Measuring Quality of Mediated Social Communication

Navya N. Sharan[1], Alexander Toet[1], Tina Mioch[1], Omar Niamut[1],

Jan B.F. van Erp[1,2]

[1]TNO, The Hague, The Netherlands

[2]University of Twente, Enschede, The Netherlands

ABSTRACT

We recently developed the Holistic Social Presence Questionnaire (HSPQ) to measure the quality of mediated social communication experiences. Initial research confirmed the content and face validity of the HSPQ. This study investigates the convergent validity and sensitivity of the HSPQ. Participants completed a decision-making task in groups, using either Microsoft Teams or Mibo to communicate. Upon completing the task, participants rated items from the HSPQ and the validated Networked Minds Questionnaire (NMQ). We expected that Mibo would induce a stronger sense of social presence than MS Teams, since it involves self-movement, orientation, group-forming and spatial audio. The HSPQ showed convergent validity with the NMQ: the ratings on both questionnaires were significantly correlated. However, both the NMQ and HSPQ indicated that participants experienced no significant difference in social presence between both conditions. Hence, further research involving a more immersive communication tool that induces a stronger sense of social presence is needed to assess the sensitivity of the HSPQ.

Keywords: Mediated Social Communication, Social Presence, Quality of Experience

INTRODUCTION

Humans have an inherent social and personal need for communication to maintain their interpersonal relationships and mental wellbeing (Kjeldskov et al. 2004). In our digital age, human social interaction is often mediated. Technologies like videoconferencing software (e.g., Zoom, Microsoft Teams, Skype, etc.) are becoming increasingly popular as they afford a new form of virtual togetherness by facilitating shared and synchronous social activities, thereby substituting real-life interactions (Hacker et al. 2020, Shah et al. 2020). However, these tools still do not provide the affective experience of in-personal social interactions. Moreover, currently available videoconferencing tools are unable to reliably and intuitively convey social and spatial cues between the communication partners (Nadler 2020). This not only results in a lack of social presence (i.e., the feeling of being in the presence of, and having an affective and intellectual connection with, other persons (Biocca et al. 2003)), but also leads to physical and mental exhaustion (Hacker et al. 2020), a phenomenon that is also known as Zoom-fatigue (Nadler 2020).

New (VR, AR or MR-based) communication systems may afford more immersive social interactions. In comparison to regular video- or audio-conferencing tools, these new communication technologies may better approximate the experience of face-to-face (F2F) meetings by eliciting both a sense social presence and a sense of shared environment. To evaluate and optimize these systems, there is a need for measures that efficiently and completely evaluate users' quality of experience. We, therefore, developed a general social presence questionnaire called the Holistic Social Presence Questionnaire (HSPQ: Toet et al. 2020). The tool is general in the sense that it applies to any kind of mediate social communication setting (e.g., VR, AR, MR). In addition, it is concise and holistic as it addresses user experience at all relevant psychological levels while using only a single item to measure each relevant outcome. As a result, it can provide all the necessary information while being minimally disruptive to user experience.

Initial validation studies (Toet et al. 2020) confirmed the content and face validity of the HSPQ. The aim of this study is to establish the convergent validity and sensitivity of the HSPQ. The first objective of this study is to answer the research question: to what extent does the HSPQ measure holistic social presence, for different mediated social communication systems, in comparison to existing validated social presence questionnaires? The second objective is to answer the research question: is the HSPQ sufficiently sensitive to measure differences in holistic social presence experienced with different mediated social communication systems with varying levels of immersion?

In this study, participants completed a collaborative decision-making task using two different mediated communication tools: Microsoft Teams or Mibo. After the task, participants responded to items from the HSPQ and another validated social presence questionnaire. We expected that Mibo would yield a higher sense of social presence than MS Teams, because it affords a more natural and intuitive user interaction (spatial orientation and movement, group forming, spatial audio, etc.; see

Figure 1). To measure the convergent validity of the HSPQ, the correlation between the responses from both questionnaires and both environments was determined. To investigate the sensitivity of the HSPQ, the difference in responses between both environments was determined.

Figure 1. In MS Teams (left) users are represented in 2D in a rectangular grid layout, whereas in Mibo (right) they can walk around in a 3D world with their webcam as their head

METHOD

Design

Participants performed the experiment in groups of three, which were formed based on availability. To avoid group effects, participants who knew each other personally or who had frequently interacted before were not grouped together. Groups were randomly assigned to either the MS Teams or Mibo condition.

The independent variable is the level of immersion associated with the mediated communication tool (i.e., MS Teams or Mibo). The dependent variables are the measures in the HSPQ and Networked Minds Questionnaire (NMQ). Two items were adapted from the Networked Minds Social Presence Measure (Harms & Biocca 2004) and ten items were adapted from the Networked Minds Social Presence Inventory (Biocca & Harms 2003). Together, these formed the NMQ used in this experiment. Eleven items were adapted from the HSPQ (Toet et al. 2020). All questionnaire items were rated on a scale of 1 to 7 from strongly disagree to strongly agree.

Participants

A total of 36 participants (16 males, 20 females) participated in this experiment, with 6 groups of 3 participants in each of the two conditions. Participants ranged from 20 to 44 years (Mean age: 27.19 years, SD age: 4.77 years).

Stimulus and Apparatus

Participants completed a collaborative decision-making task, using either MS Teams or Mibo, using their own computer and webcam from their own home. The task involved reading positive and negative attributes of four potential candidates and selecting the best candidate for a job (Schulz-Hardt et al. 2006). Within a group, each participant received slightly different details about the same four candidates. The overall best candidate was not identifiable based on the individual information from each of the participants and could only be found by pooling and integrating their unique information. After the experiment, participants filled out the HSPQ and the NMQ that were presented online.

Procedure

After all three participants had joined the virtual meeting (hosted either on MS Teams or Mibo) and had opened an online survey in a different browser window, they filled out an informed consent form and reported some demographics (age and gender).

Next, each participant, individually, received a document with information about all four candidates. Then, they had 10 minutes to read the candidate information they received and rank the candidates in order of preference for the job. Once participants had ranked their preference in the online survey, they were instructed to try, as a group, to find the best candidate for the job based on all the information they had collectively available. They had 15 minutes to reach a consensus about which of the four candidates they would hire. After the group discussion, they indicated their final choice on the same online survey.

In the final section of the experiment, participants rated items from the HSPQ and NMQ about the quality of their experience in the current mediated communication situation. These items were presented in a random order. Once participants completed this section, they were debriefed about the aim of the experiment. The experimenter also answered any questions they had about the experiment. They were thanked for their time and participation.

RESULTS

Convergent Validity of the HSPQ

The ratings for different items of the HSPQ and NMQ were averaged to calculate one value for each questionnaire per participant. Since this data was normally distributed, a paired samples *t*-test was conducted. This showed that the difference between the average of all HSPQ items (M = 5.55, SD = 0.80) was not significantly different from the average of all NMQ items (M = 5.42, SD = 0.64). Furthermore, the two variables significantly correlated, with a correlation coefficient of 0.67 ($p < .001$; see Figure 2).

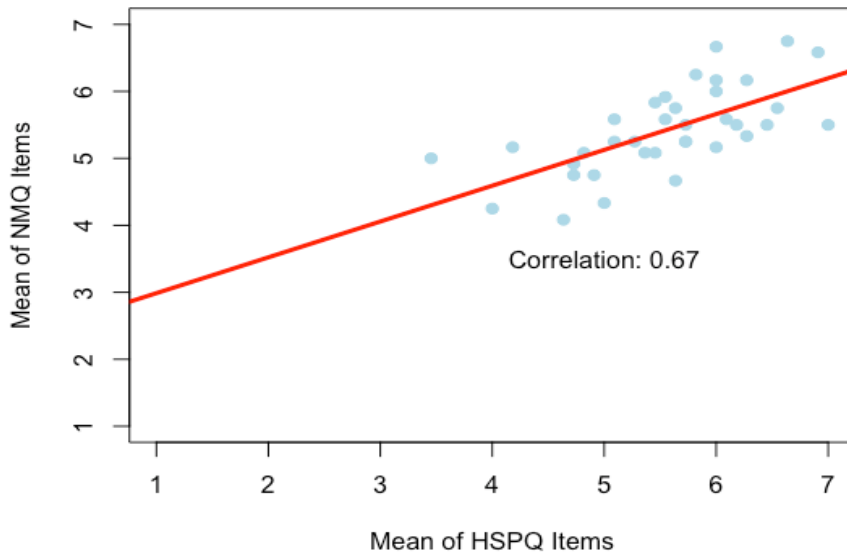Figure 2. Correlation between the average of HSPQ items and the average of NMQ items

Interrater Reliability of HSPQ

To assess the interrater reliability we computed the Intraclass Correlation Coefficient (ICC) based on a mean-rating, consistency, two-way mixed-effects model (Koo & Li 2016), both for Mibo and MS Teams. The ICC value was .64 ($p < .001$) for Mibo and .60 ($p < .001$) for MS Teams, indicating a statistically significant moderate interrater agreement for both conditions.

Sensitivity of the HSPQ

To measure the sensitivity of the HSPQ to varying levels of immersion, we compared the ratings from the two conditions: MS Teams and Mibo. Since the independent variable has two levels and the dependent variable is ordinal and not normally distributed, a Mann-Whitney U test was used. This test was applied individually to each of the 11 HSPQ items.

We found no significant difference between both communication conditions across individual items of the HSPQ. The only exception to this is HSP9, i.e., "*My appearance seemed normal to the other participants*", which is rated higher in MS Teams than in Mibo ($W = 79.5$, $p = .006$).

DISCUSSION

In this section, the results will be interpreted in line with the three main objectives of this study. It should be noted that this discussion only serves as a concise interpretation of the results. There is, of course, room for further analyses and an in-depth interpretation of the results.

Firstly, an aim of this research is to compare the HSPQ to the NMQ. Overall, the ratings for the HSPQ correlated significantly with the ratings for the NMQ. Breaking down the questionnaires into clusters of related items (i.e., social presence, mutual understanding, affective understanding, and behavioral engagement) showed that the HSPQ and NMQ items correlated significantly for social presence and affective understanding and not for mutual understanding and behavioral engagement. This indicates that items related to themes like social presence and affective understanding are correlated more strongly than the items in other clusters. A possible explanation could be that the items in these clusters are phrased more similarly between both questionnaires than those in other clusters.

Next, the interrater reliability of the HSPQ was computed. The ICC values indicated a moderate reliability for both conditions (MS Teams and Mibo). However, repeated tests are needed to assess the test-retest reliability of the HSPQ.

Finally, the sensitivity of the HSPQ was measured by comparing the ratings for individual items across conditions. The results showed no significant difference between the conditions for any of the individual items, except for an item related to appearance (HSP9). A possible reason for this lack of difference in social presence between both conditions could be the nature of the task, which required the full attention of the participants. Hence, they may have paid insufficient attention to their fellow participants. Unfortunately, we did not collect qualitative data from the participants and this hypothesis is based on informal observations of the experimenter. It is recommended that future research will use a communication tasks that stimulates participants to pay more attention to each other.

CONCLUSION

The objective of this research was to assess the convergent validity of the HSPQ and the sensitivity in comparison with the NMQ, which is an existing validated questionnaire that measures social presence. The HSPQ ratings correlated significantly with the NMQ ratings, supporting previous research in establishing the construct validity of the HSPQ. Additionally, moderate interrater reliability was observed. In conclusion, this research should be viewed as a pilot validation study for the HSPQ. Future studies involving more immersive communication tools and tasks that require participants to pay more attention to each other's social behavior are needed to fully assess the sensitivity of the HSPQ.

REFERENCES

Biocca, F., and Harms, C. (2003), "Guide to the Networked Minds Social Presence Inventory (Version 1.2): Measures of co-presence, social presence, subjective symmetry, and intersubjective symmetry. Downloaded from http://cogprints.org/6743/.

Biocca, F., Harms, C., and Burgoon, J.K. (2003), "Toward a more robust theory and measure of social presence: Review and suggested criteria." *Presence: Teleoperators and Virtual Environments, 12* (5), 456-480.

Hacker, J., vom Brocke, J., Handali, J., Otto, M., and Schneider, J. (2020), "Virtually in this together – how web-conferencing systems enabled a new virtual togetherness during the COVID-19 crisis." *European Journal of Information Systems, 29* (5), 563-584.

Harms, C., and Biocca, F. (2004), "Internal consistency and reliability of the Networked Minds measure of social presence." In: M. Alcaniz, & B. Rey (Eds.), *Seventh Annual International Workshop: Presence 2004*, 246-251, Valencia, Spain:Universidad Politecnica de Valencia.

Kjeldskov, J., Gibbs, M., Vetere, F., Howard, S., Pedell, S., Mecoles, K., and Bunyan, M. (2004), "Using cultural probes to explore mediated intimacy." *Australasian Journal of Information Systems, 11* (2), 102-115.

Koo, T.K., and Li, M.Y. (2016), "A guideline of selecting and reporting intraclass correlation coefficients for reliability research." *Journal of Chiropractic Medicine, 15* (2), 155-163.

Nadler, R. (2020), "Understanding "Zoom fatigue": Theorizing spatial dynamics as third skins in computer-mediated communication." *Computers and Composition, 58*, 102613.

Schulz-Hardt, S., Brodbeck, F.C., Mojzisch, A., Kerschreiter, R., and Frey, D. (2006), "Group decision making in hidden profile situations: Dissent as a facilitator for decision quality." *Journal of Personality and Social Psychology, 91* (6), 1080-1093.

Shah, S.G.S., Nogueras, D., van Woerden, H.C., and Kiparoglou, V. (2020), "The COVID-19 Pandemic: A Pandemic of Lockdown Loneliness and the Role of Digital Technology." *Journal of Medical Internet Research, 22* (11), e22287.

Toet, A., Mioch, T., Gunkel, S.N.B., Sallaberry, C., van Erp, J.B.F., and Niamut, O. (2020), "Holistic quality assessment of mediated immersive multisensory social communication." In: P. Bourdot, V. Interrante, R. Kopper, A.H. Olivier, H. Saito, & G. Zachmann (Eds.), *Virtual Reality and Augmented Reality. EuroVR 2020.*,Vol Lecture Notes in Computer Science, vol 12499, 209-215, Cham:Springer.