It Takes Two: Assessing collaborative brainstorming tasks in Virtual Reality and desktop environments

Jenny Horridge, Bibek Khattri,Maite Frutos-Pascual

and Ian Williams[1]

[1]College of Computing, STEAMHouse, Birmingham City University, Birmingham, Birmingham, UK, B4 7RQ

Abstract

Collaboration builds the collective development of new ideas, perspectives, and ways of working. By engaging in shared collaborative tasks, teams can strengthen professional relationships and improve creative outcomes. Since the COVID-19 pandemic, the adoption of digital tools for remote collaboration has grown rapidly, with desktop-based applications now widely used to manage teams, conduct workshops, and support ideation. In parallel, advances in Virtual Reality (VR) technology and the availability of affordable commercial headsets have enabled immersive tools for collaborative work. While prior research has shown that VR can effectively support team-based activities such as training and simulation, its comparative value for creative collaboration relative to traditional desktop platforms remains under explored

This study addresses this gap by evaluating collaborative brainstorming across Desktop (DT) and VR platforms, investigating the extent to which VR enables collaboration comparable to that in traditional desktop environments. Using a triangulated methodology combining task performance, sentiment analysis, and thematic evaluation, we examine how teams collaborate across these modalities. Participants (n = 56) worked in pairs (n_{pairs} = 28) to collectively generate ideas. Results show no significant performance differences between DT and VR, with both conditions producing comparable levels of idea generation and communication. However, team workload was higher in VR, suggesting increased perceived task effort. Overall, findings indicate that VR can support collaborative brainstorming comparable to DT while offering an immersive alternative. The study contributes a mixed-method framework for assessing collaboration across immersive and non-immersive systems, providing empirical insight into methodological triangulation in VR research.

Corresponding author: Ian Williams, ian.williams@bcu.ac.uk

Introduction

Collaboration is a fundamental component of knowledge development and innovation (To & Ko, 2016). It supports collective decision making (Brownell & Chung, 2001) the exchange of diverse perspectives (Stahl & Maznevski, 2021), and can evoke positive emotions such as enjoyment and pride among team members (Chen & Li, 2021).

Recent advances in consumer-ready Virtual Reality (VR) have introduced opportunities for more immersive and interactive collaborative experiences. VR now enables teams to meet, share ideas, learn (van der Meer, van der Werf, Brinkman, & Specht, 2023), and train (Naranjo et al., 2020) within shared virtual environments. Prior studies suggest that VR collaboration can enhance users' sense of communication and engagement compared to desktop (DT) based video calls (Sanaei, Machacek, Gilbert, Wu, & Oliver, 2023), offering levels of immersion and social presence approaching "face to face" interaction (Doroudian, 2023; Joo, Han, Park, & Chung, 2024). Nevertheless, VR still poses challenges such as increased cognitive load, cybersickness (Hove & Watson, 2022), and difficulties interpreting non-verbal cues (Karl, Peluchette, & Aghakhani, 2022; Tomprou, Kim, Chikersal, Woolley, & Dabbish, 2021).

Given the continued shift toward digital collaboration, including the growing use of VR (Kaiser, Walters, Sheehy, Murray, & Spencer, 2024), and the limited understanding of its advantages and limitations (Hamad & Jia, 2022), there is a timely requirement to examine how VR compares to established DT platforms in supporting creative teamwork. Furthermore, as "face to face" collaboration is known to evoke strong feelings of partnership working and shared accomplishment (Chen & Li, 2021; Leng & Guo, 2018), it is important to determine whether VR can elicit these, or comparable experiences.

This study therefore investigates the question: *To what extent does VR enable collaboration comparable to that in traditional DT environments during brainstorming tasks?* We contribute a triangulated methodological framework that integrates performance analysis, computational sentiment assessment, and qualitative thematic evaluation to examine collaborative behavior. The study offers an empirical comparison of collaboration quality between VR and DT systems and identifies design and workload implications for immersive collaborative tools.

The paper is structured as follows: Initially we present the current literature in collaborative systems illustrating lack of work directly comparing immersive VR and DT collaboration for creative brainstorming tasks. We then present the methodology for the study, illustrating the task, protocol, environment and metrics. We then report on the results where we triangulate the quantitive performance of the brainstorming task, the sentiment analysis of the user conversations and qualitative analysis of the conversations. A formal discussion and conclusion is then given where the the findings explore the commonalties found in the DT and VR conditions.

Related Work

Digital collaboration

Effective collaboration supports creative idea generation and innovation (Morgan & Castle, 2023). To achieve this, teams must engage in interaction, information sharing, and establish clear roles and responsibilities (Thompson et al., 2021). Such dynamics are well supported in traditional face to face settings but are increasingly now facilitated through digital collaboration platforms. The growing accessibility of consumer-grade Virtual Reality (VR) systems, such as the Meta Quest 3, now extends these opportunities further, with VR platforms now enabling teams to collaborate, share virtual spaces, and even simulate office-like working environments.

Collaboration in VR

Studies have shown that collaboration in virtual environments can strengthen working relationships and promote positive interdependence among team members (Laal, 2013). In particular, users who may not typically assume leadership roles have demonstrated greater willingness to take initiative when collaborating through head-mounted displays (HMDs) (Slater, Sadagic, Usoh, & Schroeder, 2000). Additionally, VR is well known for its ability to help users visualize complex, abstract, or otherwise inaccessible concepts and support learning experiences that would be difficult or too costly to replicate in reality (Doroudian, 2023; Dunmoye, May, & Hunsu, 2022). Consequently, VR has been widely adopted within science, technology, engineering, and mathematics (STEM) disciplines to facilitate group projects and collaborative problem-solving (Thompson et al.,

2021; Xu, Zhang, Wang, Liu, & Zhu, 2023). Beyond education, VR is increasingly used to simulate professional workspaces, notably virtual offices and simulated meeting environments, therefore allowing teams to brainstorm, present, and collaborate remotely in shared immersive settings (Brendel & Williams, 2013; He, Du, & Perlin, 2020; Sharma, Shroff, & Dewan, 2011; Tate, 2011).

Despite these benefits, VR presents several challenges for effective collaboration. Firstly the technical proficiency of VR users can vary between, therefore limiting participation (Doroudian, 2023). Secondly, system latency may cause delays or missed cues during conversation (Roberts, Duckworth, Moore, Wolff, & O'Hare, 2009) and the absence of clear environmentally spatialized audio or body language can make it difficult to identify speakers or interpret non-verbal signals (Dickson, Lee-Partridge, Limayem, & Desanctis, 1996). Thirdly, prolonged VR use can cause cybersickness (Caserman, Garcia-Agundez, Gámez Zerban, & Göbel, 2021; Kim, Won, Jang, & Ko, 2022; Oh & Son, 2022) and complex task demands or interface interactions, can increase user cognitive load (Dickson et al., 1996; Makransky, Terkildsen, & Mayer, 2019; Parmar et al., 2016).

Therefore, while VR offers a clear potential for enhancing collaboration through immersion and shared spatial environments, it also introduces usability and accessibility barriers. Prior research examining the relationship between system richness and collaboration outcomes has produced mixed results (Cummings & Bailenson, 2016; Steed, Pan, Zisch, & Steptoe, 2016). Building on this the present study empirically evaluates whether differences between VR and desktop (DT) platforms translate into measurable variations in teamwork experience and creative output quality.

Collaborative brainstorming on digital platforms

With growing technological advancement and market competition, there is increasing pressure on organizations to generate a large number of high-quality ideas (Amabile, 1997; Seeber, de Vreede, Maier, & Weber, 2017). Brainstorming remains one of the most established methods for creative ideation, however for brainstorming to be effective, ideas should be generated freely either individually or collectively as part of a team (Galdámez, Sanz, & Armando De Giusti, 2012). Generating a high quantity of ideas increases the likelihood of producing

valuable solutions (Osborn, 1963), while idea quality tends to improve when sessions are goal oriented rather than informal discussions (Nunamaker, Dennis, Valacich, Vogel, & George, 1991).

Digital platforms have increasingly been used to support brainstorming, often replicating traditional tools such as whiteboards and sticky notes (Moulin, Sugawara, Kaeri, Fujita, & Abel, 2014). These tools are widely adopted in educational and professional contexts for collecting and organizing ideas (Siegle, 2020). They are considered advantageous for promoting a focus on idea generation rather than evaluation (Kettler, Lamb, & Mullet, 2019), and some studies suggest that technology supported collaboration can enhance originality and novelty in outcomes (Dennis & Wixom, 2002; Fjermestad & Hiltz, 1998). However, despite these benefits, limited technical support and usability challenges continue to restrict the widespread adoption of digital collaboration tools (Kowalkiewicz, 2012).

Building on this foundation, recent work by Khattri, Chantratita, Frutos-Pascual, and Williams (2025) examined the use of Virtual Reality (VR) as a platform for collaborative brainstorming, exploring sentiment based measures as indicators of team performance. Their study involved small collaborative pairs ($N_{pairs} = 7$) working in VR and investigated the correlation between spoken sentiment polarity and perceived team workload, measured via the Team Workload Questionnaire (TWLQ) (Sellers, Helton, Naswall, Funke, & Knott, 2014). A positive correlation between sentiment and workload indicated the potential value of integrating language and performance based metrics to assess collaboration quality.

While prior research has explored the usability of collaborative tools, there remains a clear gap in understanding user experience and collaborative performance during brainstorming across VR and DT platforms. The present study addresses this gap by applying a triangulated evaluation of performance, sentiment, and thematic factors to compare collaboration in both environments.

Evaluating digital collaboration

A criticism in VR research is the over reliance on post study evaluations to measure user experience (Paulsen, Dau, & Davidsen, 2024). Specifically, that when user data are collected solely through questionnaires or surveys after immersion, results may fail to accurately represent the user when "in experience" (Joo et al., 2024).

Moreover, participants may struggle to recall or report emotions once removed from the immersive environment (Kozachek, Yaqin, Prasad, & Wang, 2023). To address these limitations, researchers have begun employing alternative methods such as speech transcription, facial-expression tracking, and physiological sensing to capture real-time user responses (Dzedzickis, Kaklauskas, & Bucinskas, 2020; Soleymani, Asghari-Esfeden, Fu, & Pantic, 2016).

Slater et al. (2022) recently emphasised this issue, arguing that post task questionnaires, while convenient, do not fully encapsulate the complexity of VR experience. They issued a "call to action" which requested the triangulation of multiple evaluation approaches that integrate quantitative and qualitative data, including sentiment-based measures (Slater et al., 2023) be used in immersive technology (VR) research. Responding to this, Khattri et al. (2025) demonstrated a positive correlation between spoken sentiment polarity and team workload, as measured by the Team Workload Questionnaire (TWLQ) (Sellers et al., 2014), reinforcing the potential of sentiment analysis as a complementary assessment method in VR studies.

Sentiment analysis involves evaluating the polarity of language to determine whether statements are positive, negative, or neutral (Taboada, 2016). In user experience research, this often involves preprocessing conversational data and using computational models to identify frequently occurring terms. Although this technique effectively captures general attitudes (van Atteveldt, van der Velden, & Boukes, 2021), its accuracy can depend on the contextual fit of the sentiment dictionaries used (Bordoloi & Biswas, 2023) with often domain-specific lexicons enhancing precision.

Beyond sentiment polarity, several studies combine sentiment analysis with thematic or emotion analysis to provide richer insights into user experience (Joo et al., 2024; Makki, Alhalabi, & Adham, 2019). Thematic analysis enables the categorisation of opinions by human factor considerations, while emotion analysis identifies feelings such as happiness, frustration, or empathy, which influence group dynamics and cooperation (Eisenberg & Miller, 1987; Shen, Zhang, Ou, Zhao, & Zhou, 2021; Stephan & Finlay, 1999). However, accurately classifying emotion remains difficult, and some participants may hesitate to express genuine feelings in unfamiliar social settings (Jing et al., 2023).

Existing research demonstrates the promise of sentiment and emotion analysis for evaluating VR experiences, yet little is known about how user sentiment during collaboration in VR compares to that in traditional DT environments. Addressing this gap is essential, given the increasing shift from "face to face" to virtual team based collaboration. The present study therefore examines how the platform (VR or DT) influences the usability and experience of collaborative brainstorming, and proposes design recommendations for optimising immersive teamwork.

Methodology

This study presents a comprehensive triangulation of objective task performance, sentiment, and spoken conversation in collaborative brainstorming across Desktop (DT) and Virtual Reality (VR) platforms. This approach responds to the recent call to action for immersive technology research to adopt sentiment analysis as a complementary method for evaluating user experience in immersive systems (Slater et al., 2022). The study design extends the methodology of Khattri et al. (2025), which identified relationships between sentiment and team workload in VR collaboration, by incorporating a comparable Desktop condition for cross-platform analysis alongside detailed conversational analysis, task performance and user sentiment.

Apparatus and materials

Within this work we apply collaboration in a coordinated paired activity where users must cooperate to complete a brainstorming task. The applications Miro [1] and Noda [2] were chosen for the DT and VR conditions. Both applications offer comparable functionality for brainstorming (see Figure 2). In the DT condition, participants used a standard 15-inch windows based computer equipped with a touchpad and keyboard interface. In the VR condition, participants used Meta Quest 2 head-mounted displays with handheld controllers for object selection and placement. Both conditions were run on identical network configurations to eliminate hardware performance bias.

[1] https://miro.com/
[2] https://noda.io/

Participants in both conditions communicated verbally and voice communication occurred through the physical room audio rather than spatialized in-world channels. This setup was chosen to maintain consistent voice transmission between DT and VR participants while ensuring latency and fidelity were the same. Audio and video recordings of the participants were captured on the device being used (i.e., VR or DT) and were transcribed via Microsoft Copilot [3].

Task

In both the DT and VR conditions, participants were instructed to work together to brainstorm advantages and disadvantages of a given topic. In the DT condition users completed the brainstorming using the Miro application [4] on MS Windows and in the VR condition used the Noda App [5] on the meta Quest. Both Miro and Noda support real time collaboration and comparable functionality for brainstorming tasks (see Figures 2b and 2c), thus both participants interacted together in the creation of one shared brainstorm visualization of labeled objects based on the topic.

Four topics were choses for the brainstorming task, namely *the advantages and disadvantages of going to university* and *the advantages and disadvantages of public transport in Birmingham, UK*, *the advantages and disadvantages of mobile phones* and *the advantages and disadvantages of remote working*. Participant pairs had 10 minutes to complete the task with topics randomized across pairs.

Each brainstorming task involved generating and grouping ideas using circular (advantages) and rectangular (disadvantages) nodes (see Figure 2a). The visual differentiation aided in identifying idea categories and relationships during collaborative discussion and later automated content analysis. This structured visual approach for the task aligns with prior research on guided brainstorming, which suggests that providing facilitation and structure can support idea organization in users and reduce their cognitive load during collaborative ideation (Osborn, 1963; Seeber et al., 2017). The use of consistent shapes also ensured uniformity across the VR

[3] https://www.microsoft.com/en-gb/microsoft-365/microsoft-stream
[4] https://miro.com/
[5] https://noda.io/

and DT conditions, facilitating cross-condition comparison. To ensure comparability across the different conditions, no requirement to connect the created shapes was given to ensure used focused on brainstorming and not mind mapping.

Participants were briefed to use the think aloud protocol to vocalise their process of brainstorming and were encouraged to interact with their partner to generate ideas and improve the collaboration. Audio and video of the shared interaction were captured by the device used (i.e., VR or DT). Furthermore, following completion of the task participants engaged in an open discussion about their experience, this was used alongside the task recording for task analysis.

Participants

A within subjects design is applied where users (n = 56) are allocated into pairs (n_{pairs} = 28) (see Figure 1) and completed one of the primary conditions DT (n_{pairs} = 14) and VR (n_{pairs} = 14). To avoid biases or variations within participant groupings, pairs were allocated who had no prior experience of collaborative working together.

Fifty-six participants (n_{pairs} 28) took part in the study (27 male, 27 female, 1 non-binary and 1 preferred not to disclose). Participants were postgraduate and final year undergraduate students recruited from the university's design and computing programmes through email and in class announcements. All participation was voluntary and uncompensated. Inclusion criteria required basic computer literacy and conversational English proficiency. Aiming to represent a sample which considered different cultural perspectives (Stahl & Maznevski, 2021), participants represented a diverse range of nationalities including British (13), Nigerian (11), Indian (10), Pakistani (5), Thai (4), Indonesian (2), Taiwanese (2), Kenyan (2), Malaysian (2), Swedish (1), Bulgarian (1), Greek (1), Iranian (1) and Spanish (1). Regarding VR experience, 37 participants reported to be somewhat familiar with VR, 4 were very familiar and 15 were not at all familiar. Finally experience in group based working was considered, with 27 reporting frequent experience, 28 reporting some experience, and 1 participant indicating they rarely did this.

Protocol

Participants were introduced to the test environment and completed a data consent form and pre test questionnaire which determined their experience in collaborative systems and VR. Binocular visual acuity was assessed using a Snellen chart (Snellen, 1862) and color blindness evaluated using and Ishihara test (Ishihara, 1917) via the online tool EnChroma [6]. Participants wore any prescribed visual correction devices required for the duration of the test.

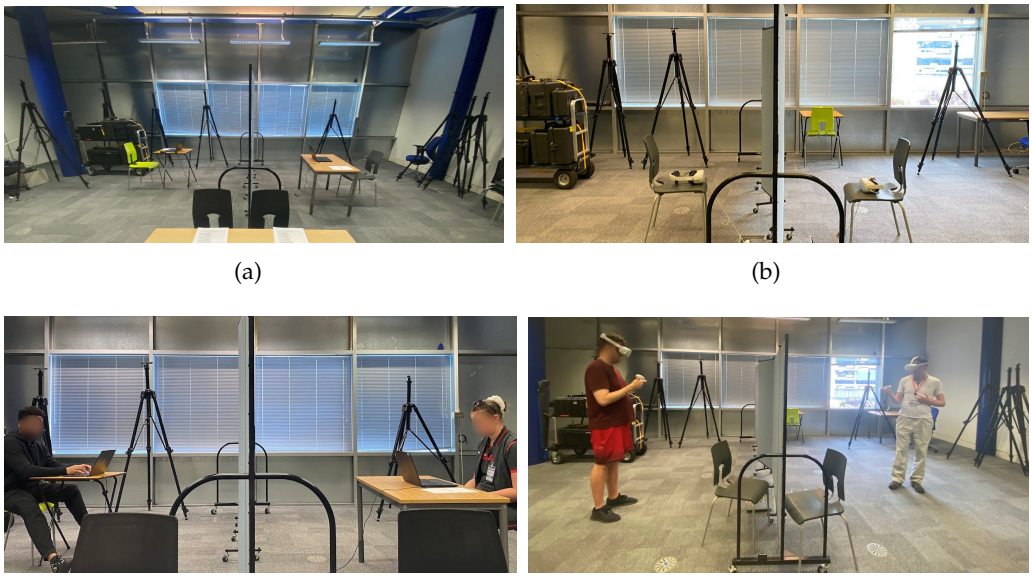
Following the recommendations of Tan, Luyten, Van Den Bergh, and Schöning (2014) and the methodology of Khattri et al. (2025), to support collaboration and establish familiarization in the the participant pairs, prior to the test ice breaking exercise was conducted where participants introduced themselves, discussed their background, study experience and hobbies in a five-minute conversation. This helped build an initial relationship deemed necessary for collaboration.

Prior to the test, and to normalise across experience levels (He et al., 2020), participants underwent a training phase on the system (VR or DT) and the application (Noda or Miro). Once participants had successfully completed the training the test coordinator instructed them on the task.

Environment

All participants were co-located in the same laboratory space and completed the task simultaneously. Each pair was seated (for DT) or stood (VR) in their own space separated by a lightweight divider to prevent direct visual cues (see Figure 1). Participants communicated verbally within the space. The VR and DT conditions were conducted on the same hardware network, ensuring identical latency and communication setup across both modalities.

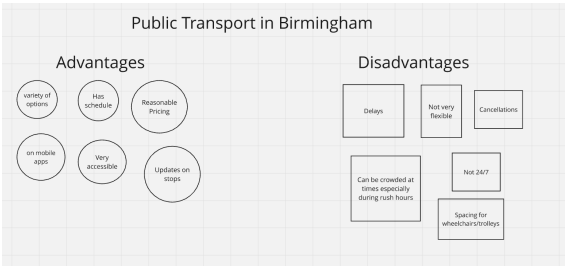
[6] https://enchroma.co.uk/pages/colour-blind-test



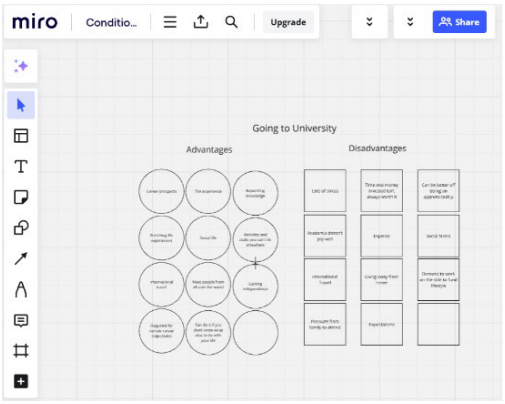
(a) (b)

(c) (d)

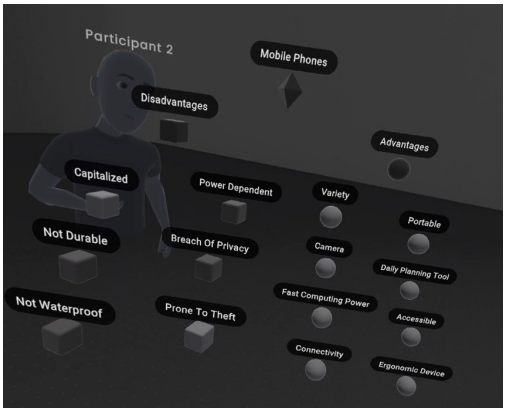
Figure 1: The environment for the study in both the Desktop (DT) (a) and Virtual Reality (VR) (b) conditions. To prevent external visual cues between participants and ensure that they completed the collaborative task on their own devices (DT and VR) pairs were separated by a dividing screen in both the DT (c) and VR (d) conditions.



(a)



(b)



(c)

Figure 2: Example brainstorming template shown to participants before the session (a), including visual guidance on structured shapes (circular, rectangular) for idea grouping. Examples of completed outputs produced collaboratively by participant pairs in the Desktop (b) and VR (c) conditions respectively.

Metrics

Within this work we triangulate results from a variety of performance metrics, notably, task performants, spoken sentiment, post study questionaries and a thematic analysis of user communication.

Task performance

We assess task performance in both the DT and VR conditions (Figure 1) based on objective measures of task completion. Since both the DT and VR conditions had a fixed duration of ten-minutes, we assess task completion based on objective measures of the brainstorming, therefore, the *Total Advantages* (TA), *Total Disadvantages* (TD) and *Total Ideas* (TI).

Sentiment analysis

Sentiment is a personal feeling or opinion from a user and is based on human emotion (Taboada, 2016). Complementary to the task performance we also assess the utterances within the pairs of users. For this we use the methodology of Khattri et al. (2025) who built on the work of Slater et al. (2023). The work of Khattri et al. (2025) however presented a limited evaluation of sentiment analysis which mapped sentiment scores to TWLQ only and there was no evaluation of the sentiment of users presented or comparison of VR against DT. We therefore greatly expand this by providing not only the quantitative measures of task performance but also qualitative thematic analysis on the user dialogue within the DT and VR conditions.

We provide objective measures of the conversation as the *Total Words* (TW) and the sentiment within the conversation. To analyse sentiment in the conversation we apply VADER (Valence Aware Dictionary and sEntiment Reasoner) (Gilbert, 2014) and determine emotional tone (positivity, negativity) and subjectivity in participant pairs conversations. As VADER was originally developed for social media text, its ability to operate effectively on short, informal statements means it application within paired conversations is suitable and aligns to the work of Khattri et al. (2025). While more advanced sentiment processes are possible, notably BERT (Devlin, Change, Lee, & Toutanova, 2019) or RoBERTa (Liu et al., 2019)) which provide increased interpretation of

sentiment they typically require fine-tuning on large, domain-specific datasets and are therefore unfeasible application in this study.

$$PS = \frac{|PW| - |NW|}{(|PW| + |NW|)} + \lambda \qquad (1) \qquad\qquad SS = \frac{(|PW| + |NW|)}{|TW|} + \lambda. \qquad (2)$$

We measure the total number of positive and negative words expressed as the *Polarity Score* (PS) and the *Subjectivity Score* (SS). PS and SS are defined in Equation 2 and Equation 1, where $|PW|$ is the total positive words, $|NW|$ is the total negative words, $|TW|$ is the total overall words and λ is a constant set to 0.000001. PS occupies a scale range between -1 to $+1$, where a -1 indicates complete negativity, a $+1$ indicates absolute positivity and 0 indicates neutrality. SS occupies a range from 0 to 1 indicating subjectivity level.

When calculating sentiment (PS and SS) word dictionaries are used that contain a corpus of both positive words (i.e., wonderful, glad, happy, impressive), negative words (i.e., wrong, sick, impede, broken) and stop words, which represent neutral or pauses (i.e., am, an, I, like, yes) [7]. While these dictionaries are commonly designed for general applications of sentiment, they also contain words common to user performance of systems (i.e., buggy, lag, judder, enjoy, viewable, seamless).

The classical process to apply sentiment analysis is to firstly extract the stop words from the passage of text using the stop word dictionaries. This is completed before the calculation of PS and SS meaning that any words within the stop word dictionaries will not be considered in the sentiment calculation. As stop word dictionaries can contain words which may be valuable in assessing sentiment of digital systems (i.e., DT and VR) we also present PS and SS *without stop word removal* and *with stop word removal*.

Post-study questionnaires

Upon completing the experiment, two post-study questionnaires were given to participants to assess user experience. Owing to the different hardware and software platforms under study, we report the system usability scale (SUS) (Brooke, 1996) as a measure of comparable usability between the DT and VR conditions. SUS is therefore not presented as a formal assessment of a complete system functionality but as a measure of consis-

[7] https://www.kaggle.com/datasets/vivek603/text-analysis?resource=download

tency and to illustrate no usability biases of each system. To measure the collaboration between the user pairs we report on the Team Workload Questionnaire TWLQ (Sellers et al., 2014) for both DT and VR conditions. TWLQ is captured per user as a measure of the user's experience within the collaborative task (Sellers et al., 2014).

Thematic Analysis

We conducted a thematic analysis of user dialogue in both DT and VR conditions to explore participants' qualitative reflections on system use and collaboration. Themes were derived inductively from the data following the approach of Braun and Clarke (2006). Initial open coding was performed on all transcribed dialogue and post-task comments, after which conceptually similar codes were grouped into overarching themes. Through iterative refinement, five primary themes were identified, notably; Usability, Devices, Environment, Collaboration, and User Suggestions (see Table 1).

Each coded segment was also assigned a sentiment label (positive, negative, or neutral) to capture the evaluative tone of participant comments within each theme. This enabled analysis of both the frequency and polarity of user reflections across conditions.

Label Name	Label Description	Example Comment
Usability	Comments regarding specific tools, features and affordances within the applications or user satisfaction (positive or negative) with the application	*"I was wondering if you can copy the square, the circles or squares."*
Devices	Comments regarding the computers and VR headsets, i.e., negative or positives comments participants have with the devices	*"I keep like, I keep adjusting the headsets, you know, like, because, because like, I don't know, it's blurry sometimes, you know, I think my head is just too big."*
Environment	Comments regarding the space they are in which refer to the application as a whole and not to specific tools or features (i.e., boundaries, presence, cyber sickness, infinite canvas, whiteboard, general layout)	*"The space could be more lively and things could be happening, maybe more realistic, maybe like have a pet, have a dog come up to you."*
Collaboration	collaborative conversation between participants interacting with each other to complete the task assigned.	**P1:***"Let me try deleting that one first."*, **P2:***"Do we just hang on?"*, **P1:***"Do I just delete this?"* ...
User Suggestion	Suggestions from users on how the system or application can be improved	*"And also, like what could have helped in that bit was like there could have been like option for like suggested words that we when we're typing, but there wasn't."*
Positive	Comments which show a strong sense of positivity, satisfaction, excitement or when the application, device or feature exceeds expectations.	*"So personally, like I like you, like we've already mentioned, like it was way more easier to type than I expected."*
Negative	Comments which show a strong sense of negativity, frustration or disappointment or when the application, device or feature falls below expectations.	*"But then the problem is it's like we've got to move them around and I can't remember to do that."*
Neutral	Comments which do not show either a strong positive or negative sentiment and without personal opinion.	*"There's delete node in the middle of the keyboard."*

Table 1: Description of thematic analysis labels including sentiment themes (positive, negative and neutral) applied to the user utterances in both the Desktop (DT) and Virtual Reality (VR) conditions.

Results

Task performance

We present task performance as a measure of overall similarity in the task. We present task performance in both DT and VR as measures of the number of ideas generated and the individual total number of advantages and of disadvantages presented in the task.

Number of ideas generated

Performance data for ideas generated was non-normally distributed using the Shapiro-Wilk test (VR, $p = 0.035$), we therefore report the Mann-Whitney U test ($\alpha < 0.05$), with 95% confidence. There was no statical difference regarding number of ideas generated, $U = 72.00, Z = -1.204, p = 0.246$, with a small effect size ($r =$ –0.16) despite the DT having a slightly higher number of ideas ($M = 12.93, SD = 4.20$) than VR ($M = 11.71, SD = 5.17$).

Number of advantages created

Performance data for advantages discussed was normally distributed using the Shapiro-Wilk test (DT, $p = 0.478$; VR, $p = 0.113$), we therefore report the independent samples *T-Test ($\alpha < 0.05$)*, with 95% confidence. There was no statistical difference in the number of advantages generated, $t(26) = 0.565, p = 0.577$ with the DT ($M = 6.93, SD = 2.65$) and VR ($M = 6.36, SD = 2.71$) showing similar numbers of advantages with a small effect size ($d = 0.21$).

Number of disadvantages created:

Performance data for disadvantages generated was non-normally distributed using the Shapiro-Wilk test (VR, $p = 0.015$), we therefore report the Mann-Whitney U test ($\alpha < 0.05$), with 95% confidence. There was no statical difference between in the number of disadvantages created, $U = 71.00, Z = -1.260, p = 0.227$ with the DT ($M = 6.00, SD = 2.00$) and VR ($M = 5.36, SD = 2.82$) showing similar numbers of disadvantages with a small effect size ($r =$ –0.17).

Sentiment analysis

We present the sentiment analysis of both conditions using classical stop word removal and without classical stop word removal. We present, the polarity score (PS), subjectivity score (SS) and the total words (TW).

With stop word removal: Polarity Score (PS) TF1:

PS data were normally distributed across each condition using the Shapiro-Wilk test (DT, $p = 0.461$; VR, $p = 0.726$), we therefore report the independent samples *T-Test* ($\alpha < 0.05$), with 95% confidence. Regarding PS, no statistical difference was observed $t(26) = -1.617, p = 0.118$, with the DT condition illustrating lower positivity ($M = 0.02, SD = 0.18$) than the VR ($M = 0.14, SD = 0.23$) condition. Although the effect size was moderate ($d = 0.58$), the non-significant result indicates insufficient evidence to conclude a reliable difference between conditions.

With stop word removal: Subjectivity Score (SS) TF1:

SS data were non-normally distributed using the Shapiro-Wilk test (DT, $p = 0.043$), we therefore report the Mann-Whitney U test ($\alpha < 0.05$), with 95% confidence. Regarding the SS, no statical difference was found between DT and VR $U = 114.00, Z = -0.735, p = 0.482$ with the DT ($M = 0.08, SD = 0.03$) and VR ($M = 0.08, SD = 0.01$) presenting comparable subjectivity with a small effect size ($r = –0.10$).

With stop word removal: Total Words (TW) TF1:

TW data were normally distributed across each condition using the Shapiro-Wilk test (DT, $p = 0.293$; VR, $p = 0.911$), we therefore report the independent samples *T-Test* ($\alpha < 0.05$), with 95% confidence. There was no statical difference between total words, $t(26) = 0.275, p = 0.785$, although the DT ($M = 433.00, SD = 187.92$) used slightly more words overall than the VR ($M = 414.47, SD = 179.75$) condition and the effect size was small ($d = 0.10$).

Without Stop Word Removal: Polarity Score (PS) TF0:

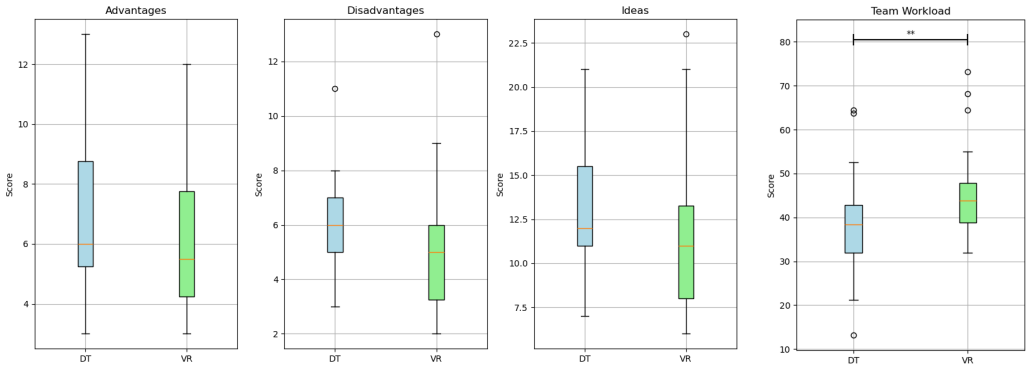
PS data were normally distributed across each condition using the Shapiro-Wilk test (DT, $p = 0.101$; VR, $p = 0.683$), we therefore report the independent samples *T-Test* ($\alpha < 0.05$), with 95% confidence. A statistical difference was observed on PS, $t(26) = -1.58, p = 0.128$, with the DT condition having a lower level of positivity ($M = 0.13, SD = 0.19$) than the VR condition ($M = 0.24, SD = 0.19$). The effect size was moderate ($d = 0.58$), suggesting a potentially meaningful difference in emotional tone between conditions despite the lack of significance.

Without Stop Word Removal: Subjectivity Score (SS) TF0:

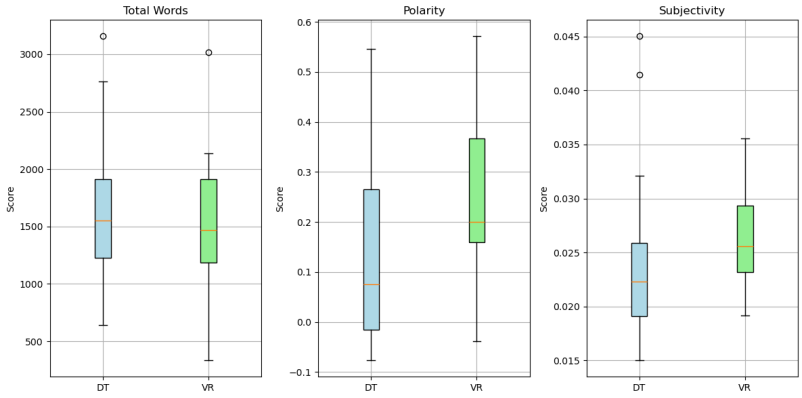
SS data were non-normally distributed using the Shapiro-Wilk test (DT, $p = 0.015$), we therefore report the Mann-Whitney U test ($\alpha < 0.05$), with 95% confidence. Regarding the SS, no statical difference was found between DT and VR, $U = 128.00, Z = 1.378, p = 0.178$ with DT ($M = 0.03, SD = 0.008$) and VR ($M = 0.03, SD = 0.005$) again presenting comparable subjectivity with a small effect size ($r = 0.18$).

Without Stop Word Removal: Total Words (TW) TF0:

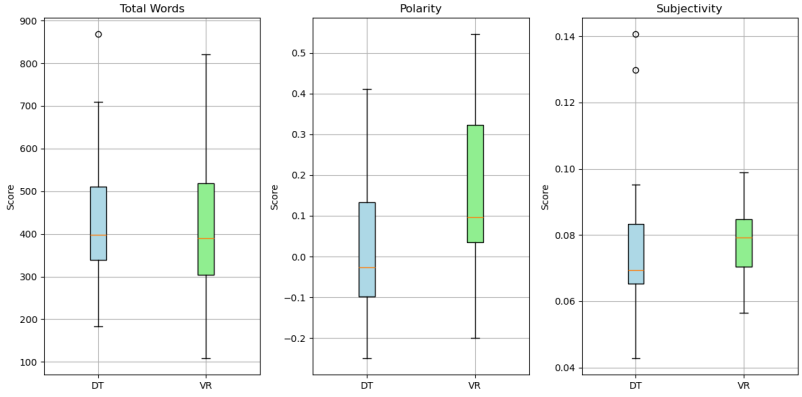
TW data were normally distributed across each condition using the Shapiro-Wilk test (DT, $p = 0.388$; VR, $p = 0.931$), we therefore report the independent samples *T-Test* ($\alpha < 0.05$), with 95% confidence. No statical difference in the TW was observed, $t(26) = 0.433, p = 0.668$, despite the DT ($M = 1619.50, SD = 707.79$) using more words ideas on average than VR ($M = 1504.79, SD = 692.57$) with a small effect size ($d = 0.16$).



(a) (b)



(c)



(d)

Figure 3: Box plots of (a) Task outcomes and (b) Team workload alongside Sentiment (c) without stop word removal and (d) with stop word removal , for each condition Desktop (DT) and Virtual Reality (VR). Significance is indicated with $p < .001$***,$p = .001$**, $p = .05$**

Condition	Measure	Median (IQR)	Mean (SD)	95%CI (LL,UL)	Skewness	Kurtosis
Task Performance						
DT	Ideas	12 (6)	12.93 (4.20)	10.51, 15.35	0.58	-0.14
VR	Ideas	11 (7)	11.71 (5.17)	8.73, 14.70	1.19	0.71
DT	Advantages	6.00 (4)	6.93 (2.65)	5.40, 8.46	0.82	0.73
VR	Advantages	5.50 (5)	6.36 (2.71)	4.79, 7.92	0.89	-0.18
DT	Disadvantages	6.00 (2)	6.00 (2.00)	4.85, 7.15	1.01	1.97
VR	Disadvantages	5.00 (3)	5.36 (2.82)	3.73, 6.98	1.66	3.55
Sentiment - Without Stop Word Removal						
DT	Polarity	0.08 (0.33)	0.13 (0.19)	0.02, 0.24	0.98	0.20
VR	Polarity	0.20 (0.27)	0.24 (0.19)	0.14, 0.35	0.24	-0.71
DT	Subjectivity	0.02 (0.01)	0.03 (0.01)	0.02, 0.03	1.42	1.31
VR	Subjectivity	0.03 (0.01)	0.03 (0.01)	0.02, 0.02	0.37	-0.50
DT	Total Words	1555.00 (796)	1619.50 (707.79)	1210.83, 2028.17	0.79	-0.62
VR	Total Words	1468.00 (913)	1504.79 (692.57)	1104.91, 1904.67	0.352	0.55
Sentiment - With Stop Word Removal						
DT	Polarity	-0.03 (0.30)	0.02 (0.18)	-0.09, 0.12	0.76	0.26
VR	Polarity	0.10 (0.37)	0.14 (0.23)	0.01, 0.27	0.18	-0.86
DT	Subjectivity	0.07 (0.02)	0.08 (0.03)	0.06, 0.09	1.27	1.30
VR	Subjectivity	0.08 (0.02)	0.08 (0.01)	0.07, 0.09	-0.04	-0.75
DT	Total Words	397.50 (198)	433.79 (187.92)	325.28, 542.29	0.91	1.06
VR	Total Words	390.00 (258)	414.64 (179.75)	310.86, 518.43	0.56	0.79
Team Workload						
DT	TWLQ	38.43 (11.56)	38.46 (10.69)	34.31, 43.60	0.41	1.72
VR	TWLQ	43.75 (9.69)	45.50 (9.79)	41.71, 49.30	1.41	1.97

Table 2: Descriptive data of performance and sentiment for each of the conditions Desktop (DT) and Virtual Reality (VR)

Post study questionaries

We present the post study questionnaire results using the TWLQ. We also present SUS as a measure of overall comparable usability of each system.

TWLQ data were non normally distributed using the Shapiro-Wilk test (VR, $p = 0.003$), we therefore report the Mann-Whitney U test ($\alpha < 0.05$). with 95% confidence. Regarding TWLQ, a statistical difference was observed in TWLQ responses, $U = 561.00, Z = 2.77, p = 0.006^{*}$, with the TWLQ having lower team workload in the DT ($M = 38.46, SD = 10.69$) than the VR ($M = 45.51, SD = 9.79$) condition. The effect size was moderate ($r = 0.37$), suggesting that collaborative work in VR may impose a higher subjective workload than on desktop. SUS was

Condition	Label	Total	Positive	Negative	Neutral
DT	Usability	94	22	44	28
VR	Usability	105	53	30	22
DT	Devices	87	19	53	15
VR	Devices	40	9	28	3
DT	Environment	102	55	28	19
VR	Environment	111	55	40	16
DT	Collaboration	154	13	26	115
VR	Collaboration	241	15	26	200
DT	User Suggestion	9	4	2	3
VR	User Suggestion	25	17	4	4
DT	Total	446	113	153	180
VR	Total	522	149	128	245

Table 3: Results of the thematic analysis including sentiment (positive, negative and neutral) themes for the conditions Desktop (DT) and Virtual Reality (VR).

reported to be 82.85 ($SD = 10.68$) in the DT and 71.58 ($SD = 17.63$) in the VR condition with both systems in the excellent category (> 70).

Thematic analysis

Based on the methodology of Braun and Clarke (2006) the initial open coding revealed the five primary themes of Usability, Devices, Environment, Collaboration, and User Suggestions (see Table 1). As discussed prior we also applied a thematic sentiment categorisation to each comment using three categories of positive, negative, or neutral. Therefore a user comment relating to device could be further labeled as either positive, negative or neutral. The distribution of coded comments across conditions is presented in Table 3 and Figure 4. Overall, there were 76 more user comments related to elements of the system and application in the VR condition, despite participants speaking fewer words overall on average (see Table 2), suggesting greater focus on system interaction in VR than on the brainstorming task itself.

Thematic analysis: Usability

There were 94 comments in the usability category for DT and 105 comments for VR. In DT, the majority of the comments (44) were negative. Users reported that they were unable to identify their pair: *"I see the circle, but I don't see you...but I can't see the cursor"* (Pair 3, DT). In addition, some users found it difficult to create, select and manipulate shapes. For example, *"there's no option to align stuff"* (Pair 5 DT) and *"(I keep) not dropping the shape in the right place"* (Pair 6, DT). The positive comments in DT (22) show that users appreciate that you can use Miro to work collaboratively: *"we can see each other working at the same time"* (Pair 7, DT). A few pairs also note that they find Miro easy to use because they are familiar with other similar design software. For example, *"I think it's quite easy to use because I'm more familiar with the UI"* (Pair 11, DT) and *"I think if you're also familiar with these (kinds of) software, there's a lot of shortcuts..."* (Pair 13, DT).

For VR, the majority of comments (53) in this category were positive. Participants say that learning to use Noda was *"interactive and intuitive"* (Pair 3, VR) and most of them *"picked up (creating shapes and typing) easily"* (Pair 13, VR). The negative comments regarding VR in this category (30) were mostly complaints regarding the typing experience. For instance, *"I guess the backspace key is very weird. It's like a backwards play button rather than the standard one"* (Pair 7, VR) and *"So I would get annoyed maybe because you have to point to each and every letter"* (Pair 12, VR). Pair 6, (VR) also described feeling disorientated as *"all the dots are everywhere, like the squares and the circles"*.

Thematic analysis: Devices

There were 87 comments in the devices category for DT and 40 for VR. Most of the comments for DT in this category were negative (53). Participants generally disliked having to use a computer trackpad and said that they would have preferred a mouse: *"I don't like using trackpads at all"* (Pair 2, DT) and *"I like a mouse, not a mouse pad"* (Pair 5, DT). Others had complaints about the keyboards the layout was different to what they were used to on their personal devices : *"It's all mapped wrong...some of the letters aren't matching properly"* (Pair 9, DT) and *"although I did get used to the keyboard in the end, I still never could understand how to copy and paste, which is what I was looking for..."* (Pair 14, DT). In contrast, the positive comments in this category (19) referred to the fact that participants found the laptops easy to use because they are familiar with these devices. For example, *"I've been us-*

ing laptops so it's easy to use" (Pair 3, DT) and *"I'm more used to using a mouse, but I can use a trackpad. It wasn't as bad as I thought it would be"* (Pair 13, DT).

Most of the comments in the devices category are also negative for VR (28). Some participants criticised the weight of the headset: *"the headset is very tiring"* (Pair 4, VR) and *"Like for small projects it's okay to work with this, but if you have a longer time period you have to work with then it would be uncomfortable"* (Pair 8, VR). Similarly to DT, users also expressed negative sentiment towards the typing experience in VR. For instance, *"I wasn't the biggest fan of the typing though"* (Pair 9, VR) and *"I can't find any symbols on the keyboard...and the option to go to the next paragraph... those two things I could not find"* (Pair 9, VR). In contrast, there are only 9 positive comments for VR devices. There was only one positive comment regarding the headset: *"I like the headset, pretty comfortable"* (Pair 7, VR). Additionally, the same participant pair found that *"it was pretty self-explanatory, pretty easy controls like the little handheld thing"*. There were four comments which stated that typing in the VR headset was easier than they expected. Some examples of this include, *"...I expected the typing to be worse actually"* (Pair 7, VR) and *"I also think the typing was surprisingly easy because I thought that this would be, I mean, I honestly thought that this would take me longer to type over here as opposed to my laptop or my phone"* (Pair 9, VR).

Thematic analysis: Environment

There were 102 comments in the environment category for DT and 111 for VR. The majority of comments in DT were positive (55). Participants mentioned that Miro was *really straightforward* (Pair 1 DT) and that *there was no learning curve* (Pair 11 DT). This is because they were *"used to applications like Figma and other design applications"* (Pair 3, DT). There were 28 negative comments in this category for DT. While users acknowledge that Miro was useful for simple tasks, they explained that the software does not have the adequate functionality needed to do more complex design tasks. For example, *"Like for the like the quick and dirty design where it's not, it's not intended to be like design perfect, like the fact that it's not easy to align stuff"* (Pair 5, DT) and *"I've used the FigJam app before, I would say it's more convenient or more easy to use"* (Pair 9, DT). Some participants also commented on the fact that they would have preferred to see their pair visually represented throughout the collaboration process. For instance, *"you don't really have a good view of what other people are doing"* (Pair 2, DT) and *"for me to work it out, I need*

to see someone's face so that if they've closed their mouth and has stopped talking for a few seconds, I know I can speak" (Pair 13, DT).

Most of the comments for the VR environment were also positive (55). Participants generally enjoyed the infinite space and the fact that they had creative freedom to do whatever they liked: *"I can like drag and drop things and I feel like it's my own space"* (Pair 12, VR) and *"Okay, so basically it's very good kind of thing like where we can do meeting and we can go creative and plus we have like lot of things we can do over here"* (Pair 14, VR). There were 40 negative comments for the VR environment. Some participants thought that the *"virtual space could be a bit livelier"* (Pair 5, VR) rather than being in a *"big blank warehouse looking area"* (Pair 2, VR). A few others also stated that the space became more cluttered throughout the task. For example, *"Because everything is in the space, it's not well arranged...the advantages look like everything is just going through each other"* (Pair 1, VR) and *"I'm spending more time getting distracted by all the shapes"* (Pair 7, VR).

Thematic analysis: Collaboration

There were 154 comments in the collaboration category for DT and 241 for VR representing the highest number of comments in any category. For both conditions, this was the category with the most comments and the majority were neutral (DT = 115 and VR = 200). For DT, participants typically discussed how they wanted to carry out the task and how to use Miro's functionalities and the keyboard. *"Duplicate it and then change it. Change the shape that it is"* (P1, Pair 2, DT), *"Oh, wait. So, like highlight it and then"* (P2, Pair 2, DT), *"Highlighted and control D, I think"* (P1, Pair 2, DT).

There was a similar trend in the VR condition. The comments are generally of participants conversing with each other to assist their pair in using Noda's features. *"do you know how to delete this note?"* (P1, Pair 8, VR), *"Can you do it?"* (P2, Pair 8, VR), *"Oh yeah, I got it now"* (P1, Pair 8, VR).

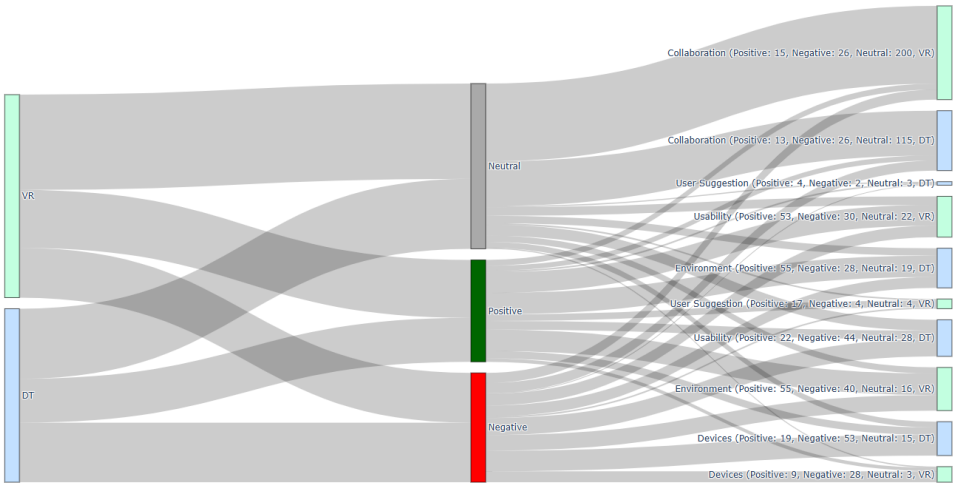


Figure 4: A Sanky diagram of the themes extracted via the thematic analysis for the conditions Desktop (DT) and Virtual Reality (VR).

Thematic analysis: User suggestions

There were 9 comments for user suggestion for DT and 25 comments for VR. This category contained the least number of comments overall. The positive user suggestions for DT (4) address improvements that could be made to the laptops' functionalities. For example, they suggested that having a *"touch screen...could also work quite well"* (Pair 5, DT). The negative comments (2) arise after users experiencing problems with aligning the shapes in Miro: *"It'd be much better if it had snapping to grid"* (Pair 5, DT).

There were more positive user suggestions for VR (17) than DT. Participants tended to express their excitement for using Noda and suggested features that could improve their user experience. They suggested that it *"will be fun...(if) you can customise the environment"* (Pair 3, VR) and that having *"voice recognition would make it faster to enter the text"* (Pair 7, VR). One pair also suggested other use cases for Noda, such as *"information modeling"* (Pair 7, VR) and *"visualizing how a house would look"* (Pair 7, VR).

Discussion

The aim of the paper was to evaluate collaborative brainstorming in desktop (DT) and Virtual Reality (VR) environments and via triangulation of multiple experience metrics investigate the research question: *To what extent does VR enable collaboration comparable to that in traditional DT environments during brainstorming tasks?*

Desktop and Virtual Reality present comparable performance

While the input modalities and software differed between the DT and VR conditions, the interface and interaction design of the task was suitably matched for functionally. No significant statistical differences were found in task performance, with both the DT and VR conditions resulting in comparable numbers of ideas generated within the time given.

Additionally, no significant differences were found in the sentiment of user conversation, notably in the number of words uttered between pairs and the polarity or subjectivity of their discussion. This illustrates that DT and VR can offer comparable performance regarding personal spoken sentiment and task completion for collaborative brainstorming tasks. Additionally while the team workload (TWLQ) illustrated a statistical difference with VR resulting in a higher team workload, the overall thematic analysis presented comparable data trends for both the DT and VR conditions. This mean that while users may perceive their individual workload to be higher in VR condition, their vocalisation, verbalisation and task success illustrated comparability in the use of DT and VR systems.

Analysis of the conversations between user pairs also highlighted how participants were more interested in the VR condition and while this may be a result of the novelty effect with VR, it may also be a result of increased social presence in this condition. Even though both systems simulated a remote collaboration and users were visually separated in space from their user, the VR condition incorporated virtual avatars. As studies have shown that increased social presence, or the use of avatars, is associated with a variety of positive communication outcomes (Fogg & Tseng, 1999) (Fox et al., 2015), this could be attributable to the increased communication and interest displayed by users for the VR condition.

No significant difference was observed in the number of ideas created, however, DT was on average slightly higher than VR ($M = 12.93, M = 11.71$). The number of disadvantages and advantages created by user pairs in the task were also comparable, with DT resulting in $M = 6.93, M = 6.00$, and VR resulting in $M = 6.36, M = 5.36$. Although both conditions compared similarly regarding the number of ideas, there was an observed tendency for some user pairs to attempt to link up the ideas (via interconnecting lines) and differentiate between the advantages and disadvantages (using colors) in the VR condition. This variation in the task approach may illustrate an increased engagement in the VR condition, which aligns to prior work highlighted how individuals using VR commonly exhibit higher levels of engagement in tasks than they do in a 2D (or DT) condition (Pizzolante, Bartolotta, Sarcinella, Chirico, & Gaggioli, 2024). This finding may also be related to a potential increased spatial awareness in the VR condition and the presentation of a 3D representation of the task alongside a 3D UI may increase their engagement (Pizzolante et al., 2024). However users did vocalise more technical challenges in the VR condition (i.e., difficulties in typing, virtual avatar management, and headset weight), although the reported comparability in sentiment scores of polarity and subjectivity implies that both systems may be appropriate for collaborative brainstorming.

Despite the VR condition providing greater environmental richness and potentially increased user (and co-worker) presence, the results demonstration only small descriptive differences in collaborative performance between the conditions. This suggests that for collaborative tasks such as brainstorming which rely on a formal structure and mutual communication, the added immersion in reality may not provide measurable performance gains.

This aligns to work by Slater (2009) who argues that presence alone is not a proxy for effectiveness and that task design and correct use of technology is essential.

Similarity in spoken sentiment is observed between Desktop and Virtual Reality

Recent work by Slater et al. (2022) reports on the *"Problem of Measurement"* highlighting that questionnaires are simple to use but do not encapsulate the whole user experience. Slater et al. (2022) states that *"ideal measurements should involve this type of triangulation between several completely different approaches, such as a combination of*

configuration transitions and qualitative approaches, including sentiment analysis". Therefore following the methodology of Khattri et al. (2025) and building on the recommendations of Slater et al. (2023), sentiment analysis was applied within the task to support triangulation results for the DT and VR conditions against the task performance and the post task questionaries.

The results illustrate no significant difference in the sentiment polarity (PS), subjectivity (SS) or total number of words uttered in DT and VR. However, less words were uttered in the VR condition and this partially supports the idea that VR has the capacity to fully immerse users (Joo et al., 2024). Participants may communicate less with their pairs in VR because they are more focused on exploring the virtual world that they have been placed in (Balcerak Jackson & Balcerak Jackson, 2024) or extracting information on how their physical body relates and interacts with their identity in the new environment. While higher immersion often leads to higher levels of motivation, curiosity and reflective thinking (Guerra-Tamez, 2023), these results could also highlight certain disadvantages of using VR technology in collaboration. Specifically, that participants may be speaking to each other slightly less because their attention is directed to identifying their pair in the VRCS (Roberts et al., 2009) and solving technical issues with their device and avatar (Dickson et al., 1996). While this is evident in some user comments regarding the VR devices, for example, *"Oh, no, I can't see you, but I can see you doing it"*, or *"I'm spending more time getting distracted by all the shapes."* the VR condition performed comparably to the DT condition in the PS, SS and total words scores. The observed comparisons in team sentiment may have been influenced by the use of open room audio for partner communication rather than spatialized device centric sound. Therefore, future research should explore the impact of device centric spatial audio cues on team sentiment, communication and co presence.

Commonalties in paired user conversations

Aligned to the comparable performance and the sentiment similarity between the DT and VR conditions, commonalities were also uncovered within the conversation detail between user pairs. Firstly, there was a higher number of common negative words than positive words in both conditions. This finding does not necessarily

mean that both conditions contained more negative than positive words overall and could simply imply that more participants shared similar opinions on the negative aspects of both the VR and DT conditions.

For both conditions, it is likely that the words *advantage* and *disadvantage* and their plural forms had the highest frequency, this owing to these words being central in the brainstorming task. Additionally, it is important to consider that some of the other words may also describe the task elements rather than the system of application (i.e., positive or negative utterances within the specific task). For this reason, within specific team based tasks like brainstorming, these common words should be examined in the context of how they appear in the user comments regarding the system and PS and SS not be used in complete isolation.

Apart from *"advantages"*, the positive words that both conditions had in common included*"easy"* and *"good"*. In the DT condition, participants commented that the platform was a *"good platform"* and a *"good tool for what we are trying to do"*, which was *"good for collaboration"* especially in large groups. In the VR condition, individuals said that the application was *"good"* for brainstorming and whiteboarding in meeting settings. One pair said, *"I thought it was actually better than I expected"* followed with *"yeah, I thought it was pretty good"*. This shows that some people may have had negative preconceptions of VR prior to this study and may have assumed collaboration and creative exploration would be limited in VR. In addition, they may have previously thought that the VR condition was difficult to use especially if they do not have the required technical experience (Baniasadi, Ayyoubzadeh, & Mohammadzadeh, 2020). These opinions were in some way demystified as participants commented that the VR system was *"pretty self-explanatory"* with *"pretty easy controls"*. In some cases, the words *"easy"* and *"good"* appeared in negative contexts. For example, *"it's not easy to align stuff"* (DT) or *"you don't really have a good view of what other people are doing"* (DT). This could suggest that the positive sentiment ($PS = 0.2$) may be skewed by this form of comment and more comprehensive sentiment analysis using valence shifting could be considered in future work (Slater et al., 2023).

Variations of the word *"work"* or *"worked"* also appeared frequently in both the DT and VR conditions. Commonly this was used in a general context to describe the way the systems work or that it could be used in work tasks. In VR, however, the use of *"work"* is seen more commonly as a confirmation of success in the task. For

example, *"Does it work now?"*. This could imply that in VR, more discussion between the pairs is needed to confirm that the system is working for the collaborative task.

Lastly, words which appeared only in VR included *"fun"* and *"interesting"*. Participants may have conveyed a higher level of excitement for VR due to the novelty of VR environments (Taufik, Kunz, & Onwezen, 2021). The fact that the word *"cool"* appeared more in VR than DT supports this, and may illustrate that most participants did not experience the same level of novelty with the desktop condition.

In terms of common negative words, the words with the highest frequency in both conditions were *"hard"* and *"difficult"*. In VR, participants used these words to comment on the usability of the VR system as well as typing or moving objects with controllers (Foy, Dudley, Gupta, Benko, & Kristensson, 2021). In contrast, *"hard"* and *"difficult"* were used more in DT within a context which conveys simplicity of the system. For example, *"it's not too difficult for me to get used to it"*. However, there are some exceptions to this. Examples include, *"it is really quite difficult to manoeuvre"* and *"changing the size of the font is hard"*.

The word *"Problem"* occurred in DT and VR conditions. In both conditions, it is generally used to convey limitations of the systems or reasons why they have expressed negativity. For example, *"that's a problem"* (DT) and *"...only problem is probably can't wear this for a long period of time"* (VR). Furthermore, variations of the word *"limitation"* appeared in DT and VR. Common limitations of mentioned in both conditions focused on hardware (i.e., battery life, cost, input) rather than the environment or application.

Overall, the analysis of the most common positive words showed that the users felt both applications were appropriate to use for collaborative brainstorming task. In terms of the most common negative words, participants focused on hardware considerations in both conditions.

Higher team workload was found in Virtual Reality

Post-study questionnaire results revealed a significant difference in team workload (TWLQ) between the DT and VR conditions, with VR participants reporting higher perceived workload on average. This suggests that participants considered the DT condition easier and experienced fewer challenges in collaboration than those

working in VR. For example *"I've been using laptops so it's easy to use."* and *"it's not difficult, but it's like it's not exactly as it's as smooth"*. Although sentiment scores and overall task performance were comparable across both conditions, the increased workload reported in VR indicates potential usability and coordination challenges unique to immersive collaboration.

Participants in the VR condition often commented on issues related to navigation, input mechanics, and system usability, which may have contributed to their increased workload perception. For example, *"oh my God, it's hard to type in"* and *"when you're trying to, like (read the ideas that we put down), it's quite hard"*. While this does not render VR unsuitable for collaborative tasks compared with DT platforms, it highlights that users may benefit from additional familiarization tasks or further training before engaging in complex teamwork scenarios (He et al., 2020).

It is also important to note that the brainstorming task employed in this study used structured visual elements, prioritising analytic consistency over open-ended creativity. Consequently, greater variability in workload perceptions might emerge in less structured, more creative tasks where interaction demands and cognitive load differ more substantially between DT and VR environments.

Communication themes reveal subtle differences between Desktop and Virtual Reality

The thematic analysis of paired user conversations revealed that more comments were found in the usability category for the VR condition. Most of these comments were positive for VR while most were negative for DT.

Neutral usability comments commonly consisted of participants describing what they were going to do or have done. In some cases, they also sought confirmation from their pair that what they have done is visible. Examples of this include, *"Can you see my mouse? Can you see my point(er) already?"* (DT) and *"I was wondering if you can copy the square, the circles"* (VR). Additionally, given that users vocalised difficulties in typing in the VR condition and using the controllers, could imply that the VR system is less usable than the desktop in this aspect. This finding aligns to the conferable body of work exploring alternatives to typing in VR (Wang, Yan, Zhang, Liu, & Wang, 2023) and alternative input paradigms (Wan et al., 2024). Therefore, it is recognised that further research is needed to improve the usability text entry in VR.

The negative comments regarding devices in VR also confirmed many of the disadvantages of VR. Participants sometimes struggled to identify their pair due to avatar and device management issues (Dickson et al., 1996). The thematic analysis also revealed more specific issues regarding the weight of the headset, typing and blurriness which in reality could increase cognitive load (Parmar et al., 2016) (Makransky et al., 2019). The majority of comments relating to collaboration were neutral, commonly focusing on task leadership or supportive comments task (Laal, 2013) (Thompson et al., 2021) and did not highlight specific challenges to collaborating. This finding was also evidenced with the limited comments relating to improvements and suggestions for both conditions. In the VR condition participants stated that they were impressed with the features and suggested ways of improving them further, notably relating to typing and avatar visibility.

Triangulation of multiple measures is recommended for collaborative performance assessment

A recent critique of VR research concerns the over reliance on post-experiment questionnaires rather than methods that capture participants' actual experiences (Paulsen et al., 2024). In this study, post-study questionnaires were used to assess user impressions and perceived workload but were complemented by sentiment analysis of participants' dialogue during the collaborative task, extending the approach proposed by Khattri et al. (2025). This triangulated methodology of combining objective performance metrics, sentiment analysis, and thematic evaluation, is novel to this work and provides a richer understanding of collaborative brainstorming in VR and DT conditions.

While performance results alone suggested similar outcomes across conditions, post-study workload ratings indicated that collaboration in VR was perceived as more demanding. Moreover, sentiment analysis revealed comparable positivity (PS) and subjectivity (SS) between conditions, yet differences emerged in the total words spoken and in the frequency of negative comments identified through thematic coding. These findings reinforce the value of incorporating sentiment and user dialogue based measures alongside conventional task metrics and questionnaires to capture richer aspects of user experience.

Although more advanced sentiment analysis models could yield finer grained insights, specifically those that include valence shifting (Slater et al., 2023) the present study demonstrates the feasibility and potential of integrating sentiment analysis within evaluation frameworks of user experience. However, depending on the task, research should consider how task related dialogue can overlap with experiential dialogue, i.e., when participants alternated between discussing the task and commenting on the system. In this case sentiment scores may partially reflect both task content and user attitude and therefore highlights an important avenue for refinement in future research.

Limitations

While we consider this study to be a success and believe the results offer valuable insights to illustrate the comparability of VR and DT for collaborative brainstorming tasks, we acknowledge some limitations and future avenues for research.

Firstly, in this paper, we use the term collaboration to describe coordinated group activity toward a shared goal. However, given the structure of the brainstorming task, participant interaction primarily reflected cooperative behavior, characterised by parallel idea generation with intermittent coordination rather than deeply interdependent collaboration. While the study examined teamwork and communication, the task itself focused on idea generation rather than joint problem-solving or decision-making. As such, the depth of collaboration observed may have been limited to cooperative interaction rather than fully interdependent teamwork (Salas, Cooke, & Rosen, 2008). Additionally, the open ended nature of the brainstorming task used may have introduced greater variability for task success, therefore leading to less pronounced differences between conditions, more heterogenous responses from participants, and wider confidence intervals (Table 2). Future research should therefore employ more structured design or decision based tasks that require a cooperation in strategy and negotiation, which may reveal deeper insights into the potential differences DT and VR offer to collaborative processes. Additionally, as participants were co-located in the same space but physically separated with a screen rather than geographically remote, communication dynamics may have been influenced by shared physical context and

this may not represent remote working or remote device use. Future work should explore remote VR collaboration to explore distributed team scenarios.

This study also demonstrated the feasibility of incorporating sentiment analysis into evaluations of VR, thus building on work of Slater et al. (2023) and Khattri et al. (2025). However, as the VADER (Gilbert, 2014) method implemented in this study does not consider valance shifting, meaning that statements like *"this is not very good"* or *"this is not bad"* will return incorrect polarity, more sophisticated methods, such as BERT (Devlin et al., 2019) or RoBERTa (Liu et al., 2019), could be employed in future. Additionally, future work could move beyond individual utterance sentiment to analyze team based interpersonal dynamics such as, roles, leadership, turn-taking, distribution of conversation or shared emotional similarity. Additionally visualization methods for presenting real-time sentiment feedback in collaborative systems could also enhance team awareness and potentially improve emotional communication within teams of mixed abilities or backgrounds. Additionally, terms of participant recruitment, we aimed to recruit a large population sample ($N = 56, Npairs = 28$) with considerable diversity in background, however we acknowledge the sample comprised primarily of students. Therefore future work should include professional and multidisciplinary teams to assess generalisability. Finally, we would welcome replication of this study across different task types which offer unique challenges not found in brainstorming (i.e., assembly tasks or creative design). This wider exploration would help create a more generalised set of findings to inform the design of immersive collaborative systems.

Conclusions

The aim of this study was to examine user experience when performing collaborative brainstorming tasks in a traditional computer-based environment (DT) and Virtual Reality environment (VR). User pairs ($Npairs = 28$) were tasked to collaboratively brainstorm ideas for a chosen topic for ten-minutes. To overcome the reported problems with user experience assessment, notably the over reliance on post task questionnaires for VR, we presented an in depth triangulation of quantitative and qualitative performance, sentiment analysis and team workload. Task completion accuracy was measured using the number of ideas generated within the brainstorming captured. Additionally, participant conversations were recorded during the collaboration and sentiment

analysis via the VADER method was applied to determine the degree of positivity and negativity in the experience. We illustrate the level of commonality in both the task performance and the user sentiment between conditions. Lastly, to identify the general trend of user opinion in both conditions, a thematic analysis was carried out structured against the labels: usability, devices, environment, collaboration and user suggestion and user team workload assessed using the Team Workload Questionnaire (TWLQ).

The results of both the sentiment analysis and task performance illustrated no significant difference between collaborative experience and task performance when completing brainstorming in a VR or DT environment. This illustrates comparability in paired working across both platforms and therefore offers insights to the potential use of VR for collaborative team-working. Additionally, the thematic analysis revealed that participants in the VR condition were more inclined to experiment with tools presented to them in the virtual environment. However, participants did report on more technical considerations for VR technology, notably relating to typing, avatar management and headset weight. This suggests that whilst both systems VR and DT systems are suitable for collaborative brainstorming, there may be more technical considerations which still require further research for VR. Results of the Team Workload Questionnaire (TWLQ) confirmed this outcome with a statistical difference being observed between the VR and DT conditions and a higher workload reported for VR.

This work demonstrates that collaborative brainstorming in VR can achieve outcomes comparable to desktop (DT) settings across multiple measures of performance, communication sentiment, and perceived workload. The triangulated approach highlights the value of combining behavioral, linguistic, and experiential data in evaluating collaborative systems. While VR introduced higher team workload, the overall collaboration patterns remained consistent, suggesting its viability for creative teamwork rather than superiority over desktop tools.

By presenting a triangulation between multiple measures of collaborative task completion our results illustrate that quantitative data cannot be analyzed in isolation and the nuances in user sentiment and opinion should be considered for a more in depth insight of their experiences. We further recommend that future studies should consider this triangulation of objective measures alongside in experience sentiment and post tasks analysis to evaluate VR and collaborative computing systems in future.

References

Amabile, T. (1997). Motivating creativity in organizations: On doing what you love and loving what you do. *California Management Review*, *40*(1), 39-58. doi: 10.2307/41165921

Balcerak Jackson, M., & Balcerak Jackson, B. (2024). Immersive experience and virtual reality. *Philosophy and Technology*, *37*(19). doi: 10.1007/s13347-024-00707-1

Baniasadi, T., Ayyoubzadeh, S. M., & Mohammadzadeh, N. (2020). Challenges and practical considerations in applying virtual reality in medical education and treatment. *Oman Medical Journal*, *35*(3). doi: 10.5001/omj.2020.43

Bordoloi, M., & Biswas, S. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. *Artificial Intelligence Review*, *56*, 12505-12560. doi: 10.1007/s10462-023-10442-2

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. doi: 10.1191/1478088706qp063oa

Brendel, D., & Williams, R. (2013). Methods of effective information exchange in the virtual world. In *2013 international conference on collaboration technologies and systems (cts)* (p. 553-560). IEEE. doi: 10.1109/CTS.2013.6567286

Brooke, J. (1996). Sus - a quick and dirty usability scale. In B. A. W. P. W. Jordan B. Thomas & I. L. McClelland (Eds.), *Usability evaluation in industry* (p. 189–194). Taylor Francis.

Brownell, J., & Chung, B. G. (2001). The management development program: A competency-based model for preparing hospitality leaders. *Journal of Management Education*, *25*(2), 124-145. doi: 10.1177/10525629010250020

Caserman, P., Garcia-Agundez, A., Gámez Zerban, A., & Göbel, S. (2021). Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. *Virtual Reality*, *25*(4), 1153 - 1170. doi: 10.1007/s10055-021-00513-6

Chen, D., & Li, W. (2021). Analyzing college students' social cognitive construction and emotion in different types of online synchronous collaborations of a creative task. In *9th international conference on information and education technology (iciet)* (p. 311-325). IEEE. doi: 10.1109/ICIET51873.2021.9419637

Cummings, J., & Bailenson, J. (2016). How immersive is enough? a meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, *19*(2), 272–309. doi: 10.1080/15213269.2015.1015740

Dennis, A. R., & Wixom, B. H. (2002). Investigating the moderators of the group support systems use with meta-analysis. *Journal of Management Information Systems*, *18*(3), 235-257. doi: 10.1080/07421222.2002.11045696

Devlin, J., Change, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the north American chapter of the association for computational linguistics: Human language technologies, volume 1 (long and short papers)* (p. 4171-4186). Association

for Computational Linguistics. doi: 10.18653/v1/N19-1423

Dickson, G., Lee-Partridge, J., Limayem, M., & Desanctis, G. (1996). Facilitating computer-supported meetings: A cumulative analysis in a multiple-criteria task environment. *Group Decision and Negotiation*, *5*, 51-72. doi: 10.1007/BF02404176

Doroudian, S. (2023). Collaboration in immersive environments: Challenges and solutions. *ArXiv*. doi: 10.48550/arXiv.2311.00689

Dunmoye, I., May, D., & Hunsu, N. (2022). An exploratory study of social presence in a collaborative desktop virtual reality (vr) land surveying task. In *Ieee frontiers in education conference (fie)* (p. 1-5). IEEE. doi: 10.1109/FIE56618.2022.9962641

Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). Human emotion recognition: Review of sensors and methods. *Sensors*, *20*(3). doi: 10.3390/s20030592

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. *Psychological Bulletin*, *101*(1), 91-119. doi: 10.1037/0033-2909.101.1.91

Fjermestad, J., & Hiltz, S. R. (1998). An assessment of group support systems experimental research: Methodology and results. *Journal of Management Information Systems*, *15*(3), 7-149. doi: 10.1080/07421222.1998.11518216

Fogg, B. J., & Tseng, H. (1999). The elements of computer credibility. In *Chi '99: Proceedings of the sigchi conference on human factors in computing systems* (p. 80-87). Association for Computing Machinery. doi: 10.1145/302979.303001

Fox, J., Ahn, S. J., Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N. (2015). Avatars versus agents: A meta-analysis quantifying the effect of agency on social influence. *Human Computer Interaction*, *30*(5), 401-432. doi: 10.1080/07370024.2014.921494

Foy, C., Dudley, J., Gupta, A., Benko, H., & Kristensson, P. (2021). Evaluating collaborative ideation in immersive systems. In *Chi '21: Proceedings of the 2021 chi conference on human factors in computing systems* (p. 1-11). ACM. doi: 10.1145/3411764.3445671

Galdámez, N., Sanz, C., & Armando De Giusti, C. (2012). Storm: Brainstorming collaborative web environment oriented to the educational context. In *2012 international conference on collaboration technologies and systems (cts)* (p. 275-279). IEEE. doi: 10.1109/CTS.2012.6261061

Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*, *8*(1), 216-225. doi: 10.1609/icwsm.v8i1.14550

Guerra-Tamez, C. (2023). The impact of immersion through virtual reality in the learning experiences of art and

design students: The mediating effect of the flow experience. *Education Sciences*, *13*(2). doi: 10.3390/educsci13020185

Hamad, A., & Jia, B. (2022). How virtual reality technology has changed our lives: An overview of the current and potential applications and limitations. *International Journal of Environmental Research and Public Health*, *19*(18). doi: 10.3390/ijerph191811278

He, Z., Du, R., & Perlin, R. (2020). Collabovr: A reconfigurable framework for creative collaboration in virtual reality. In *2020 ieee international symposium on mixed and augmented reality (ismar).* IEEE. doi: 10.1109/ISMAR50242.2020.00082

Hove, D., & Watson, B. (2022). The shortcomings of video conferencing technology, methods for revealing them, and emerging xr solutions. *PRESENCE: Teleoperators and Virtual Environments*, *31*, 283-305. doi: 10.1162/pres$_{a0}$0398

Ishihara, S. (1917). *Tests for colour-blindness*. Tokyo: Hongo Harukicho.

Jing, A., Frederick, M., Sewell, M., Karlson, A., Simpson, B., & Smith, M. (2023). How visualising emotions affects interpersonal trust and task collaboration in a shared virtual space. In *2023 ieee international symposium on mixed and augmented reality (ismar)* (p. 849-858). IEEE. doi: 10.1109/ISMAR59233.2023.00100

Joo, J., Han, S., Park, I., & Chung, T. (2024). Immersive emotion analysis in vr environments: A sensor-based approach to prevent distortion. *Electronics*, *13*(8). doi: 10.3390/electronics13081494

Kaiser, K., Walters, S., Sheehy, K., Murray, E., & Spencer, K. (2024). Implementing the technology shift from 2d to 3d: Insights and suggestions for umpire educators. *Frontiers in Virtual reality*, *5*. doi: 10.3389/frvir.2024.1368648

Karl, K., Peluchette, J., & Aghakhani, N. (2022). Virtual work meetings during the covid-19 pandemic: The good, bad, and ugly. *Small Group Research*, *53*(3), 343-365. doi: 10.1177/10464964211015286

Kettler, T., Lamb, K., & Mullet, D. (2019). *Developing creativity in the classroom: Learning and innovation for 21st-century schools*. New York: Routledge. doi: DOI:10.4324/9781003234104

Khattri, B., Chantratita, P., Frutos-Pascual, M., & Williams, I. (2025). I've got a feeling: Sentiment analysis for collaborative performance in vr. In *2025 ieee conference on virtual reality and 3d user interfaces abstracts and workshops (vrw)* (p. 1386-1387). IEEE. doi: 10.1109/VRW66409.2025.00339

Kim, Y., Won, J., Jang, S., & Ko, J. (2022). Effects of cybersickness caused by head-mounted display-based virtual reality on physiological responses: Cross-sectional study. *JMIR Serious Games*, *10*(4). doi: 10.2196/37938

Kowalkiewicz, M. (2012). Ideawall: Bridging the digital and non-digital worlds to facilitate distant collaboration. In *2012 6th ieee international conference on digital ecosystems and technologies (dest)* (p. 1-5). IEEE. doi: 10.1109/DEST.2012.6227942

Kozachek, D., Yaqin, D., Prasad, K., & Wang, S.-M. (2023). Evaluating the outcome of collaborative vr mind mapping

sessions with sentiment analysis and emotional intelligence. In *Human-computer interaction* (p. 245-263). Springer Nature Switzerland. doi: 10.1007/978-3-031-35599-8_16

Laal, M. (2013). Positive interdependence in collaborative learning. *Procedia - Social and Behavioral Sciences*, *93*, 1433-1437. doi: 10.1016/j.sbspro.2013.10.058

Leng, J., & Guo, R. (2018). Research on discourse analysis of critical thinking in online collaborative platforms: An overview. *E-education Research*, *39*(2), 26-31. doi: 10.13811/j.cnki.eer.2018.02.004

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint*. doi: 10.48550/arXiv.1907.11692

Makki, I., Alhalabi, W., & Adham, R. (2019). Using emotion analysis to define human factors of virtual reality wearables. *Procedia Computer Science*, *163*, 154-164. doi: 10.1016/j.procs.2019.12.097

Makransky, G., Terkildsen, T., & Mayer, R. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, *60*, 225-236. doi: 10.1016/j.learninstruc.2017.12.007

Morgan, J., & Castle, S. (2023). Arts–research collaboration: Reflections on collaboration as creative method. *Qualitative Inquiry*, *30*(3-4), 291-300. doi: 10.1177/10778004231176280

Moulin, C., Sugawara, K., Kaeri, Y., Fujita, S., & Abel, M. (2014). Collaborative brainstorming activity results and information systems. In *Lecture notes in computer science.* Springer, Berlin. doi: 10.1007/978-3-662-45550-0_40

Naranjo, J. E., Sanchez, D. G., Robalino-Lopez, A., Robalino-Lopez, P., Alarcon-Ortiz, A., & Garcia, M. V. (2020). A scoping review on virtual reality-based industrial training. *Applied Sciences*, *10*(22). doi: 10.3390/app10228224

Nunamaker, J., Dennis, A., Valacich, J., Vogel, D., & George, J. (1991). Electronic meeting systems. *Communications of the ACM*, *34*(7), 40-61. doi: 10.1145/105783.105793

Oh, H., & Son, W. (2022). Cybersickness and its severity arising from virtual reality content: A comprehensive study. *Sensors*, *22*(4). doi: 10.3390/s22041314

Osborn, A. (1963). *Applied imagination: Principles and procedures of creative problem-solving*. New York: Scribner.

Parmar, D., Isaac, J., Babu, S., D'Souza, N., Leonard, A., & Jörg, S. (2016). Programming moves: Design and evaluation of applying embodied interaction in virtual environments to enhance computational thinking in middle school students. In *Ieee virtual reality (vr)* (p. 131-140). IEEE. doi: 10.1109/VR.2016.7504696

Paulsen, L., Dau, S., & Davidsen, J. (2024). Designing for collaborative learning in immersive virtual reality: A systematic literature review. *Virtual Reality*, *28*(63). doi: 10.1007/s10055-024-00975-4

Pizzolante, M., Bartolotta, S., Sarcinella, E. D., Chirico, A., & Gaggioli, A. (2024). Virtual vs. real: Exploring perceptual, cognitive and affective dimensions in design product experiences. *BMC Psychology*, *12*(1). doi: 10.1186/s40359-023-01497-5

Roberts, D., Duckworth, T., Moore, C., Wolff, R., & O'Hare, J. (2009). Comparing the end to end latency of an immersive collaborative environment and a video conference. In *13th ieee/acm international symposium on distributed simulation and real time applications* (p. 89-94). IEEE. doi: 10.1109/DS-RT.2009.43

Salas, E., Cooke, N., & Rosen, M. (2008). On teams, teamwork, and team performance: discoveries and developments. *Hum Factors*, *50*(3), 540-547. doi: 10.1518/001872008X288457

Sanaei, M., Machacek, M., Gilbert, S., Wu, P., & Oliver, J. (2023). Comparing perceptions of performance across virtual reality, video conferencing, and face-to-face collaborations. In *Ieee international conference on systems, man, and cybernetics (smc)* (p. 4556-4561). IEEE. doi: 10.1109/SMC53992.2023.10394218

Seeber, I., de Vreede, G. J., Maier, R., & Weber, B. (2017). Beyond brainstorming: Exploring convergence in teams. *Journal of Management Information Systems*, *34*(4), 939-969. doi: 10.1080/07421222.2017.1393303

Sellers, J., Helton, W., Naswall, K., Funke, G., & Knott, B. (2014). Development of the team workload questionnaire (twlq). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*, 989-993. doi: 10.1177/1541931214581207

Sharma, G., Shroff, G., & Dewan, P. (2011). Workplace collaboration in a 3d virtual office. In *Ieee international symposium on vr innovation* (p. 3-10). IEEE. doi: 10.1109/ISVRI.2011.5759582

Shen, L., Zhang, J., Ou, J., Zhao, X., & Zhou, J. (2021). Constructing emotional consensus and utilizing unpaired data for empathetic dialogue generation. In *Findings of the association for computational linguistics: Emnlp 2021* (p. 3124-3134). Association for Computational Linguistics. doi: 10.18653/v1/2021.findings-emnlp.268

Siegle, D. (2020). I have an idea i need to share: Using technology to enhance brainstorming. *Gifted Child Today*, *43*(3), 205-211. doi: 10.1177/1076217520919967

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549-3557. doi: 10.1098/rstb.2009.0138

Slater, M., Banakou, D., Beacco, A., Gallego, J., Macia-Varela, F., & Oliva, R. (2022). A separate reality: An update on place illusion and plausibility in virtual reality. *Frontiers in Virtual Reality*, *3*. doi: 10.3389/frvir.2022.914392

Slater, M., Cabriera, C., Senel, G., Banakou, D., Beacco, A., Oliva, R., & Gallego, J. (2023). The sentiment of a virtual rock concert. *Virtual Reality*, *27*, 651-675. doi: 10.1007/s10055-022-00685-9

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-group behavior in a virtual and real environment: A comparative study. *Presence Teleoperators and Virtual Environments*, *9*(1), 37-51. doi: 10.1162/105474600566600

Snellen, H. (1862). *Letterproeven, tot bepaling der gezigtsscherpte* (No. v. 1). J. Greven. Retrieved from `https://books.google.co.uk/books?id=wn9jAAAAcAAJ`

Soleymani, M., Asghari-Esfeden, S., Fu, Y., & Pantic, M. (2016). Analysis of eeg signals and facial expressions for continuous emotion detection. *IEEE Transactions on Affective Computing*, *7*(1), 17-28. doi: 10.1109/TAFFC.2015.2436926

Stahl, G., & Maznevski, M. (2021). Unraveling the effects of cultural diversity in teams: A retrospective of research on multicultural work groups and an agenda for future research. *Journal of International Business Studies*, *52*(1), 4-22. doi: 10.1057/s41267-020-00389-9

Steed, A., Pan, Y., Zisch, F., & Steptoe, W. (2016). Getting the message across in collaborative virtual environments: An empirical study. *Presence: Teleoperators and Virtual Environments*, *25*(1), 1–15. doi: 10.1162/PRES$_{a0}$0238

Stephan, W., & Finlay, K. (1999). The role of empathy in improving intergroup relations. *Journal of Social Issues*, *55*(4), 729-743. doi: 10.1111/0022-4537.00144

Taboada, M. (2016). Sentiment analysis: An overview from linguistics. *Annual Review of Linguistics*, *2*(1). doi: 10.1146/annurev-linguistics-011415-040518

Tan, S., Luyten, K., Van Den Bergh, J., & Schöning, J. (2014). The role of physiological cues during remote collaboration. *Presence*, *23*(1), 90-107. doi: 10.1162/PRES$_{a0}$0168

Tate, A. (2011). I-room: Augmenting virtual worlds with intelligent systems. *IEEE Internet Computing*, *15*(5), 56-61. doi: 10.1109/MIC.2011.71

Taufik, D., Kunz, M., & Onwezen, M. (2021). Changing consumer behaviour in virtual reality: A systematic literature review. *Computers in Human Behavior Reports*, *3*. doi: 10.1016/j.chbr.2021.100093

Thompson, M., Bilgin, C., Eberhart, R., Cho, L., Anteneh, M., & Klopfer, E. (2021). Visualizing the collaborative problem solving process in an immersive cross platform game. In *7th international conference of the immersive learning research network (ilrn)* (p. 1-6). IEEE. doi: 10.23919/iLRN52045.2021.9459336

To, C., & Ko, K. (2016). Problematizing the collaboration process in a knowledge-development context. *Journal of Business Research*, *69*(5), 1604-1609. doi: 10.1016/j.jbusres.2015.10.025

Tomprou, M., Kim, Y., Chikersal, P., Woolley, A., & Dabbish, L. (2021). Speaking out of turn: How video conferencing reduces vocal synchrony and collective intelligence. *PLOS ONE*, *18*(6). doi: 10.1371/journal.pone.0247655

van Atteveldt, W., van der Velden, M., & Boukes, M. (2021). The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. *Communication Methods and Measures*, *15*(2), 121-140. doi: 10.1080/19312458.2020.1869198

van der Meer, N., van der Werf, V., Brinkman, W.-P., & Specht, M. (2023). Virtual reality and collaborative learning: a systematic literature review. *Frontiers in Virtual Reality*, *4*. doi: 10.3389/frvir.2023.1159905

Wan, T., Shi, R., Xu, W., Li, Y., Atkinson, K., Yu, L., & Liang, H.-N. (2024). Hands-free multi-type character text entry

in virtual reality. *Virtual Reality*, *28*(8). doi: https://doi.org/10.1007/s10055-023-00902-z

Wang, K., Yan, Y., Zhang, H., Liu, X., & Wang, X. (2023). Eye-shaped keyboard for dual-hand text entry in virtual reality. *Virtual Reality and Intelligent Hardware*, *5*(5), 451-469. doi: 10.1016/j.vrih.2023.07.001

Xu, Q., Zhang, B., Wang, J., Liu, X., & Zhu, M. (2023). A collaborative learning and support system for stem education and learning analytics. In *Ieee integrated stem education conference (isec)* (p. 66-72). IEEE. doi: 10.1109/ISEC57711.2023.10402298