.148	0.256

Notes: Results: $F(52) = 0.948$, $p = 0.256$, Adj. $R^2 = 0.0060$
* p < 0.05, ** p < 0.01, *** p < 0.001

To check the robustness of the above finding, a multiple linear regression (R8) was conducted, controlling for all relevant covariates (see Appendix E1 for inclusion reasoning).

The R8 results were significant, $F(41) = 2.27$, $p = 0.0256$, $R^2 = 0.3992$, Adj. $R^2 = 0.2233$, but the model did not meet the homoscedasticity and linearity assumptions (see Appendix E4.8). VR Treatment continued to remain a non-significant predictor, $\beta = 0.441$, $p = 0.249$, however significant decreases in the change in tiredness (feeling less tired) were predicted by levels of

tiredness at the start of the task, β = -0.455, p = 0.007, and group dynamics, β = -0.589, p = 0.029 (see Appendix E3, Table E3.5).

As a further robustness check, highly non-significant predictors (p>0.50) were dropped for a better fitting model. The resulting model (R9) was significant overall, F(44) = 3.156, p = 0.005, R2= 0.3923, Adj. R2= 0.2680, but did not meet the homoscedasticity and linearity assumptions (see Appendix E4.9). The model confirmed that VR treatment does not have a significant effect on change in tiredness, β = 0.277, p = 0.272, and that tiredness at the start of the task, β = -0.472, p = 0.004, and group dynamics, β = -0.619, p = 0.016, have a negative effect. Wearing prescription glasses also significantly predicted a decrease in the change in tiredness, β = -0.489, p = 0.04 (see Appendix E3, Table E3.6). Given that R9 tests multiple (9) hypotheses, adjusted significance levels using the Holm-Bonferroni method should be used (Holland & Copenhaver, 1988). With the adjustments, only feeling tired at the start of the task remains a significant predictor, however as R9 did not meet the homoscedasticity assumption, which increases the possibility of a Type I error (Osborne & Waters, 2002), the result should be interpreted with caution.

Given that VR Treatment did not have a significant effect on change in tiredness within any regression model (R7-R9), there is no evidence in support of H_3: negative effect of VR treatment on tiredness, and the null hypothesis cannot be rejected (Table 8). This finding is inconclusive due to the low power of this study.

Table 8. VR Treatment coefficients, p-values, and sensitivity analyses for all regressions of Change in tiredness

Regression model	β for VR Treatment	*p* for VR Treatment	Tested predictors	Model p	Adj. R^2
R7: Simple linear regression	0.296	0.256	1	0.256	0.0060
R8: Multiple linear regression, with selected covariates [1]	0.441	0.249	12	0.026 *	0.2233
R9: Multiple linear regression, highly non-significant predictors (p>0.5) removed	0.277	0.272	9	0.005 **	2680

Notes: * p < 0.05, ** p < 0.01, *** p < 0.001
[1] Reasoning for covariates selection available in Appendix E1

4.3. Additional Analysis

The flow of the conversation during the group task was different across the two conditions, with a higher percentage of time spent on figuring out how to work with the technology in the VR condition, at the expense of having a productive conversation (Figure 5).

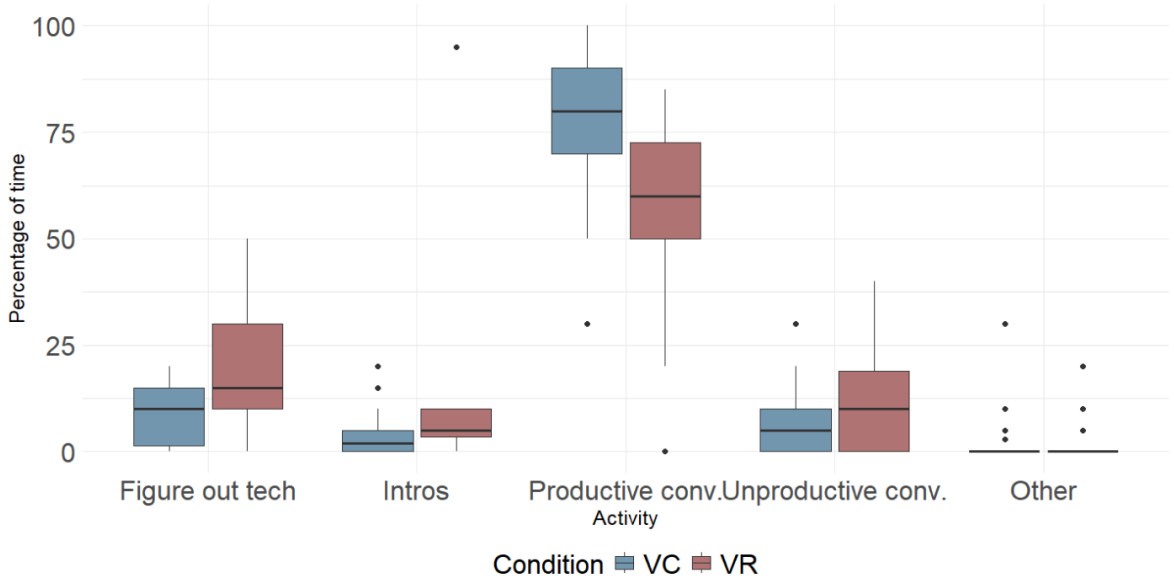


Figure 5. Breakdown of time spent on different activities during group task

To determine if there were significant differences across conditions in how much time (in minutes) groups spent on each activity, and in total, t-tests were conducted (Table 9). Results showed that participants in the VR condition spent significantly more time on the group task in total, *MD* = 8.8 mins, p <0.001, on figuring out the technology, *MD* = 3.3 mins, p <0.001, and on unproductive activities (i.e. interrupting each other, arguing), *MD* = 1.5 mins, p = 0.011. Significant findings are illustrated in Figure 6.

Table 9. Differences between VR and VC condition on duration of time spent on different activities during group task, and in total

	VR Treatment (N = 27)		VC Treatment (N = 27)					
Time spent on…	**M**	**SD**	**M**	**SD**	**df**	**t**	***p***	**Adj. α***
Group task	**24.522**	**3.223**	**15.689**	**4.379**	**~48**	**-8.440**	**<0.001**	**0.008**
Figuring out the tech	**4.619**	**3.093**	**1.337**	**1.234**	**~34**	**-5.121**	**<0.001**	**0.010**
Unproductive activities	**2.585**	**2.489**	**1.126**	**1.392**	**~41**	**-2.659**	**0.011**	**0.013**
Introductions	2.215	4.435	0.622	0.878	~28	-1.831	0.078	0.017
Productive activities	14.448	5.151	12.244	4.306	~50	-1.706	0.094	0.025
Other activities	0.656	1.456	0.363	1.387	~52	-0.756	0.453	0.050

Notes: Adjusted significance level $\alpha_i^*=\alpha/(k-i+1)$, where significance level $\alpha=0.05$, number of hypotheses $k=6$, and i is the order of p-values from smallest to largest, $i \in [1, k]$ (Holland & Copenhaver, 1988).
All reported time is measured in minutes. Significant results are in bold.

It is important to note that the VR meeting room lacked a clock, providing no objective means for participants to track time. As a result, nearly all sessions had to be interrupted by the researcher (myself) approximately 15 minutes after they began, informing participants that time was up and requesting them to conclude their discussion and finalise their answers. This, coupled with the additional time spent learning to use the technology and engaging in unproductive conversations, likely explains the longer average duration of the group task in the VR condition.

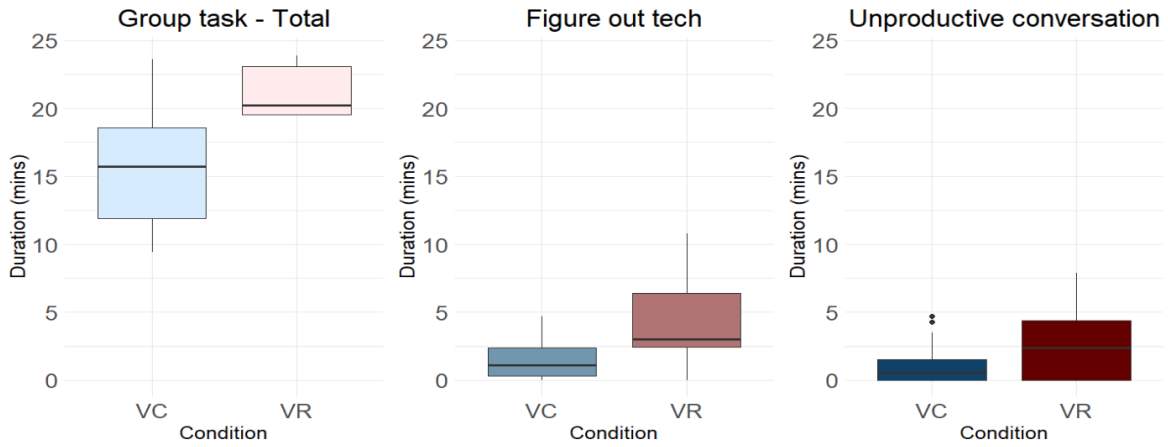


Figure 6. Significant differences between conditions in duration of activities (in minutes)

5. Discussion

5.1. Overview of findings

Since this study did not achieve the required number of participants to be sufficiently powered, its results provide inconclusive evidence on the tested hypotheses and should be interpreted with caution.

One of the main hypothesised benefits of conducting meetings over VR is that the environment could foster better group dynamics. This study, however, finds no significant differences among conditions, which could be due to the low power of the study, thus providing no conclusive evidence on the topic. This is in line with the available literature (Abramczuk et al., 2023; Aliman et al., 2023; Macchi & De Pisapia, 2024).

Another important pillar of VR's case as an alternative to videoconferencing is that it leads to similar group performance. Contrary to the available literature (Abdullah et al., 2021; Macchi & De Pisapia, 2024), this study found significant evidence against that - on average groups meeting in VR achieved between 5.6 and 11.6 points lower scores (out of maximum score of 112), compared to

those interacting over videoconferencing. This could be explained by the differences in the discussion patterns between the two conditions. Groups interacting in VR spent on average 3.3 minutes longer on trying to figure out how to use the technology and 1.5 minutes longer in unproductive conversation such as arguing or interrupting each other. Even though they also spent 8.8 mins longer in total to complete the group task, there was no significant difference in the time spent on productive conversation, indicating that the reduced performance could have been caused by the disruptions associated with the unfamiliarity with the VR technology. Unsurprisingly, the study also found that the average expertise in a group was a strong predictor of group performance.

In order for VR meetings to be effective alternatives to videoconferencing, they should not lead to higher levels of tiredness, contrary to the findings in the literature (Abramczuk et al., 2023; Aliman et al., 2023; Hennig-Thurau et al., 2023). The findings of this study are inconclusive as there were no significant differences between conditions, which could have been due to the low power of the study and higher chance of a false negative, or due to the short durations of the interactions in VR (between 19.5 and 29.7 minutes, with an average of 24.5).

5.2. Limitations and future work

This study had many limitations, mostly arising from the budget constraints associated with an unfunded experimental study and the time constraints associated with the tight deadlines of an MSc dissertation. Those limitations should be considered to ensure responsible interpretation of the results and to guide future research.

Firstly, significantly less than the required number of participants were recruited, which meant that the study did not achieve the required statistical power to lead to conclusive results. Moreover, all recruited participants were LSE students, between the ages of 19 and 38. This is not a representative sample of the knowledge worker population who may have different experiences with and attitudes to VR technologies (Aburbeian et al., 2022) which decreases the external validity of this study. However, it can be argued that although the behaviour of current students may not be generalizable to the wider population, it is still in itself interesting as they will soon be entering the workforce.

Secondly, given the unpaid nature of the study, the planned duration of the experimental sessions had to be capped at 30 minutes to make recruitment feasible. Thus, a short enough scale that holistically measures all aspects of group dynamics was required. Such a scale could not be found in the literature, and a new scale was created, which had high internal consistency (Cronbach's $\alpha = 0\ .92$), but was not validated prior to the study due to budget constraints. An interesting avenue of future research though would be to test and measure group dynamics

empirically, rather than rely on self-reports. Some ideas for more robust and holistic findings could include measuring relevant hormonal changes (i.e. oxytocin, cortisol), neuroimaging (Alós-Ferrer & Farolfi, 2019) or performing sentiment analysis on recorded conversations (Menon et al., 2022).

Thirdly, due to the limited duration of the experimental sessions, comprehensive training on how to use the VR technology was not provided before the start of the study. This might have impacted the performance of those groups as 69% of participants were using such technology for the first time, 40% of groups encountered some form of a technical issue (i.e. having to restart a headset), and based on self-reports, participants in the VR condition spent on average 3.3 minutes longer trying to figure out how to use the technology. In future studies, researchers should incorporate a training session and make sure participants can comfortably use the technology before the start of the study.

Similarly, the short duration of the experimental sessions and group discussions — averaging 15.7 minutes for the control and 24.5 minutes for the treatment group — may not have been sufficient for differences in group dynamics to manifest, given that group formation is a complex process (Arrow et al., 2000). Future studies could address this by incorporating multiple longer group tasks with breaks in between (Abdullah et al., 2021). It will be of particular interest to also observe if the group dynamics change for different types of tasks or with each following task.

Last but not least, budget constraints necessitated the use of a free version of VR software and older model VR headsets, which provided lower quality audio and video (Yenduri et al., 2024) and limited functionalities (e.g. users could not customise their avatars). This may have affected the participants' experience in the VR treatment, and therefore, future studies should aim to use the latest hardware and software to offer current insights for businesses managing remote workforces.

Furthermore, most existing research, including this study, focuses on replicating face-to-face office settings in a virtual environment, without exploring the additional capabilities the technology could provide. McVeigh-Schultz & Isbister (2022) argue that the critical question researchers should be asking is not whether VR is a better alternative to videoconferencing, but rather what unique benefits VR can offer that would entice even individuals in the same location to prefer it.

6. Conclusion

Since the coronavirus pandemic, working from home has become increasingly prevalent, leading to challenges such as reduced social interactions, diminished sense of belonging, reduced collaboration, and 'Zoom fatigue' among remote workers. Conducting some of the group activities in virtual reality could potentially address these issues.

This study contributes to the very limited body of literature, investigating the effects of meetings in VR on 1) group dynamics, 2) group performance and 3) levels of tiredness caused by the technology. As one of the few lab studies examining this topic, its robust experimental design offers a solid foundation for future research to build upon. Due to its limited sample size, the study provides inconclusive results regarding the effect of VR meetings on group dynamics and users' levels of tiredness. However, in contrast to the existing literature, it finds that groups interacting in VR perform worse than those using videoconferencing. This can be attributed to participants' unfamiliarity with the novel technology and the technical issues faced by groups, likely disrupting their discussions. These findings highlight the importance of training if VR technology is to be widely implemented across organisations. Still further research on the topic is required and this study outlines the exciting research path ahead.

Shall we meet in VR? Not yet.

References

Abbey, J. D., & Meloy, M. G. (2017). Attention by design: Using attention checks to detect inattentive respondents and improve data quality. *Journal of Operations Management*, *53–56*, 63–70. https://doi.org/10.1016/j.jom.2017.06.001

Abdullah, A., Kolkmeier, J., Lo, V., & Neff, M. (2021). Videoconference and Embodied VR: Communication Patterns across Task and Medium. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2). https://doi.org/10.1145/3479597

Abramczuk, K., Bohdanowicz, Z., Muczyński, B., Skorupska, K. H., & Cnotkowski, D. (2023). Meet me in VR! Can VR space help remote teams connect: A seven-week study with Horizon Workrooms. *International Journal of Human-Computer Studies*, *179*, 103104. https://doi.org/10.1016/j.ijhcs.2023.103104

Aburbeian, A. M., Owda, A. Y., & Owda, M. (2022). A Technology Acceptance Model Survey of the Metaverse Prospects. *AI (Switzerland)*, *3*(2), 285–302. https://doi.org/10.3390/ai3020018

Alexander, A., De Smet, A., & Mysore, M. (2020). *Reimagining the postpandemic workforce.* McKinsey Quarterly. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/reimagining-the-postpandemic-workforce

Aliman, D. N., Hennig-Thurau, T., & Henke, A. (2024). Navigating the enterprise metaverse: How virtual reality affects business agility and meeting outcomes. *Business Horizons*. https://doi.org/10.1016/j.bushor.2024.05.007

Alós-Ferrer, C., & Farolfi, F. (2019). Trust Games and Beyond. *Frontiers in Neuroscience*, *13*, 887. https://doi.org/10.3389/fnins.2019.00887

Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). *Small Groups As Complex Systems: Formation, Coordination, Development, and Adaptation*. SAGE Publications, Incorporated. https://doi.org/10.4135/9781452204666

Biener, V., Kalamkar, S., Nouri, N., Ofek, E., Pahud, M., Dudley, J. J., Hu, J., Kristensson, P. O., Weerasinghe, M., Pucihar, K. Č., Kljun, M., Streuber, S., & Grubert, J. (2022). Quantifying the Effects of Working in VR for One Week. *IEEE Transactions on Visualization and Computer Graphics*, *28*(11), 3810–3820. https://doi.org/10.1109/TVCG.2022.3203103

Campbell, A., Holz, T., Cosgrove, J., Harlick, M., & O'Sullivan, T. (2020). Uses of Virtual Reality for Communication in Financial Services: A Case Study on Comparing Different Telepresence Interfaces: Virtual Reality Compared to Video Conferencing. In *Lecture Notes in Networks and Systems* (pp. 463–481). https://doi.org/10.1007/978-3-030-12388-8_33

Carifio, J., & Perla, R. J. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*, *3*(3), 106–116. https://doi.org/10.3844/jssp.2007.106.116

Cattell, R. B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. Springer US. https://doi.org/10.1007/978-1-4684-2262-7

Chang, E., Kim, H. T., & Yoo, B. (2020). Virtual Reality Sickness: A Review of Causes and Measurements. *International Journal of Human-Computer Interaction*, 1658–1682. https://doi.org/10.1080/10447318.2020.1778351

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009

Chen, S.-Y., Feng, Z., & Yi, X. (2017). A general introduction to adjustment for multiple comparisons. *Journal of Thoracic Disease*, *9*(6), 1725–1729. https://doi.org/10.21037/jtd.2017.05.34

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cooney, G., Mastroianni, A. M., Abi-Esber, N., & Brooks, A. W. (2020). The many minds problem: Disclosure in dyadic versus group conversation. *Current Opinion in Psychology*, *31*, 22–27. https://doi.org/10.1016/j.copsyc.2019.06.032

Costa, A. C., & Anderson, N. (2011). Measuring trust in teams: Development and validation of a multifaceted measure of formative and reflective indicators of team trust. *European Journal of Work and Organizational Psychology*, *20*(1), 119–154. https://doi.org/10.1080/13594320903272083

Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, *949*(1), 012009. https://doi.org/10.1088/1742-6596/949/1/012009

De Jong, B. A., Dirks, K. T., & Gillespie, N. (2016). Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of Applied Psychology*, *101*(8), 1134–1150. https://doi.org/10.1037/apl0000110

Dirks, K. T., & Ferrin, D. L. (2001). The Role of Trust in Organizational Settings. *Organization Science*, *12*(4), 450–467. https://doi.org/10.1287/orsc.12.4.450.10640

Elbogen, E. B., Lanier, M., Griffin, S. C., Blakey, S. M., Gluff, J. A., Wagner, H. R., & Tsai, J. (2022). A National Study of Zoom Fatigue and Mental Health during the COVID-19 Pandemic: Implications for Future Remote Work. *Cyberpsychology, Behavior, and Social Networking*, *25*(7), 409–415. https://doi.org/10.1089/cyber.2021.0257

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Feitosa, J., Grossman, R., Kramer, W. S., & Salas, E. (2020). Measuring team trust: A critical and meta-analytical review. *Journal of Organizational Behavior*, *41*(5), 479–501. https://doi.org/10.1002/job.2436

Geue, P. E. (2018). Positive Practices in the Workplace: Impact on Team Climate, Work Engagement, and Task Performance. *The Journal of Applied Behavioral Science*, *54*(3), 272–301. https://doi.org/10.1177/0021886318773459

Gould, S. J. J., Rudnicka, A., Cook, D., Cecchinato, M. E., Newbold, J. W., & Cox, A. L. (2023). Remote Work, Work Measurement and the State of Work Research in Human-Centred Computing. *Interacting with Computers*, *35*(5), 725–734. https://doi.org/10.1093/iwc/iwad014

Gross, D. R. (1984). Time allocation: A tool for the study of cultural behavior. *Annual Review of Anthropology. Vol. 13*, 519–558. https://doi.org/10.1146/annurev.anthro.13.1.519

Grossman, R., Nolan, K., Rosch, Z., Mazer, D., & Salas, E. (2022). The team cohesion-performance relationship: A meta-analysis exploring measurement approaches and the changing team landscape. *Organizational Psychology Review*, *12*(2), 181–238. https://doi.org/10.1177/20413866211041157

Grove, L. (2023). *Research Ethics Policy and Procedures*. The London School of Economics and Political Science. https://info.lse.ac.uk/staff/services/Policies-and-procedures/Assets/Documents/resEthPolPro.pdf

Güzel, A. H., Beyazian, J., Chakravarthula, P., & Akşit, K. (2023). ChromaCorrect: Prescription correction in virtual reality headsets through perceptual guidance. *Biomedical Optics Express*, *14*(5), 2166–2180. https://doi.org/10.1364/BOE.485776

Hamlyn-Harris, J. H., Hurst, B. J., Baggo, K. von, & Bayley, A. J. (2006). Predictors of Team Work Satisfaction. *Journal of Information Technology Education: Research*, *5*(1), 299–315. Retrieved August 17, 2024 from https://www.learntechlib.org/p/111547/

Hennig-Thurau, T., Aliman, D. N., Herting, A. M., Cziehso, G. P., Linder, M., & Kübler, R. V. (2023). Social interactions in the metaverse: Framework, initial evidence, and research roadmap. *Journal of the Academy of Marketing Science*, *51*(4), 889–913. https://doi.org/10.1007/s11747-022-00908-0

Hockey, R. (2013). *The Psychology of Fatigue: Work, Effort and Control*. Cambridge University Press. https://doi.org/10.1017/CBO9781139015394

Holland, B. S., & Copenhaver, M. D. (1988). Improved Bonferroni-type multiple testing procedures. Psychological Bulletin, 104(1), 145–149. https://doi.org/10.1037/0033-2909.104.1.145

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Innami, I. (1994). The Quality of Group Decisions, Group Verbal Behavior, and Intervention. *Organizational Behavior and Human Decision Processes*, *60*(3), 409–430. https://doi.org/10.1006/obhd.1994.1092

Jamieson, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, *38*(12), 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

Jasgur, C. (2023). Tips for managing a remote workforce: Journal of Business Continuity & Emergency Planning. *Journal of Business Continuity & Emergency Planning*, *16*(4), 346–358.

Joseph, R. P., Keller, C., & Ainsworth, B. E. (2016). Recruiting Participants into Pilot Trials: Techniques for Researchers with Shoestring Budgets. *Californian Journal of Health Promotion*, *14*(2), 81–89.

Kirchner, K., Ipsen, C., & Hansen, J. P. (2021). COVID-19 leadership challenges in knowledge work. *Knowledge Management Research & Practice*, *4*, 493–500. https://doi.org/10.1080/14778238.2021.1877579

Kobayashi, M., Ueno, K., & Ise, S. (2015). The Effects of Spatialized Sounds on the Sense of Presence in Auditory Virtual Environments: A Psychological and Physiological Study. *Presence: Teleoperators and Virtual Environments*, *24*(2), 163–174. https://doi.org/10.1162/PRES_a_00226

Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Validation of the Virtual Reality Neuroscience Questionnaire: Maximum Duration of Immersive Virtual Reality Sessions Without the Presence of Pertinent Adverse Symptomatology. *Frontiers in human neuroscience*, 13, 417. https://doi.org/10.3389/fnhum.2019.00417

Kurzweg, M., Reinhardt, J., Nabok, W., & Wolf, K. (2021). Using Body Language of Avatars in VR Meetings as Communication Status Cue. *Proceedings of Mensch Und Computer 2021*, 366–377. https://doi.org/10.1145/3473856.3473865

Kyaw, B. M., Saxena, N., Posadzki, P., Vseteckova, J., Nikolaou, C. K., George, P. P., Divakar, U., Masiello, I., Kononowicz, A. A., Zary, N., & Car, L. T. (2019). Virtual reality for health professions education: Systematic review and meta-analysis by the digital health education collaboration. *Journal of Medical Internet Research*, *21*(1). https://doi.org/10.2196/12959

Lafferty, J. C., & Pond, A. W. (1974). *The Desert Survival Problem: A Group Decision Making Experience for Examining and Increasing Individual and Team Effectiveness: Manual*. Experimental Learning Methods.

Lee, L.-H., Braud, T., Zhou, P., Wang, L., Xu, D., Lin, Z., Kumar, A., Bermejo, C., & Hui, P. (2021). *All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda* (arXiv:2110.05352). arXiv. https://doi.org/10.48550/arXiv.2110.05352

Leuteritz, J.-P., Navarro, J., Czakert, J. P., & Berger, R. (2022). Validation of the German group development (GD) questionnaire. *Current Psychology*, *41*(7), 4321–4328. https://doi.org/10.1007/s12144-020-00945-4

Li, B. J., & Yee, A. Z. H. (2023). Understanding videoconference fatigue: A systematic review of dimensions, antecedents and theories. *Internet Research*, *33*(2), 796–819. https://doi.org/10.1108/INTR-07-2021-0499

List, J. A., Sadoff, S., & Wagner, M. (2011). So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics*, *14*(4), 439–457. https://doi.org/10.1007/s10683-011-9275-7

Littlepage, G. E., Schmidt, G. W., Whisler, E. W., & Frost, A. G. (1995). An input-process-output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology*, 69(5), 877–889. https://doi.org/10.1037/0022-3514.69.5.877

Ma, J. K.-H. (2021). The digital divide at school and at home: A comparison between schools by socioeconomic level across 47 countries. *International Journal of Comparative Sociology*, *62*(2), 115–140. https://doi.org/10.1177/00207152211023540

Macchi, G., & De Pisapia, N. (2024). Virtual reality, face-to-face, and 2D video conferencing differently impact fatigue, creativity, flow, and decision-making in workplace dynamics. *Scientific Reports*, *14*(1), 10260. https://doi.org/10.1038/s41598-024-60942-6

McGrath, J. E. (1984). *Groups: Interaction and Performance*. Prentice-Hall.

Mclean, S., Gee, M., Booth, A., Salway, S., Nancarrow, S., Cobb, M., & Bhanbhro, S. (2014). Targeting the use of reminders and notifications for uptake by populations (TURNUP): A systematic review and evidence synthesis. *Health Services and Delivery Research*, *2*(34), Article 34. https://doi.org/10.3310/hsdr02340

McVeigh-Schultz, J., & Isbister, K. (2022). A "beyond being there" for VR meetings: Envisioning the future of remote work. *Human–Computer Interaction*, *37*(5), 433–453. https://doi.org/10.1080/07370024.2021.1994860

Menon, V., Weger, K., Mesmer, B., & Gholston, S. (2022). *Using Big Data Analytics for Sentiment Analysis to Explore Team Communication Dynamics in Human Machine Interactions for Team Situational Awareness*. Proceedings of the 2022 IEEE International Conference on Human-Machine Systems, ICHMS 2022. https://doi.org/10.1109/ICHMS56717.2022.9980604

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers and Education*, *70*, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

Moreland, R. L., & Levine, J. M. (2002). Socialization and trust in work groups. *Group Processes & Intergroup Relations*, *5*(3), 185–201. https://doi.org/10.1177/1368430202005003001

Naef, M., & Schupp, J. (2009). Measuring Trust: Experiments and Surveys in Contrast and Combination. *SOEPpaper No 167*. http://dx.doi.org/10.2139/ssrn.1367375

Nesher Shoshan, H., & Wehrt, W. (2022). Understanding "Zoom fatigue": A mixed-method approach. *Applied Psychology*, *71*(3), 827–852. https://doi.org/10.1111/apps.12360

Osborne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research, and Evaluation*, *8(1)*(2). https://doi.org/10.7275/r222-hv23

Panda, P., Nicholas, M. J., Gonzalez-Franco, M., Inkpen, K., Ofek, E., Cutler, R., Hinckley, K., & Lanier, J. (2022). AllTogether: Effect of Avatars in Mixed-Modality Conferencing Environments. *Proceedings of the 1st Annual Meeting of the Symposium on Human-Computer Interaction for Work*, 1–10. https://doi.org/10.1145/3533406.3539658

Park, H., Ahn, D., & Lee, J. (2023). *Towards a Metaverse Workspace: Opportunities, Challenges, and Design Implications.* In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23). https://doi.org/10.1145/3544548.3581306

Perugini, M., Gallucci, M., & Costantini, G. (2018). *A Practical Primer To Power Analysis for Simple Experimental Designs* (1). *31*(1), Article 1. https://doi.org/10.5334/irsp.181

Qazi, A., Hasan, N., Abayomi-Alli, O., Hardaker, G., Scherer, R., Sarker, Y., Kumar Paul, S., & Maitama, J. Z. (2022). Gender differences in information and communication technology use & skills: A systematic review and meta-analysis. *Education and Information Technologies*, *27*(3), 4225–4258. https://doi.org/10.1007/s10639-021-10775-x

Sagnier, C., Loup-Escande, E., & Valléry, G. (2020). Effects of Gender and Prior Experience in Immersive User Experience with Virtual Reality. In T. Ahram & C. Falcão (Eds.), *Advances in Usability and User Experience* (pp. 305–314). Springer International Publishing. https://doi.org/10.1007/978-3-030-19135-1_30

Saredakis, D., Szpak, A., Birckhead, B., Keage, H. A. D., Rizzo, A., & Loetscher, T. (2020). Factors associated with virtual reality sickness in head-mounted displays: A systematic review and meta-analysis. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00096

Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. (2017). *Understanding Regression Analysis: An Introductory Guide*. SAGE Publications, Inc. https://doi.org/10.4135/9781506361628

Sell, J., Knottnerus, J. D., & Adcock-Azbill, C. (2013). Disruptions in Task Groups. *Social Science Quarterly*, *94*(3), 715–731. https://doi.org/10.1111/j.1540-6237.2012.00879.x

Staddon, R. V. (2020). Bringing technology to the mature classroom: Age differences in use and attitudes. *International Journal of Educational Technology in Higher Education*, *17*(1), 11. https://doi.org/10.1186/s41239-020-00184-4

Stantcheva, S. (2023). How to Run Surveys: A Guide to Creating Your Own Identifying Variation and Revealing the Invisible. *Annual Review of Economics*, *15*(Volume 15, 2023), 205–234. https://doi.org/10.1146/annurev-economics-091622-010157

Statista. (2024). *Employees working primarily remotely worldwide 2015-2023*. Retrieved 15 August 2024, from https://www.statista.com/statistics/1450450/employees-remote-work-share/

Stigler, S. M. (1969). The use of random allocation for the control of selection bias. *Biometrika*, *56*(3), 553–560. https://doi.org/10.1093/biomet/56.3.553

Tabachnick, B. G., & Fidell, L. S. (2019). Using Multivariate Statistics (7th ed.). Pearson.

Dirks, Kurt T., & Ferrin, Donald L. (2001). The Role of Trust in Organizational Settings. *Organization Science 12(4):450-467. https://doi.org/10.1287/orsc.12.4.450.10640*

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00158

Wilson, T. D., Aronson, E., & Carlsmith, K. (2010). The Art of Laboratory Experimentation. *Handbook of Social Psychology*. https://onlinelibrary.wiley.com/doi/10.1002/9780470561119.socpsy001002

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Yang, L., Holtz, D., Jaffe, S., Suri, S., Sinha, S., Weston, J., Joyce, C., Shah, N., Sherman, K., Hecht, B., & Teevan, J. (2022). The effects of remote work on collaboration among information workers. *Nature Human Behaviour*, *6*(1), 43–54. https://doi.org/10.1038/s41562-021-01196-4

Yenduri, G., M, R., Maddikunta, P. K. R., Gadekallu, T. R., Jhaveri, R. H., Bandi, A., Chen, J., Wang, W., Shirawalmath, A. A., Ravishankar, R., & Wang, W. (2024). *Spatial Computing: Concept, Applications, Challenges and Future Directions* (arXiv:2402.07912). arXiv. https://doi.org/10.48550/arXiv.2402.07912

Yu, Z. (2023). A meta-analysis of the effect of virtual reality technology use in education. *Interactive Learning Environments*, *31*(8), 4956–4976. https://doi.org/10.1080/10494820.2021.1989466

Appendices

A. Power analysis and effect size benchmark

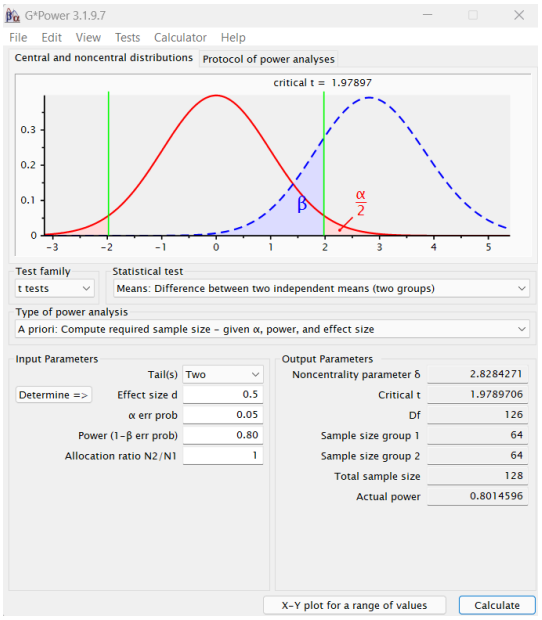


*Figure A1. A-priori power analysis in G*Power*

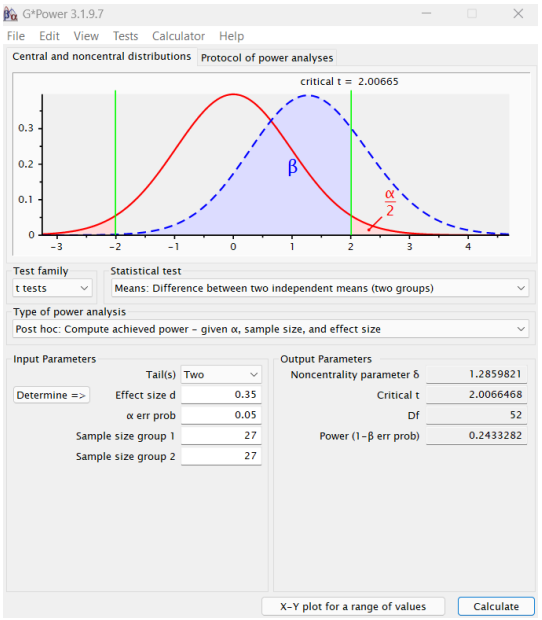


*Figure A2. Post-hoc power analysis in G*Power*

Table A1. Effect size benchmark calculations based on data from Macchi & De Pisapia (2024)

Benchmark for	*Observed measure*	*Effect size (d)*
Productivity	Perceived outcome importance (VR vs VC)	*0.73*
	Absorption by activity (VR vs VC)	0.55
	Task area: attempted answers (VR vs VC)	1.35
	Task area: questions (VR vs VC)	0.96
	Average benchmark estimate for Productivity	0.90
Trust, Cooperation, Team spirit and belonging	Social emotional area: positive reactions	1.04

Notes: Based on the figures above, a conservative assumption of a medium effect size was made for the purpose of Power analysis.

B. Full experimental questionnaires

Information Sheet and Consent Form

Information Sheet and Consent Form

Thank you for considering participating in this study. This page outlines the purpose of the study and provides a description of your involvement and rights as a participant, if you agree to take part.

1. What is the study about?
This study examines group decision-making in technology mediated meetings.

2. Do I have to take part?
Participation is voluntary. It is up to you to decide whether or not to take part. There are no negative consequences for you if you decide not to take part.

3. What will my involvement be?
You will first be asked to complete a task individually, and then you will be asked to collaborate on the same task with up to two other participants. The collaboration will take place over a meeting mediated by technology. Upon completion of the group task, you will be asked to complete a short survey.

The whole session should take **approximately 30 minutes**.

4. How do I withdraw from the study?
You can withdraw from the study at any point during this session without having to give a reason. If any questions during the survey make you feel uncomfortable, you do not have to answer them. Withdrawing from the study will have no effect on you. If you withdraw from the study, the researcher will not retain the information you have given thus far, unless you are happy for them to do so.

5. What will my information be used for?
The analyses of the collected data will be used for a Master's dissertation in LSE's MSc Behavioural Science programme.

6. Will my taking part and my data be kept confidential? Will it be anonymised?
The records from this study will be kept as confidential as possible. Only the researcher and their supervisor will have access to the data. Your data will be anonymised – your name will not be used in any reports or publications resulting from the study.

7. Who has reviewed this study?
This study has undergone ethics review in accordance with the LSE Research Ethics Policy and Procedure.

8. What if I have a question or complaint?
If you have any questions regarding this study please contact on @lse.ac.uk.

If you have any concerns or complaints regarding the conduct of this research, please contact the LSE Research Governance Manager via research.ethics@lse.ac.uk.

If you consent to take part in this study, please confirm below.

- ◯ I consent, let's begin the study.
- ◯ I do not consent.

Pre-Task Survey

Thank you for agreeing to participate in this study!

You will first be asked some questions about yourself.

What is your gender?

- ◯ Male
- ◯ Female
- ◯ Non-binary
- ◯ Prefer not to say

What is your ethnicity?

- ◯ White
- ◯ Black or African American
- ◯ Middle Eastern or North African
- ◯ Asian or Asian American
- ◯ Hispanic or Latino
- ◯ Other, please specify

What is your age?

What is the highest degree or level of school you have completed?
If currently enrolled, highest degree received.

- ◯ High school
- ◯ Post-secondary A level / International Baccalaureate
- ◯ Vocational training
- ◯ Bachelors
- ◯ Masters
- ◯ Doctorate

Do you wear prescription glasses?

- ◯ Yes
- ◯ No
- ◯ Prefer not to say

How tired do you feel at the moment?

- ◯ Not tired at all
- ◯ Slightly tired
- ◯ Moderately tired
- ◯ Very tired
- ◯ Extremely tired

I see myself as someone who...

	Scale				
	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
Relies on others	○	○	○	○	○
Trusts others	○	○	○	○	○
Please select 'Somewhat agree'	○	○	○	○	○
Believes that others have good intentions	○	○	○	○	○

Individual Task

Please take up to 5 minutes to read the instructions and complete the task on the following screen.

It is approximately 10am in mid-August and you have just crash landed in the Sonora Desert, near the Mexico-USA border. The plane has completely burnt out, only the frame remains. Miraculously, all passengers are uninjured but the pilot has been killed.

The pilot was unable to tell anyone of your position before the crash. However, ground sightings taken shortly before the crash suggest that you are about 100 km off the course listed in your flight plan. A few moments before the crash, the pilot indicated you were about 110 km south east of a mining camp. The camp is the nearest known settlement.

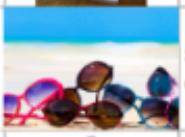
The immediate area is quite flat and, except for the occasional thorn bush and cacti, is rather barren.

Before the plane caught fire, your group was able to save 15 items from the plane.

Your task is:

- to **decide on a course of action** that will maximise your chances of survival, and
- to **rank the 15 items** according to their importance to your survival in the desert, with 1 for the most important, down to 15 for the least important.

Items

	**Torch with 4 battery-cells**		**First-aid kit**		**A book entitled 'Desert Animals That Can Be Eaten'**
	**Folding knife**		**45 calibre pistol (loaded)**		**Sunglasses (for everyone)**
	**Air map of the area**		**Parachute (red & white)**		**2 litres of vodka**
	**Plastic raincoat (large size)**		**Bottle of 1000 salt tablets**		**Overcoat (for everyone)**
	**Magnetic compass**		**2 litres of water per person**		**A cosmetic mirror**

Re-order the below items starting form the most important, down to the least important to your survival in the desert.

- Torch with 4 battery-cells
- Folding knife
- Air map of the area
- Plastic raincoat (large size)
- Magnetic compass
- First-aid kit
- 45 calibre pistol (loaded)
- Parachute (red & white)
- Bottle of 1000 salt tablets
- 2 litres of water per person
- A book entitled 'Desert Animals That Can Be Eaten'
- Sunglasses (for everyone)
- 2 litres of vodka
- Overcoat (for everyone)
- A cosmetic mirror

Group-Task - Videoconferencing (VC) Condition

Thank you for completing the Desert Survival Task!

You will now meet up to 2 other participants over Google Meet (Zoom-like Video-Conferencing cloud-based software) to **discuss as a group the course of action** that will maximise your chances of survival, and to **agree unanimously how you will rank the 15 items** according to their importance to your survival.

You have **up to 15 minutes** to complete this task.

You can now switch to the other open tab. Please **turn on your camera and wait for everyone to have turned on their camera before turning your mic on**. Keep your camera and mic on for the full duration of the meeting.

Once you have completed the task, please close the meeting tab. Refer back to this tab and confirm completion.

◯ I confirm I have completed the group task and I am ready to continue with this study.

Group-Task - Virtual Reality (VR) Condition

Thank you for completing the Desert Survival Task!

You will now be shown tutorials to familiarise yourself with how to use the VR headset and controllers.

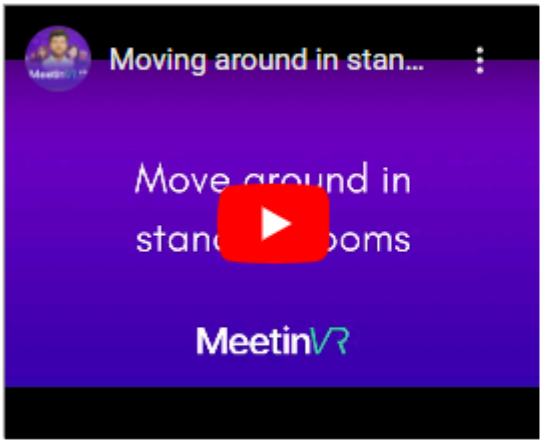
Once you are ready and set up, you will meet up to 2 other participants in VR (using MeetinVR software) to **discuss as a group the course of action** that will maximise your chances of survival, and to **agree unanimously how you will rank the 15 items** according to their importance to your survival. You will have **up to 15 minutes** to complete this task.

In case you have limited experience using VR equipment, please find below a tutorial on **how to use the VR controllers**.

○ I confirm I can correctly hold the controllers.

Please watch the following tutorials showing two key functionalities of the MeetinVR software. You will need those for the group task.

1. How to move around the room in MeetinVR?



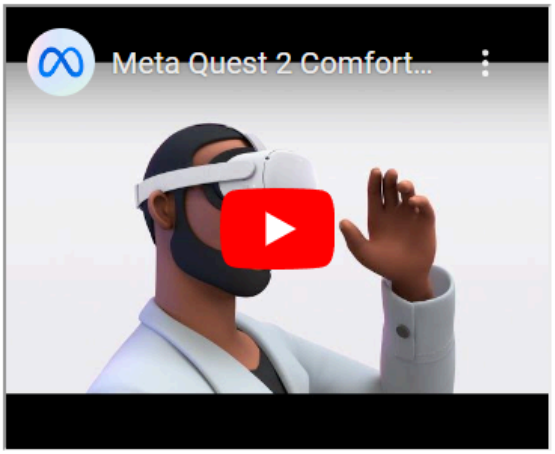
2. How to move objects in MeetinVR?

- ◯ I am now familiar with the key MeetinVR functionalities, let's continue.

You are now ready to put your VR headset on and start the group task.

Once you have adjusted the VR headset to fit comfortably (how-to guidance in video below), **patiently explore the VR environment while waiting for others to join**. Further instructions are provided in the virtual room - the screens in the centre have information about the group task and the screens on the sides have general guidance on how to use the software.

Once you have completed the task, please leave the VR headset on the table and navigate back to this tab.



- ◯ I completed the group task and had NO issues moving the ranks on the board.
- ◯ I completed the group task, but couldn't move the ranks on the board.
- ◯ I couldn't complete the group task due to technical difficulties/issues.

Post-Task Survey

Thank you for completing the group task! You will now be asked to complete a short survey.

Before this study did you know any of the participants in your group?

- ○ Yes
- ○ No

If yes, how many?

- ○ 1
- ○ 2

Do you feel everyone in your group contributed equally to the resolution of the task?

- ○ Yes
- ○ No

How tired do you feel now?

- ○ Not tired at all
- ○ Slightly tired
- ○ Moderately tired
- ○ Very tired
- ○ Extremely tired

How enjoyable did you find the group discussion?

- ○ Not at all enjoyable
- ○ Slightly enjoyable
- ○ Moderately enjoyable
- ○ Very enjoyable
- ○ Extremely enjoyable

How present did you feel during the group discussion?

- ○ Not at all present
- ○ Slightly present
- ○ Moderately present
- ○ Very present
- ○ Extremely present

How would you break down the time your group spent to complete the task? Please use percentages and make sure they sum up to 100.

Trying to figure out how to make the tech work	0
Introductions/Getting to know group members	0
Productive conversation contributing towards the task completion	0
Unproductive conversation (i.e. interrupting, arguing, repeating the same points, etc.)	0
Other	0
Total	0

Please state how much you agree with the following statements with regards to your group.

	Completely disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Completely agree
I can rely on other group members	○	○	○	○	○
I trust other group members	○	○	○	○	○
Please select 'Completely agree'	○	○	○	○	○
I believe other group members have good intentions	○	○	○	○	○
While making a decision we take each other's opinion into consideration	○	○	○	○	○
In this group we work in a climate of cooperation	○	○	○	○	○
In this group we discuss and deal with issues or problems openly	○	○	○	○	○

Please state how much you agree with the following statements with regards to your group.

	Completely disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Completely agree
I feel I am an important part of this group	○	○	○	○	○
Members are committed to the achievement of the group objectives	○	○	○	○	○
Please select 'Completely disagree'	○	○	○	○	○
I feel good about our group activity	○	○	○	○	○
My group develops good and useful ideas	○	○	○	○	○
I feel comfortable in my group	○	○	○	○	○
As a group, we like one another	○	○	○	○	○

Have you used a VR headset in the past?

- ○ Yes
- ○ No

If yes, how many times?

- ○ Only once
- ○ Only twice
- ○ 3-5 times
- ○ 5-10 times
- ○ More than 10 times

C. Desert survival task and survival experts rankings

Task Instructions

It is approximately 10 am in mid-August and you have just crash landed in the Sonoran Desert, near the Mexico-US border. The plane has completely burnt out, only the frame remains. Miraculously, all passengers are uninjured but the pilot has been killed.

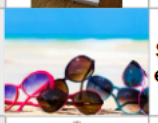
The pilot was unable to tell anyone of your position before the crash. However, ground sightings taken shortly before the crash suggest that you are about 100 km off the course listed in your flight plan. A few moments before the crash, the pilot indicated you were about 110 km south east of a mining camp. The camp is the nearest known settlement.

The immediate area is quite flat and, except for the occasional thorn bush and cacti, is rather barren.

Before the plane caught fire, your group was able to save 15 items from the plane.

You have **up to 15 minutes to discuss as a group on a course of action that will maximise your chances of survival and to unanimously agree on how you will rank the 15 items according to their importance to your survival in the desert**, with 1 for the most important, down to 15 for the least important.

Desert survival experts rankings

Rank	Item		Rank	Item		Rank	Item	
4		Torch with 4 battery-cells	10		First-aid kit	13		A book entitled 'Desert Animals That Can Be Eaten'
6		Folding knife	8		45 calibre pistol (loaded)	9		Sunglasses (for everyone)
12		Air map of the area	5		Parachute (red & white)	14		2 litres of vodka
7		Plastic raincoat (large size)	15		Bottle of 1000 salt tablets	2		Overcoat (for everyone)
11		Magnetic compass	3		2 litres of water per person	1		A cosmetic mirror

D. Data and R code

R code used for the quantitative analyses available here.

Data available upon request.

E. Regression analyses

E1. Covariates and reasons for their inclusion

Table E1.1. R2, R5, R8: List of covariates and reasons for their inclusion/exclusion

Covariates	Included			Reasoning
	R2	R5	R8	
Age	✅	✅	✅	Different age groups may have varying levels of comfort and familiarity with technology (Macchi & De Pisapia, 2024; Staddon, 2020).
Gender	❌	❌	❌	There is no evidence for significant gender differences in information and communication technology use and skills (Qazi et al., 2022).
Ethnicity	❌	❌	❌	Despite the abundant literature on the role of socio-economic status on technology use (Ma, 2021), there is limited research on the role of ethnicity in the context of higher education in the UK.
Level of Education	❌	❌	❌	All participants are LSE students (with the majority pursuing postgraduate studies). Including level of education would increase model complexity without significant explanatory power.
Group size	✅	✅	✅	Group size might impact group dynamics as communication patterns differ in dyads, triads and larger groups (Cooney et al., 2020; McGrath, 1984).
Other participants known before (#)	✅	✅	✅	Group dynamics might differ among strangers and previously known individuals, and their mix within a group (Moreland & Levine, 2002).
Prescription glasses	❌	✅	✅	Participants wearing prescription glasses might experience discomfort or blurrier vision when wearing a VR headset (Güzel et al., 2023).
Tiredness at the end of task	✅	✅	✅	Tiredness might impact performance and engagement with the group task, and the experiment as a whole (Hockey, 2013).
Change in Tiredness	❌	❌	❌	This measure is highly correlated with the measure above (0.71) and might cause multicollinearity if also included in regression analysis.
Baseline - Relies on others	✅	✅	❌	Baseline trust measures control for individual differences among participants which might impact their behaviour in the group task (Dirks & Ferrin, 2001). Not included in R8 as there is no reason to believe it affects propensity for getting tired.
Baseline - Trusts others	✅	✅	❌	
Baseline - Believes in good intentions	✅	✅	❌	
Average individual score in group	❌	✅	❌	The two variables provide a baseline for group resources, existing knowledge within each group, which prior studies have found impact the group performance (Innami, 1994). Average score not included in R2, due to high correlation with group performance (0.66) and individual score (0.56). Not included in R8 as there is no reason to believe it affects propensity for getting tired.
Best individual score in group	✅	✅	❌	
Expertise score	✅	❌	❌	Individual expertise might impact an individuals' behaviour in a group setting and affect group dynamics.
Group performance score	✅	❌	❌	Overall performance in a group might be an indicator of the dynamics within that group.
Finding group discussion enjoyable	✅	✅	✅	Having a positive experience of the group discussion might impact engagement and performance (Geue, 2018).
Feeling present	✅	✅	✅	Feeling present during group discussion might impact engagement and performance (Witmer & Singer, 1998).
Group dynamics	❌	❌	✅	Group dynamics might impact how tired participants feel after the group interactions.
Prior VR experience	✅	✅	✅	Familiarity with VR technology might impact confidence using the technology and performance (Sagnier et al., 2020).
Technical issues	✅	✅	✅	Disruptions due to technical issues might impact task performance and engagement (Sell et al., 2013).
Group task duration	✅	✅	✅	The duration of the group task might control for unobservable group dynamics and characteristics (i.e. efficiency, depth of conversation, team bonding, etc.), and impact how tired participants feel at the end.

E2. Descriptive statistics and balance checks

Table E2.1. Sample characteristics and balance checks

	Mean (SD)/ Obs. (%)			Balance Checks	
Demographic characteristics	***Full Sample (N = 54)***	***Control (VC) (N=27)***	***Treatment (VR=27)***	***Results***	***p-value***
Age	24.78 (3.53)	25.52 (4.00)	24.04 (2.86)	t(47) = 1.56	p = 0.12
Gender				$X^2(1) = 0.67$	p = 0.41
Female	26 (48.15%)	11 (40.74%)	15 (55.56%)		
Male	28 (51.85%)	16 (59.26%)	12 (44.44%)		
Education level				$X^2(3) = 3.18$	p = 0.37
High school	1 (1.85%)		1 (3.70%)		
Post-secondary A level / IB	2 (3.70%)		2 (7.41%)		
Bachelors	36 (66.67%)	19 (70.37%)	17 (62.96%)		
Masters	15 (27.78%)	8 (29.63%)	7 (25.93%)		
Ethnicity				$X^2(4) = 2.06$	p = 0.72
Asian or Asian American	16 (29.63%)	7 (25.93%)	9 (33.33%)		
Hispanic or Latino	8 (14.81%)	5 (18.52%)	3 (11.11%)		
Middle Eastern or North African	1 (1.85%)	1 (3.70%)			
White	20 (37.04%)	9 (33.33%)	11 (40.74%)		
Other	9 (16.67%)	5 (18.52%)	4 (14.81%)		
Other characteristics					
Prescription glasses				$X^2(1) = 0.08$	p = 0.78
Yes	24 (44.44%)	11 (40.74%)	13 (48.15%)		
No	30 (55.56%)	16 (59.26%)	14 (51.85%)		
Trust [1]					
Relies on others [2]	-0.22 (1.18)	-0.19 (1.21)	-0.26 (1.13)	t(52) = 0.23	p = 0.82
Trusts others [3]	0.72 (0.96)	0.70 (1.07)	0.74 (0.86)	t(50) = -0.14	p = 0.89
Believes in good intentions [4]	0.78 (0.84)	0.81 (0.83)	0.74 (0.86)	t(52) = 0.32	p = 0.75
Know other participants in same group				$X^2(1) = 0.75$	p = 0.39
Yes	18 (33.33%)	7 (25.93%)	11 (40.74%)		
No	36 (66.67%)	20 (74.07%)	16 (59.26%)		
Prior VR experience (# of uses)	1.07 (2.16)	1.11 (2.31)	1.04 (2.05)	t(51) = 0.12	p = 0.90

Notes: [1] Measured on 5-point Likert scale (-2 - Strongly disagree; 2 - Strongly agree)
[2] Measured by the question: "I see myself as someone who... - Relies on others"
[3] Measured by the question: "I see myself as someone who... - Trusts others"
[4] Measured by the question: "I see myself as someone who... - Believes that others have good intentions"

E3. Results

Table E3.1. R2: Regression analysis results of VR Treatment impact on Group dynamics, covariates included

Variables	β	SE	95% CI	t	*p*
(Intercept)	1.077	0.857	[-0.658, 2.812]	1.257	0.217
VR Treatment	-0.091	0.249	[-0.596, 0.414]	-0.366	0.717
Group size	-0.072	0.24	[-0.558, 0.413]	-0.302	0.764
Other participants known before (#)	0.017	0.108	[-0.202, 0.236]	0.16	0.874
Tiredness at the end of task	-0.15	0.072	[-0.296, -0.003]	-2.068	0.045 *
Baseline - Relies on others	0.096	0.06	[-0.026, 0.218]	1.59	0.120
Baseline - Trusts others	-0.085	0.081	[-0.25, 0.079]	-1.049	0.301
Baseline - Believes in good intentions	0.005	0.101	[-0.198, 0.209]	0.052	0.959
Technical issues	0.201	0.22	[-0.244, 0.646]	0.913	0.367
Prior VR experience	0.02	0.034	[-0.049, 0.088]	0.579	0.566
Best individual score in group	-0.006	0.015	[-0.035, 0.024]	-0.386	0.701
Expertise score	0.006	0.008	[-0.01, 0.023]	0.795	0.431
Group performance score	-0.002	0.01	[-0.022, 0.017]	-0.253	0.802
Finding group discussion enjoyable	0.094	0.102	[-0.113, 0.301]	0.918	0.364
Feeling present	0.094	0.114	[-0.138, 0.325]	0.821	0.417
Group task duration	0.019	0.021	[-0.025, 0.062]	0.865	0.392

Notes: Results: $F(38) = 1.606$, $p = 0.118$, $R^2 = 0.3881$, Adj. $R^2 = 0.1465$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E3.2. R3: Regression analysis results of VR Treatment Impact on Group dynamics, covariates included, highly non-significant predictors (p>0.5) removed

Variables	β	SE	95% CI	t	p
(Intercept)	0.644	0.524	[-0.412, 1.700]	1.230	0.225
VR Treatment	-0.090	0.197	[-0.488, 0.308]	-0.454	0.652
Tiredness at the end of task	-0.160	0.067	[-0.295, -0.026]	-2.408	0.020 *
Baseline - Relies on others	0.091	0.055	[-0.020, 0.203]	1.653	0.105
Baseline - Trusts others	-0.077	0.070	[-0.218, 0.065]	-1.094	0.280
Technical issues	0.210	0.180	[-0.153, 0.572]	1.167	0.250
Expertise score	0.003	0.007	[-0.010, 0.016]	0.474	0.638
Finding group discussion enjoyable	0.106	0.091	[-0.077, 0.289]	1.171	0.248
Feeling present	0.099	0.104	[-0.111, 0.308]	0.947	0.349
Group task duration	0.017	0.019	[-0.021, 0.055]	0.905	0.370

Notes: Results: $F(44) = 2.861$, $p = 0.010$, $R^2 = 0.3692$, Adj. $R^2 = 0.2401$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E3.3 R5: Regression analysis results of VR Treatment impact on Group performance, covariates included

Variables	β	SE	95% CI	t	p
(Intercept)	-22.259	13.373	[-49.355, 4.837]	-1.664	0.104
VR Treatment	-11.471	3.006	[-17.562, -5.38]	-3.816	<0.001 ***
Age	-0.299	0.278	[-0.861, 0.264]	-1.077	0.289
Prescription glasses	4.749	2.105	[0.484, 9.014]	2.256	0.030 *
Group size	2.995	3.108	[-3.303, 9.293]	0.964	0.342
Other participants known before (#)	2.38	1.418	[-0.493, 5.254]	1.678	0.102
Tiredness at the end of task	1.528	1.011	[-0.519, 3.576]	1.512	0.139
Baseline - Relies on others	0.48	0.809	[-1.16, 2.119]	0.593	0.557
Baseline - Trusts others	0.282	1.073	[-1.892, 2.456]	0.263	0.794
Baseline - Believes in good intentions	-3.113	1.332	[-5.813, -0.414]	-2.337	0.025 *
Average individual score in group	1.101	0.297	[0.499, 1.702]	3.707	0.001 ***
Best individual score in group	-0.053	0.237	[-0.533, 0.428]	-0.222	0.826
Technical issues	-3.823	3.093	[-10.091, 2.444]	-1.236	0.224
Prior VR experience	0.022	0.452	[-0.893, 0.937]	0.049	0.962
Finding group discussion enjoyable	1.827	1.262	[-0.729, 4.383]	1.448	0.156
Feeling present	3.374	1.603	[0.126, 6.621]	2.105	0.042 *
Group task duration	0.578	0.272	[0.028, 1.129]	2.129	0.040 *

Notes: Results: $F(37) = 6.303$, $p < 0.001$, $R^2 = 0.7255$, Adj. $R^2 = 0.6069$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E3.4. R6: Regression analysis results of VR Treatment impact on Group performance, covariates included, highly non-significant predictors (p>0.5) removed

Variables	β	SE	95% CI	t	p
(Intercept)	-21.263	12.562	[-46.633, 4.106]	-1.693	0.098
VR Treatment	-11.572	2.79	[-17.208, -5.937]	-4.147	<0.001 ***
Age	-0.342	0.257	[-0.862, 0.177]	-1.33	0.191
Prescription glasses	4.483	1.968	[0.509, 8.458]	2.278	0.028 *
Group size	2.972	2.766	[-2.614, 8.558]	1.075	0.289
Other participants known before (#)	2.195	1.285	[-0.4, 4.79]	1.709	0.095
Tiredness at the end of task	1.465	0.953	[-0.459, 3.389]	1.538	0.132
Baseline - Believes in good intentions	-3.069	1.193	[-5.477, -0.66]	-2.573	0.014 *
Average individual score in group	1.056	0.203	[0.647, 1.465]	5.214	<0.001 ***
Technical issues	-3.696	2.818	[-9.386, 1.994]	-1.312	0.197
Finding group discussion enjoyable	1.753	1.189	[-0.649, 4.155]	1.474	0.148
Feeling present	3.41	1.516	[0.349, 6.472]	2.249	0.030 *
Group task duration	0.572	0.251	[0.065, 1.079]	2.278	0.028 *

Notes: Results: $F(41) = 8.818$, $p < 0.001$, $R^2 = 0.7207$, Adj. $R^2 = 0.6390$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E3.5. R8: Regression analysis results of VR Treatment impact on Change in tiredness, covariates included

Variables	β	SE	95% CI	t	*p*
(Intercept)	2.447	1.552	[-0.686, 5.581]	1.577	0.122
VR Treatment	0.441	0.377	[-0.32, 1.201]	1.17	0.249
Age	0.016	0.036	[-0.057, 0.089]	0.449	0.656
Group size	-0.039	0.354	[-0.754, 0.677]	-0.109	0.914
Other participants known before (#)	0.16	0.178	[-0.2, 0.52]	0.896	0.376
Prescription glasses	-0.463	0.245	[-0.958, 0.033]	-1.887	0.066
Technical issues	0.343	0.342	[-0.348, 1.035]	1.002	0.322
Prior VR experience	-0.053	0.058	[-0.171, 0.064]	-0.912	0.367
Tiredness at the start of task	-0.455	0.161	[-0.78, -0.129]	-2.818	0.007 **
Group dynamics	-0.589	0.261	[-1.116, -0.062]	-2.258	0.029 *
Finding group discussion enjoyable	-0.169	0.164	[-0.501, 0.163]	-1.026	0.311
Feeling present	-0.246	0.199	[-0.648, 0.157]	-1.232	0.225
Group task duration	-0.019	0.036	[-0.091, 0.053]	-0.538	0.593

Notes: Results: $F(41) = 2.27$, $p = 0.0256$, $R^2 = 0.3992$, Adj. $R^2 = 0.2233$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E3.6. R9: Regression analysis results of VR Treatment impact on Change in tiredness, covariates included, highly non-significant predictors (p>0.5) removed

Variables	β	SE	95% CI	t	p
(Intercept)	2.524	0.653	[1.209, 3.84]	3.868	<0.001 ***
VR Treatment	0.277	0.249	[-0.225, 0.78]	1.113	0.272
Other participants known before (#)	0.153	0.164	[-0.178, 0.484]	0.93	0.358
Prescription glasses	-0.489	0.231	[-0.955, -0.024]	-2.117	0.04 *
Technical issues	0.277	0.312	[-0.351, 0.906]	0.889	0.379
Prior VR experience	-0.052	0.055	[-0.163, 0.06]	-0.935	0.355
Tiredness at the start of task	-0.472	0.154	[-0.783, -0.161]	-3.062	0.004 **
Finding group discussion enjoyable	-0.2	0.151	[-0.504, 0.104]	-1.323	0.193
Feeling present during group discussion	-0.218	0.187	[-0.595, 0.158]	-1.168	0.249
Group dynamics	-0.619	0.248	[-1.119, -0.119]	-2.497	0.016 *

Notes: Results: $F(44) = 3.156$, $p = 0.005$, $R^2 = 0.3923$, Adj. $R^2 = 0.2680$
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

E4. Assumptions checks

E4.1. R1: Group dynamics - Simple linear regression

- Linearity

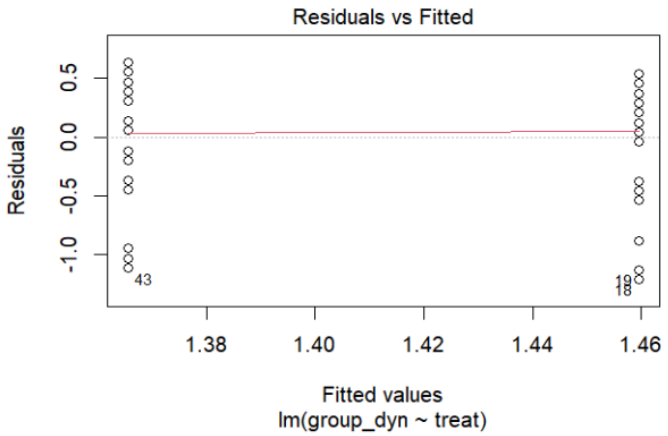


Figure E4.1.1. Residuals vs Fitted plot

Notes: The fitted line is almost horizontal at 0, indicating the linearity assumption is met.

- Independence

Durbin-Watson Test

The data met the assumption of independent errors as the Durbin-Watson value was close to 2 (Durbin-Watson value = 2.02, p-value = 0.4673).

- Homoscedasticity

Breusch-Pagan test

The data met the assumption of homoscedasticity.
(Breusch-Pagan value = 0.199, p-value = 0.6613).

- Normality

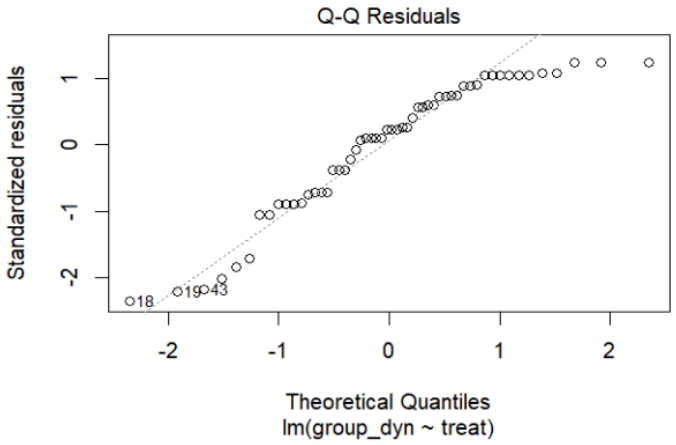


Figure E4.1.2. Normal probability plot of residuals

Notes: The data did not meet the assumption of normality as the standardised residuals did not closely follow the normality line.

- Influential outliers

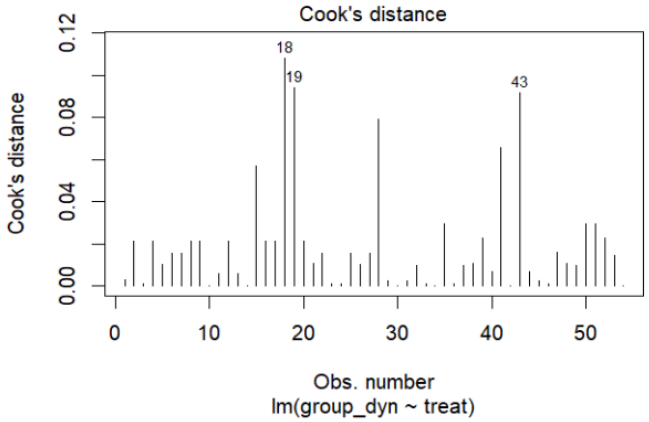


Figure E4.1.3. Cook's distances versus observation number plot

Notes: There are no influential outliers as all data points have lower Cook's distance than 0.5.

E4.2. R2: Group dynamics - Multiple linear regression, with selected covariates

- Linearity

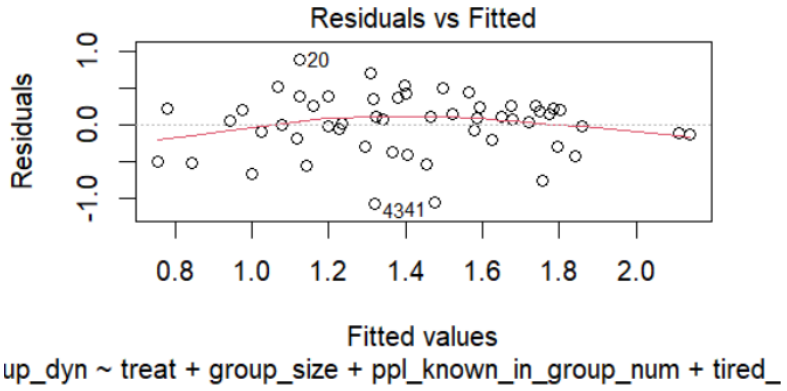


Figure E4.2.1. Residuals vs Fitted plot

Notes: The fitted line is almost horizontal at 0, indicating the linearity assumption is met.

- Independence

Durbin-Watson Test

The data met the assumption of independent errors as the Durbin-Watson value was close to 2 (Durbin-Watson value = 2.18, p-value = 0.4831).

- Homoscedasticity

Breusch-Pagan test

The data met the assumption of homoscedasticity.
(Breusch-Pagan value = 10.317, p-value = 0.7993).

- Normality

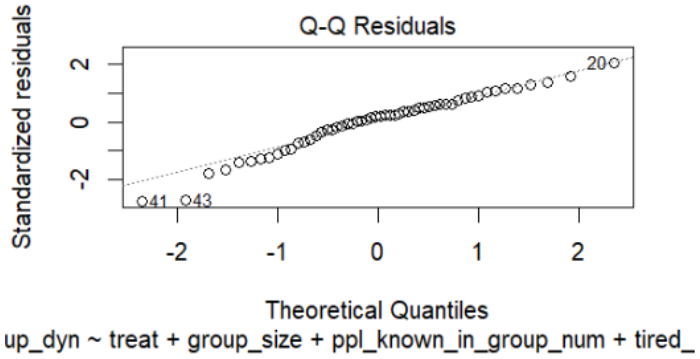


Figure E4.2.2. Normal probability plot of residuals

Notes: The data meets the assumption of normality as the standardised residuals closely follow the normality line.

- Influential outliers

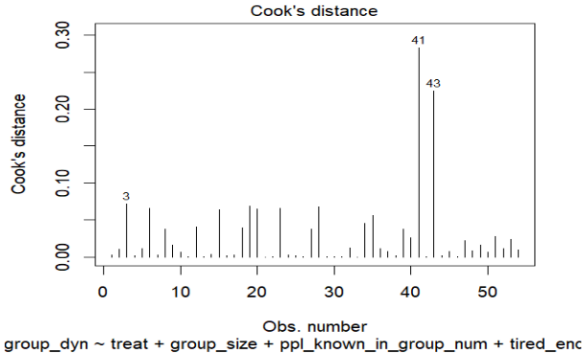


Figure E4.2.3. Cook's distances versus observation number plot

Notes: There are no influential outliers as all data points have lower Cook's distance than 0.5.

- Multicollinearity

Table E4.2.1. R2: Variance Inflation Factors (VIFs) for Group dynamic predictors

Variable	VIF
VR Treatment	3.67
Group size	1.71
Other participants known before (#)	1.4
Tiredness at the end of task	1.36
Baseline - Relies on others	1.17
Baseline - Trusts others	1.42
Baseline - Believes in good intentions	1.65
Technical issues	2.19
Prior VR experience	1.23
Best individual score in group	2.79
Expertise score	1.59
Group performance score	2.17
Finding group discussion enjoyable	1.89
Feeling present during group discussion	1.34
Group task duration (mins)	3.66

Notes: No multicollinearity assumption is met as all VIFs were below the threshold of 5 (Daoud, 2017).

E4.3. R3: Group dynamics - Multiple linear regression, with selected covariates, highly non-significant predictors (p>0.5) removed

- Linearity

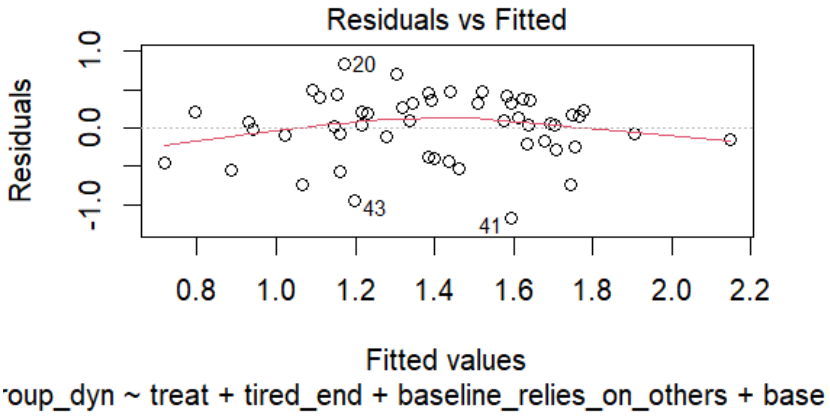


Figure E4.3.1. Residuals vs Fitted plot

Notes: The fitted line is almost horizontal at 0, indicating the linearity assumption is met.

- Independence

Durbin-Watson Test

The data met the assumption of independent errors as the Durbin-Watson value was close to 2 (Durbin-Watson value = 2.11, p-value = 0.5515).

- Homoscedasticity

Breusch-Pagan test

The data met the assumption of homoscedasticity.
(Breusch-Pagan value = 8.5676, p-value = 0.4781).

- Normality

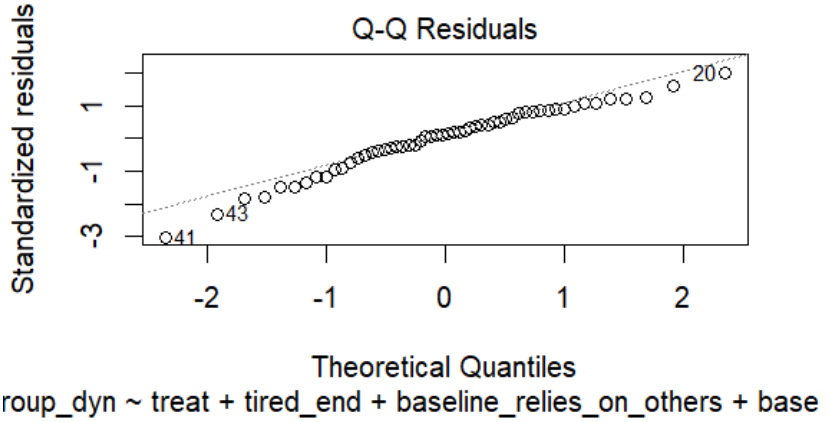


Figure E4.3.2. Normal probability plot of residuals

Notes: The data meets the assumption of normality as the standardised residuals closely follow the normality line.

- Influential outliers

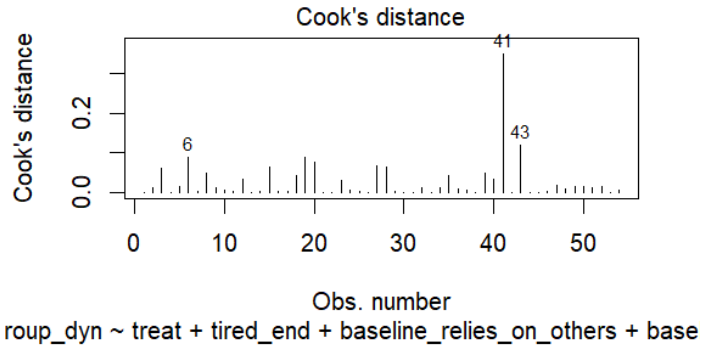


Figure E4.3.3. Cook's distances versus observation number plot

Notes: There are no influential outliers as all data points have lower Cook's distance than 0.5.

- Multicollinearity

Table E4.3.1. R3: Variance Inflation Factors (VIFs) for Group dynamic predictors

Variable	VIF
VR Treatment	2.58
Tiredness at the end of task	1.29
Baseline - Relies on others	1.1
Baseline - Trusts others	1.18
Technical issues	1.65
Expertise score	1.19
Finding group discussion enjoyable	1.68
Feeling present during group discussion	1.25
Group task duration (mins)	3.2

Notes: No multicollinearity assumption is met as all VIFs were below the threshold of 5 (Daoud, 2017).

E4.4. R4: Group performance - Simple linear regression

- Linearity

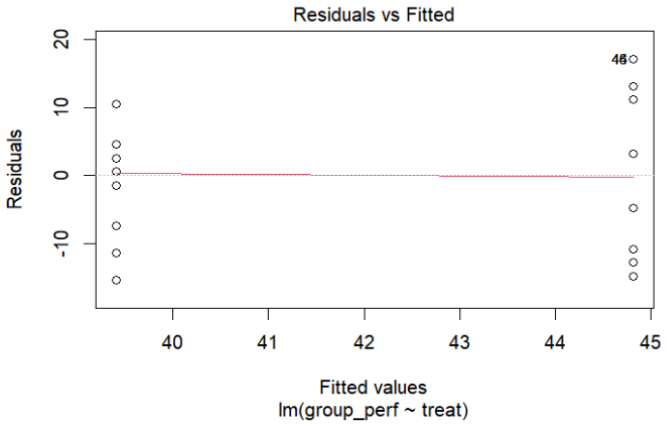


Figure E4.4.1. Residuals vs Fitted plot

Notes: The fitted line is almost horizontal at 0, indicating the linearity assumption is met.

- Independence

Durbin-Watson Test

The data did not meet the assumption of independent errors as the Durbin-Watson value was not close to 2 (Durbin-Watson value = 1.03, p-value <0.001).

- Homoscedasticity

Breusch-Pagan test

The data did not meet the assumption of homoscedasticity
(Breusch-Pagan value = 10.29, p-value = 0.001).

- Normality

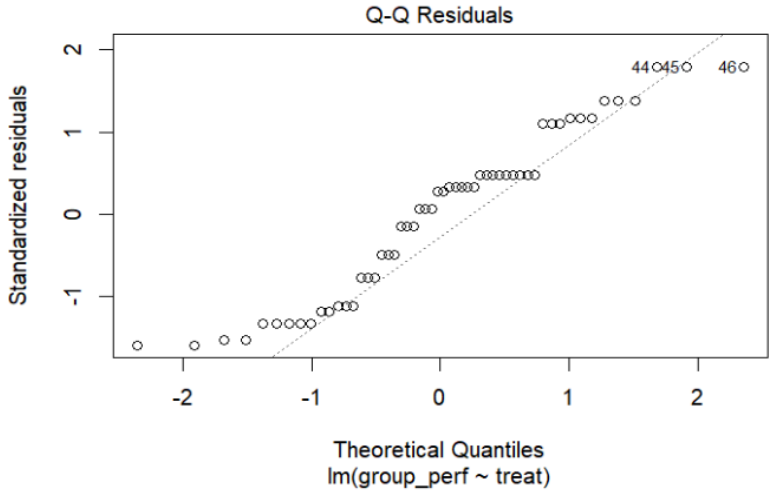


Figure E4.4.2. Normal probability plot of residuals

Notes: The data did not meet the assumption of normality as the standardised residuals did not closely follow the normality line.

- Influential outliers

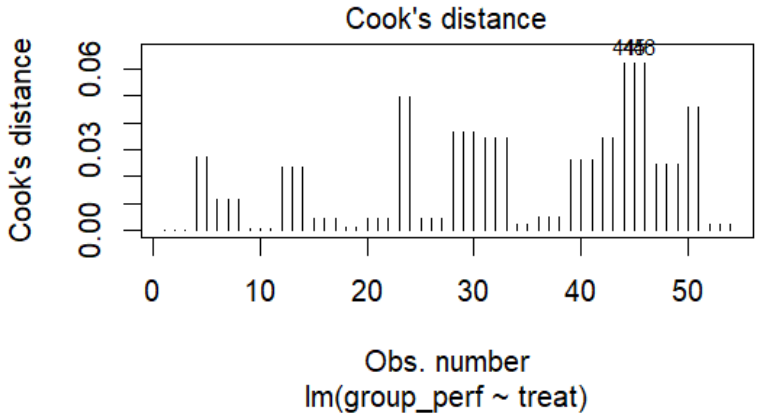


Figure E4.4.3. Cook's distances versus observation number plot

Notes: There are no influential outliers as all data points have lower Cook's distance than 0.5.

E4.5. R5: Group performance - Multiple linear regression, with selected covariates

- Linearity

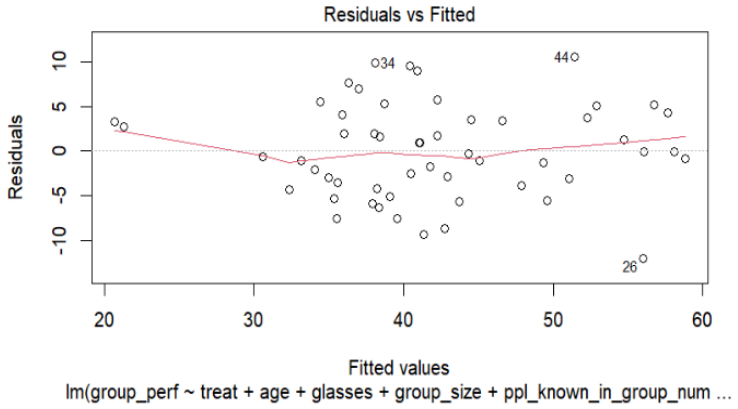


Figure E4.5.1. Residuals vs Fitted plot

Notes: The fitted line is almost horizontal at 0, indicating the linearity assumption is met.

- Independence

Durbin-Watson Test

The data did not meet the assumption of independent errors as the Durbin-Watson value was not close to 2 (Durbin-Watson value = 1.37, p-value <0.001).

- Homoscedasticity

Breusch-Pagan test

The data met the assumption of homoscedasticity
(Breusch-Pagan value = 14.54, p-value = 0.5582).

- Normality

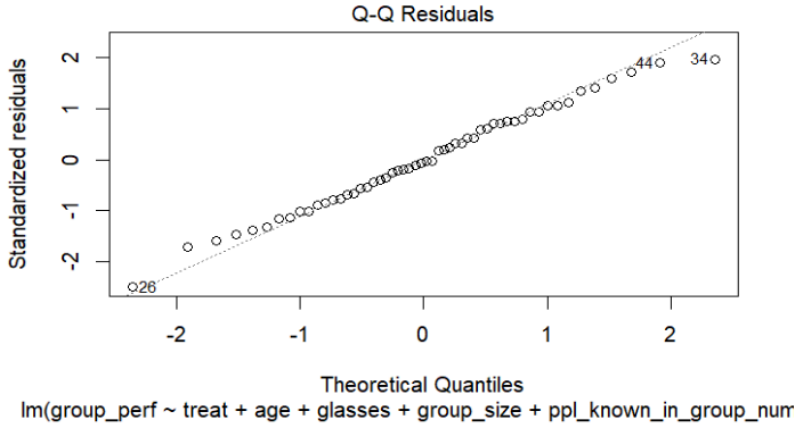


Figure E4.5.2. Normal probability plot of residuals

Notes: The data meets the assumption of normality as the standardised residuals closely follow the normality line.

- Influential outliers

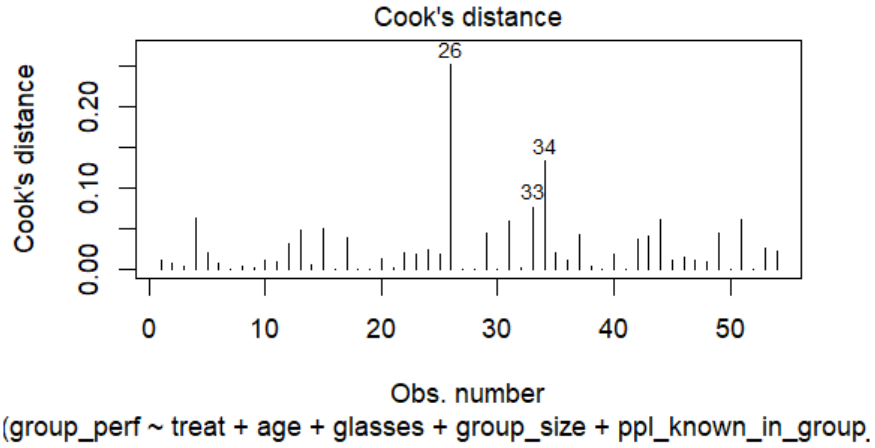


Figure E4.5.3. Cook's distances versus observation number plot

Notes: There are no influential outliers as all data points have lower Cook's distance than 0.5.

- Multicollinearity

Table E4.5.1. R5: Variance Inflation Factors (VIFs) for Group performance predictors

Variable	**VIF**
VR Treatment	3.07
Age	1.28
Prescription glasses	1.49
Group size	1.66
Other participants known before (#)	1.39
Tiredness at the end of task	1.53
Baseline - Relies on others	1.21
Baseline - Trusts others	1.42
Baseline - Believes in good intentions	1.67
Average individual score in group	3.88
Best individual score in group	4.16
Technical issues	2.50
Prior VR experience	1.27
Finding group discussion enjoyable	1.66
Feeling present during group discussion	1.52
Group task duration (mins)	3.39

Notes: No multicollinearity assumption is met as all VIFs were below the threshold of 5 (Daoud, 2017).

E4.6. R6: Group performance - Multiple linear regression, with selected covariates, highly non-significant predictors (p>0.5) removed

- Linearity

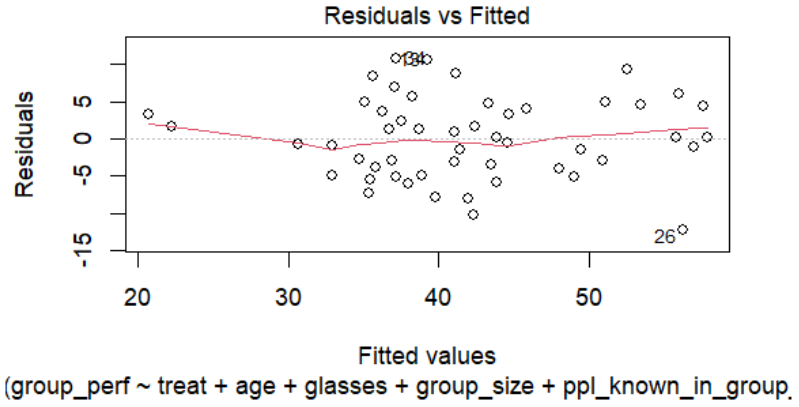


Figure E4.6.1. Residuals vs Fitted plot

Notes: The fitted line is almost horizontal at 0, indicating the linearity assumption is met.

- Independence

Durbin-Watson Test

The data did not meet the assumption of independent errors as the Durbin-Watson value was not close to 2 (Durbin-Watson value = 1.31, p-value <0.001).

- Homoscedasticity

Breusch-Pagan test

The data met the assumption of homoscedasticity.
(Breusch-Pagan value = 10.091, p-value = 0.608).

- Normality

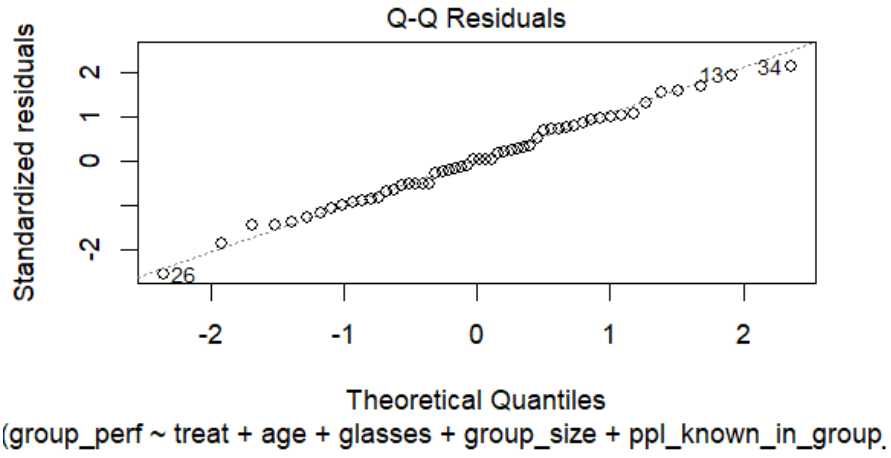


Figure E4.6.2. Normal probability plot of residuals

Notes: The data meets the assumption of normality as the standardised residuals closely follow the normality line.

- Influential outliers

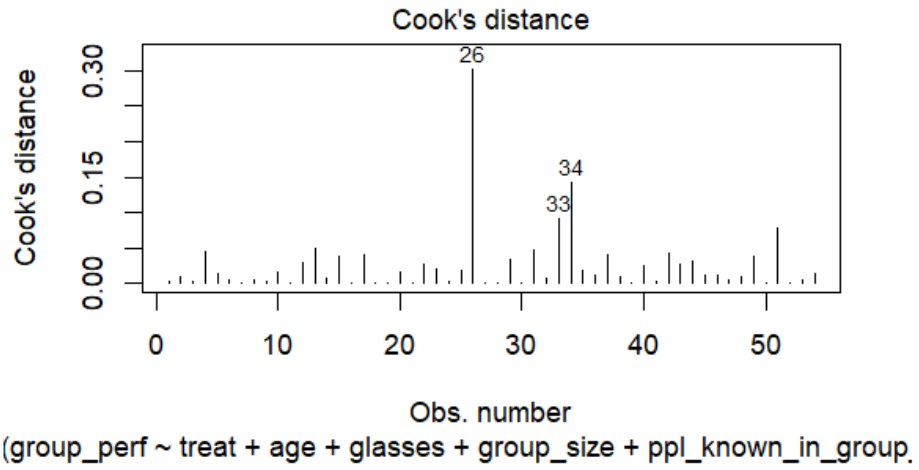


Figure E4.6.3. Cook's distances versus observation number plot

Notes: There are no influential outliers as all data points have lower Cook's distance than 0.5.

- Multicollinearity

Table E4.6.1. R6: Variance Inflation Factors (VIFs) for Group performance predictors

Variable	VIF
VR Treatment	2.88
Age	1.2
Prescription glasses	1.42
Group size	1.43
Other participants known before (#)	1.24
Tiredness at the end of task	1.48
Baseline - Believes in good intentions	1.46
Average individual score in group	1.96
Technical issues	2.26
Finding group discussion enjoyable	1.61
Feeling present during group discussion	1.48
Group task duration (mins)	3.15

Notes: No multicollinearity assumption is met as all VIFs were below the threshold of 5 (Daoud, 2017).

E4.7. R7: Change in tiredness - Simple linear regression

- Linearity

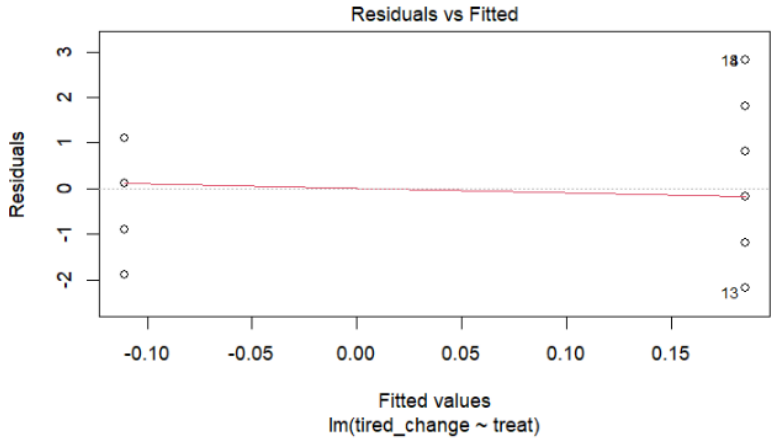


Figure E4.7.1. Residuals vs Fitted plot

Notes: The fitted line is almost horizontal at 0, indicating the linearity assumption is met.

- Independence

Durbin-Watson Test

The data met the assumption of independent errors as the Durbin-Watson value was close to 2 (Durbin-Watson value = 2.06, p-value = 0.527).

- Homoscedasticity

Breusch-Pagan test

The data did not meet the assumption of homoscedasticity
(Breusch-Pagan value = 5.391, p-value = 0.02).

- Normality

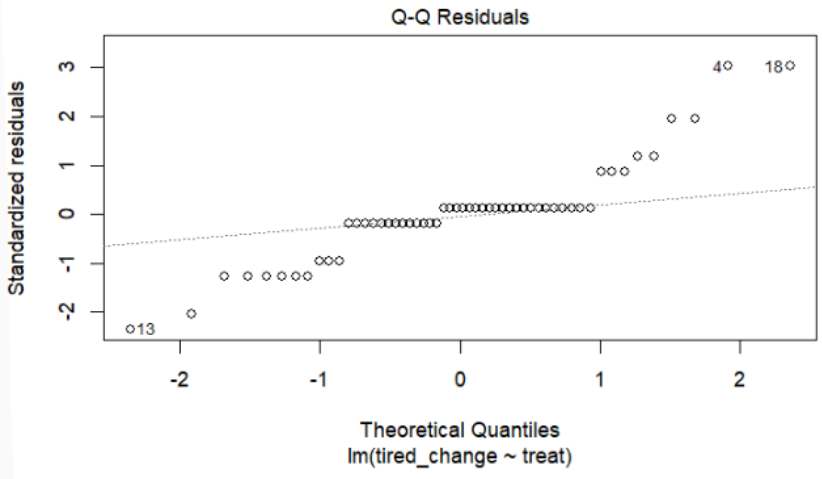


Figure E4.7.2. Normal probability plot of residuals

Notes: The data did not meet the assumption of normality as the standardised residuals did not closely follow the normality line.

- Influential outliers

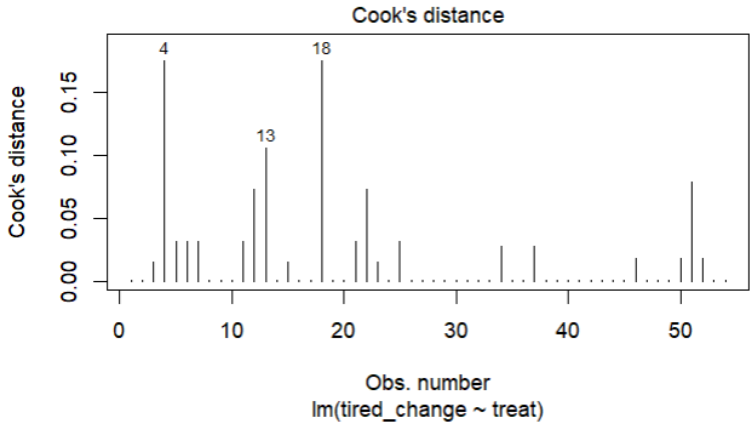


Figure E4.7.3. Cook's distances versus observation number plot

Notes: There are no influential outliers as all data points have lower Cook's distance than 0.5.

E4.8. R8: Change in tiredness - Multiple linear regression, with selected covariates

- Linearity

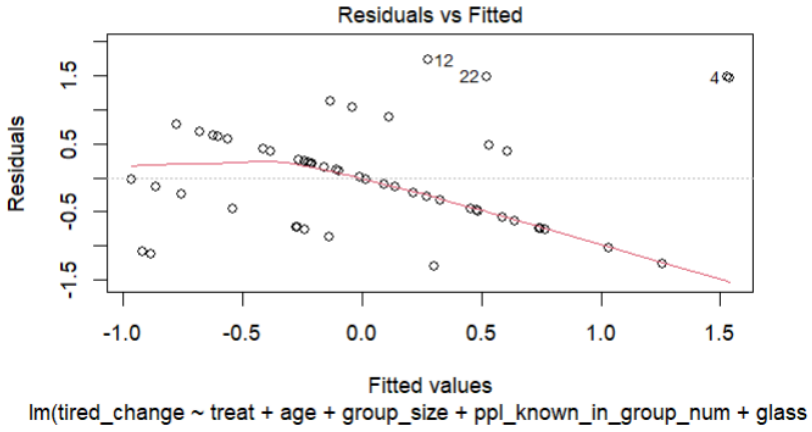


Figure E4.8.1. Residuals vs Fitted plot

Notes: The fitted line is not almost horizontal at 0, indicating the linearity assumption is not met.

- Independence

Durbin-Watson Test

The data met the assumption of independent errors as the Durbin-Watson value was close to 2 (Durbin-Watson value = 2.13, p-value = 0.4641).

- Homoscedasticity

Breusch-Pagan test

The data did not meet the assumption of homoscedasticity.
(Breusch-Pagan value = 22.11, p-value = 0.0364).

- Normality

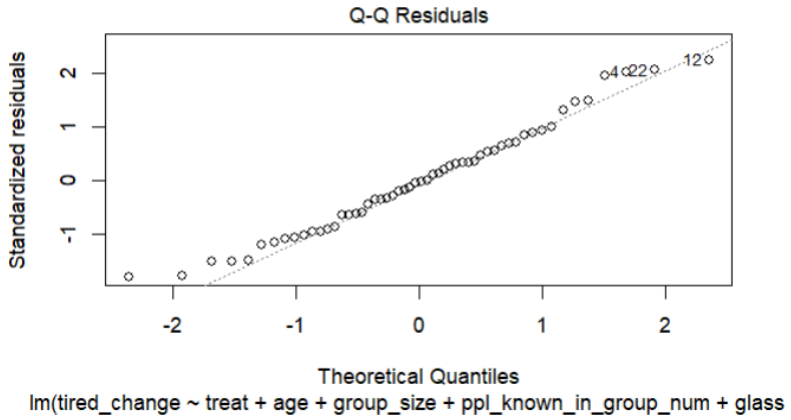


Figure E4.8.2. Normal probability plot of residuals

Notes: The data meets the assumption of normality as the standardised residuals closely follow the normality line.

- Influential outliers

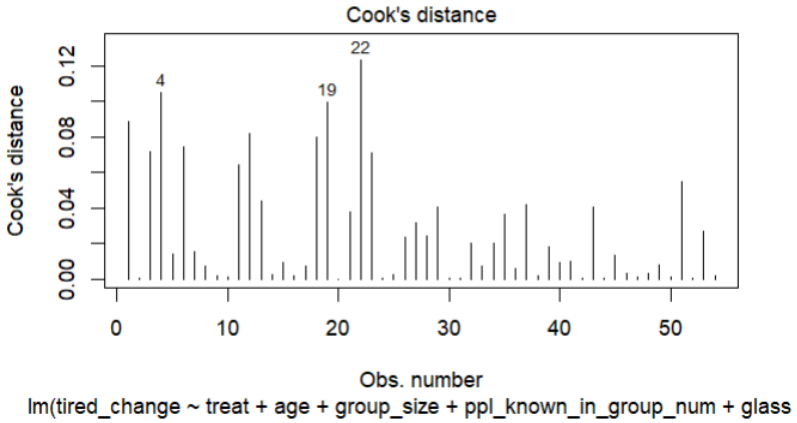


Figure E4.8.3. Cook's distances versus observation number plot

Notes: There are no influential outliers as all data points have lower Cook's distance than 0.5.

- Multicollinearity

Table E4.8.1. R8: Variance Inflation Factors (VIFs) for Change in tiredness predictors

Variable	VIF
VR Treatment	2.72
Age	1.22
Group size	1.22
Number of other participants known before	1.24
Prescription glasses	1.14
Technical issues	1.73
Prior VR experience	1.2
Tiredness at the start of task	1.17
Group dynamics	1.38
Finding group discussion enjoyable	1.6
Feeling present during group discussion	1.33
Group task duration (mins)	3.3

Notes: No multicollinearity assumption is met as all VIFs were below the threshold of 5 (Daoud, 2017).

E4.9. R9: Change in tiredness - Multiple linear regression, with selected covariates, highly non-significant predictors (p>0.5) removed

- Linearity

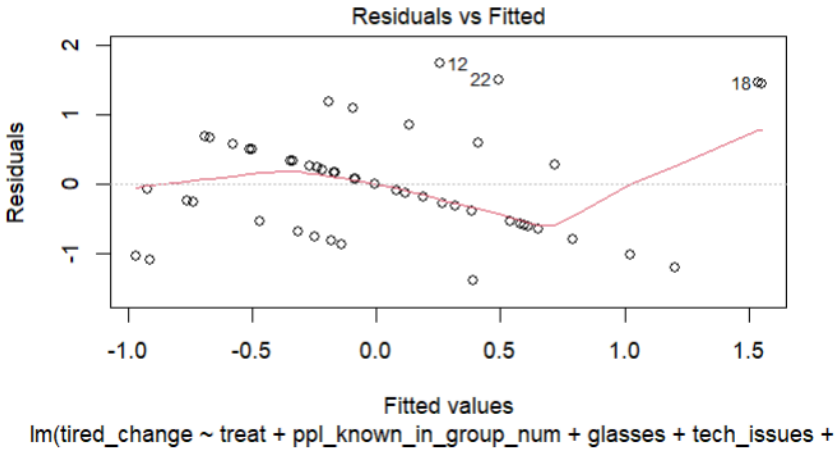


Figure E4.9.1. Residuals vs Fitted plot

Notes: The fitted line is not almost horizontal at 0, indicating the linearity assumption is not met.

- Independence

Durbin-Watson Test

The data met the assumption of independent errors as the Durbin-Watson value was close to 2 (Durbin-Watson value = 2.09, p-value = 0.5169).

- Homoscedasticity

Breusch-Pagan test

The data did not meet the assumption of homoscedasticity (Breusch-Pagan value = 20.997, p-value = 0.0126).

- Normality

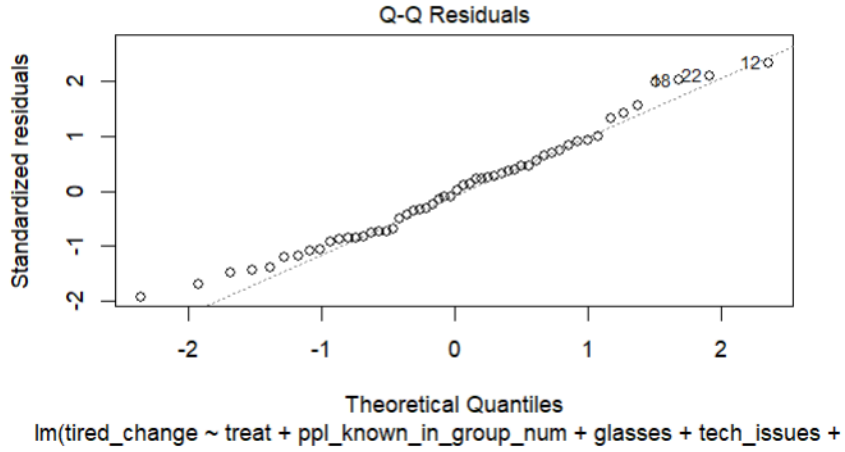


Figure E4.9.2. Normal probability plot of residuals

Notes: The data meets the assumption of normality as the standardised residuals closely follow the normality line.

- Influential outliers

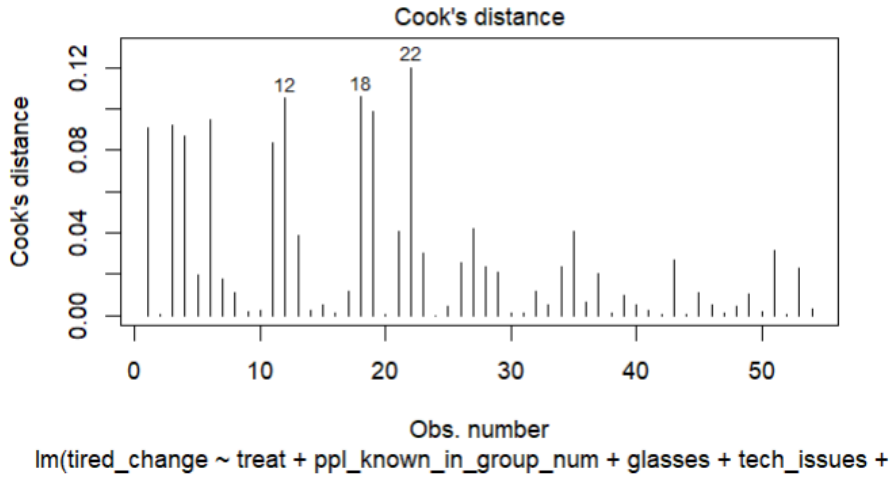


Figure E4.9.3. Cook's distances versus observation number plot

Notes: There are no influential outliers as all data points have lower Cook's distance than 0.5.

- Multicollinearity

Table E4.9.1. R9: Variance Inflation Factors (VIFs) for Change in tiredness predictors

Variable	**VIF**
VR Treatment	1.27
Number of other participants known before	1.12
Prescription glasses	1.07
Technical issues	1.52
Prior VR experience	1.15
Tiredness at the start of task	1.14
Finding group discussion enjoyable	1.43
Feeling present during group discussion	1.24
Group dynamics	1.32

Notes: No multicollinearity assumption is met as all VIFs were below the threshold of 5 (Daoud, 2017).