On the Usage of the Simulator Sickness Questionnaire for Virtual Reality Research

Pauline Bimberg* Tim Weissker† Alexander Kulik‡
Virtual Reality and Visualization Research, Bauhaus-Universität Weimar

Abstract

Originally developed for military flight simulators in the 1990s, the Simulator Sickness Questionnaire (SSQ) has been widely adopted to quantify sickness elicited by modern virtual reality systems. We illustrate the inherent challenges of applying the SSQ in virtual reality research and highlight large differences that can be found in related work. Based on our observations, we conclude by providing suggestions on how to improve the expressiveness of SSQ results in future studies and encourage researchers to consider simpler measurement methods if their research questions allow.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Empirical studies in HCI; General and reference—Document types—Surveys and Overviews

1 Introduction

Immersive interfaces and virtual environments enable expressive interactions and compelling user experiences, but they come at the risk of severely affecting the users' wellbeing. The problem is known by many names, e.g. simulator sickness and cybersickness [12], and it is an open debate whether these refer to the same or closely related issues [27]. Unless these negative effects can be avoided, they pose a major obstacle to applications of virtual reality.

A large body of research aims at a better understanding of sickness symptoms and the factors of their causation (see [18, 26] for an overview). The main motivation in our community, instead, is to control and ideally prevent such negative effects on the users' wellbeing. Even in studies on task performance or other qualities of novel immersive interfaces, sickness symptoms must be monitored since they are likely to bias most other dependent variables [19, 30]. In the highly probable case of participants reporting somewhat increased symptoms, their severity must be gauged in order to decide on the reliability of other obtained measures. This, in turn, requires interpretation on an absolute scale or at least support for coarse comparisons across different settings.

The Simulator Sickness Questionnaire (SSQ) [16] seems to meet these requirements. It was introduced as a calibrated absolute measurement tool. Its popularity over decades led to a large body of reference measures that can be consulted for comparison (the original paper counts more than 3000 citations). On closer inspection, however, the interpretation of SSQ scores as suggested by Kennedy et al. appears dubious. Even coarse comparisons of scores across different studies are impaired by large differences in the application and usage of the SSQ. Based on our paper collection gathered over the past years, we review and discuss some of the challenges and contradictions involved in the usage of this questionnaire. We then derive suggestions and guidelines for future work in our community.

*e-mail: clara.pauline.bimberg@uni-weimar.de
†e-mail: tim.weissker@uni-weimar.de
‡e-mail: kulik@uni-weimar.de

SSQ Symptom	Weight N	Weight O	Weight D
General discomfort	1	1	
Fatigue		1	
Headache		1	
Eyestrain		1	
Difficulty focusing		1	1
Increased salivation	1		
Sweating	1		
Nausea	1		1
Difficulty concentrating	1	1	
Fullness of head			1
Blurred vision		1	1
Dizzy (eyes open)			1
Dizzy (eyes closed)			1
Vertigo			1
Stomach awareness	1		
Burping	1		
Total	[1]	[2]	[3]

$$N = [1] \times 9.54$$
$$O = [2] \times 7.58$$
$$D = [3] \times 13.92$$
$$TS = ([1] + [2] + [3]) \times 3.74$$

Table 1: Computation of nausea (N), oculomotor disturbance (O), disorientation (D), and total simulator sickness (TS) scores as introduced by Kennedy et al. [16]. The red brackets were added to the original formula for clarification.

2 The Simulator Sickness Questionnaire (SSQ)

Designed as a refinement of the Pensacola Motion Sickness Questionnaire (MSQ) for computer-based simulators, the simulator sickness questionnaire (SSQ) asks participants to provide subjective severity ratings of 16 symptoms on a scale from 0 (no perception) to 3 (severe perception) after the exposure [16]. As visualized in Table 1, the ratings for the individual symptoms are grouped into three non-mutually exclusive categories representing symptoms for nausea (N), oculomotor disturbance (O), and disorientation (D). The score of each category is defined as the sum of its symptom scores multiplied by a constant scaling factor. Moreover, a total simulator sickness score (TS) combining the three sub-scales can be computed in a similar way. In general, higher scores on each scale indicate stronger perceptions of the underlying sickness symptoms and are therefore undesired. Based on a large sample of SSQ data gathered from military pilots, it is suggested that total scores can be associated with negligible (< 5), minimal ($5 - 10$), significant ($10 - 15$), and concerning ($15 - 20$) symptoms. A simulator resulting in total scores above 20 is considered "bad" [27]. Similar thresholds can be assumed for the sub-scales nausea, oculomotor disturbance, and disorientation as the scaling factors were chosen to produce scales with similar variabilities [16].

3 Challenges of using the SSQ

The SSQ has been widely adopted as a simple research tool to quantify and compare simulator sickness in virtual environments. However, there are some considerations that should be taken into account when conducting, evaluating, and interpreting the SSQ in the context of a user study.

Misleading Formula The introductory paper of the SSQ omitted the notation of brackets in the total score computation [16] although it is meant to sum up the category scores first before multiplying them with the scaling factor (see red brackets in Table 1). The (wrong) notation without brackets implies that only the disorientation-related category sum has to be multiplied by the scaling factor of 3.74, which would generally result in lower total scores than Kennedy et al. themselves reported. It would also result in a highly imbalanced weighting that overrates disorientation symptoms compared to nausea and oculomotor disturbance.

Non-Uniform Discretization Although the scaling factors were designed to ensure comparability between the sub-scales and the total score, they introduce an unequally strong discretization of potential outcome scores with different maximum values. In particular, the formulas dictate that nausea, oculomotor disturbance, disorientation, and total scores can only increment in steps of 9.54, 7.58, 13.92, and 3.74, respectively. As the four symptoms *General discomfort*, *Difficulty focusing*, *Difficulty concentrating*, and *Blurred vision* appear in two categories each, their contribution to the total score is twice as high, resulting in coarser increments of 7.48 on the TS scale. As a result, the comparability of total scores within and across studies is challenging as similar scores could be generated by a few symptoms with high influence or multiple symptoms with small influence. To mitigate these issues, related work proposed alternative symptom categorizations and scaling factors [7, 8].

Military Reference Population The absolute interpretation thresholds of SSQ scores as introduced by Stanney et al. were derived from a sample of military aviators, who might be less susceptible to simulator sickness than the general population. A study by the same authors with college students in flight-unrelated virtual environments revealed considerably higher scores on the SSQ than the aviator dataset, where all average sub-scale scores were above the "bad simulator" threshold of 20 [27]. While the authors attribute this to differences of stimuli rather than samples, average scores above 20 are frequently observed in related VR studies (e.g. [4, 11, 13, 15, 20, 24, 31]). This is not surprising since small increases of only a few symptoms can already suffice for reaching the threshold. It therefore seems that the military aviators used for the initial calibration were indeed not representative for a more general population.

Missing Baseline Scores The SSQ assumes that participants felt completely well before the exposure to the simulator. However, assuming a rating of 0 (no perception) on each of the 16 symptoms is challenging as certain symptoms could also be provoked depending on the time of the day, the journey to the experimental venue, or the current mood of a participant. Since the focus of many studies is to investigate whether an experimental manipulation has a negative effect on simulator sickness, it therefore seems reasonable to administer the SSQ before and after the exposure and only focus on the resulting differences. However, Kennedy et al. explicitly criticize the low reliability of difference scores and instead suggest asking whether participants felt "sick" and in their "usual state of fitness" prior to the experiment [16]. This procedure is further supported by studies mentioning that pre-exposure questionnaires can bias the resulting post-exposure scores negatively [32]. Nevertheless, the assumption of erroneous baseline scores could be another reason for the rather high SSQ scores observed in related work.

Subjective Ratings Similar to any other questionnaire, an absolute severity rating of sickness symptoms is highly subjective and might therefore differ between participants experiencing the same symptoms. The coarse classification of symptoms into only four severity classes might mitigate this issue slightly, but the boundaries are still volatile. While some participants could be overly sensitive, others could deliberately understate the severity of symptoms in order not to appear weak [2]. Moreover, the concrete definitions of medical terms like *stomach awareness* or *vertigo* might not be clear to all participants, which could further disperse the obtained results.

4 Diversity in SSQ Administration and Reporting

Despite the issues mentioned above, the SSQ seems to be the most common sickness measure in the our research community. However, we observe a huge diversity in its administration and reporting across publications. In the following, we would like to give a short overview of some differing approaches within the community and their influence on the interpretation of the resulting scores.

Scoring Among publications, different computation variants of the total SSQ score and its sub-scales out of the raw values can be observed. While the missing brackets in the original paper do not seem to have led to confusion in the community, other issues seem to be common. One variation of the formula that can be encountered is the computation of the total TS score by multiplying the already weighted sub-scale values N, O and D with 3.74. (e.g. [9]). This approach naturally leads to higher values as well as a change in the weighting of the different items through the corresponding sub-scale(s). In contrast, some other papers seem to use the SSQ without applying any scaling factors or sub-scores in the calculation of the total score (e.g. [3,5]). This approach, in turn, leads to very small SSQ values in comparison to the original method of calculation since it forgoes the double-counting of certain items as well as the application of the final scaling factor.

Administration Unlike recommended in the original paper, the SSQ is often administered as a pre- and post measurement (e.g. [3, 5, 14, 20, 22, 28, 29, 33]). Applying the SSQ as a post-measurement only builds on the assumption that participants who report to be in their "usual state of fitness" [16] before exposure to the stimuli would indicate absolutely no sickness symptoms in the full questionnaire. In light of several publications reporting pre-exposure values that, taken by themselves, would already indicate "significant" or "concerning" amounts of simulator sickness [5, 20, 29], this assumption appears at least questionable. We even found examples of reduced sickness symptoms after exposure to the stimuli (e.g. [3, 29]), which can be challenging to interpret. Comparing pre- and post-study questionnaires can be beneficial to better understand the effects of given stimuli on sickness symptoms. Repeating the same questionnaire, however, raises awareness to such symptoms, which can lead to their higher reported sickness scored independent of the tested stimuli [32]. Please note that such repeated filling of the same questionnaire cannot be avoided at all if different conditions are tested with repeated measures involving the same participants (within-subjects design).

Reporting The level of detail at which the results of the simulator sickness questionnaire are reported and discussed varies greatly between publications. This ranges from the full reporting of all sub-scales [1, 20] over the report of the total result [4, 14] and reporting only the difference between pre- and post measurements [1, 9] onto mentioning significances or non-significances only [10, 23, 33, 34]. Especially the latter two approaches do not disclose any information about the overall wellbeing reported by the participants of the relating study or the way in which the scores where originally computed by the authors.

The described methodological variations of applying the SSQ can most likely be attributed to practical issues, but they complicate a meaningful interpretation of results and impede the comparisons of sickness symptoms between different studies. In the following section, we therefore suggest guidelines for using the SSQ and pragmatic alternatives in research on virtual reality.

5 Suggestions on SSQ Usage in VR Research

Based on our observations of the challenges and diversity in administration and reporting, we propose the following suggestions for future studies involving the SSQ:

- When citing the original paper by Kennedy et al. [16], practitioners should use the original scoring formulas suggested by the authors. However, it should be noted that the brackets in the total score computation are mandatory but missing in the publication. Diversions from the given formulas should always be explicitly stated and justified.

- Despite the demonstration of demand characteristics [32], it seems reasonable to administer the SSQ both before and after the exposure to an experimental condition. This is especially true for within-subject studies, where participants have to complete the questionnaire multiple times anyways. In addition to the absolute values, difference scores could offer more specific insights into the effects of a condition by eliminating the scores of symptoms that were already present before the exposure. However, if the post-exposure score is smaller than before, it is unclear whether this improvement stems from the experimental manipulation or a general decay of symptoms over time. We therefore suggest considering the experimental condition to have *no negative* instead of a *positive* effect on the participants.

- Scores above 20 should not automatically be attributed to a bad simulator as the proposed threshold values by Stanney et al. [27] seem very strict when applied to non-aviators. However, stating the maximum possible total score of 235.62 to justify that a given score is comparatively low is also misleading [15, 21]. Future work should determine refined threshold values based on samples from a more general population.

- Practitioners are encouraged to report means, medians, and standard deviations for all sub-scales and the total score – even if inferential statistical tests on these values are not significant. These values help to understand the absolute magnitude of scores and allow for more detailed insights into the symptom profiles elicited by an experimental condition.

We believe that the SSQ remains a meaningful research tool and hope that our usage suggestions can help to improve the expressiveness and comparability of results. However, we also encourage researchers to consider if an administration of the SSQ is really necessary to answer their research questions. In many cases, the main motivation of applying the SSQ is not a detailed analysis of the distribution of nausea, oculomotor disorientation, and disorientation symptoms elicited by an experimental manipulation. Instead, researchers are often either interested in the overall development of their participants' wellbeing or merely aim at verifying that the wellbeing of their sample did not influence other dependent variables. In this case, it might be sufficient to repeatedly ask participants a single question about their wellbeing instead of the 16-symptom SSQ, which could highly reduce the duration and workload of the user study. Examples for single-scale questionnaires in the literature include the Fast Motion Sickness Scale [17], the Misery Scale [6], and relative ratings of wellbeing as compared to the start of the experiment [13, 25].

6 Conclusion and Future Work

In this paper, we illustrated the inherent limitations of the SSQ as introduced by Kennedy et al. and provided an overview of its diverse administration variants applied in the literature. In summary, the expressiveness of the SSQ suffers from a misleading formula, non-uniform discretizations, a military reference population, missing baseline scores, and subjective ratings. Nevertheless, the SSQ has been widely adopted to quantify simulator sickness in virtual reality research, which allows us to build on a large amount of experience from related work. To strengthen this exchange, future studies applying the SSQ should therefore provide a thorough report of their procedures and results and be cautious about their interpretation within and beyond the study context. Our suggestions could, from our point of view, improve the expressiveness of the results by a fair amount, but the honest and objective quantification of simulator sickness still remains a challenge for virtual reality research. As it has been proposed by Dennison et al., we should reach a better understanding of the factors that influence this phenomenon both on the side of the user and the side of the application [12]. This should enable us to make predictions about problematic content for certain user groups and to minimize their influence in the design phase of new applications.

In our future work, we intend to further explore the use of simple scales in addition or as an alternative to the SSQ. From our experiences as well as the experiences of other researchers with similar approaches, we hope to advise on a suitable wording of such a question (or two to three questions if needed) and to correlate the obtained scores to the SSQ.

References

[1] J. Bhandari, P. MacNeilage, and E. Folmer. Teleportation without spatial disorientation using optical flow cues. In *Proceedings of Graphics Interface*, vol. 2018, 2018.

[2] F. Biocca. Will simulation sickness slow down the diffusion of virtual environment technology? *Presence: Teleoperators & Virtual Environments*, 1(3):334–343, 1992.

[3] J. P. Bliss, P. D. Tidwell, and M. A. Guest. The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence: Teleoperators & Virtual Environments*, 6(1):73–86, 1997.

[4] B. Bolte, G. Bruder, F. Steinicke, K. Hinrichs, and M. Lappe. Augmentation techniques for efficient exploration in head-mounted display environments. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, pp. 11–18. ACM, 2010.

[5] B. Bolte, F. Steinicke, and G. Bruder. The jumper metaphor: an effective navigation technique for immersive display setups. In *Proceedings of Virtual Reality International Conference*, pp. 1–7, 2011.

[6] J. Bos, S. Mackinnon, and A. Patterson. Motion Sickness Symptoms in a Ship Motion Simulator: Effects of Inside, Outside and No View. *Aviation, space, and environmental medicine*, 76:1111–8, 01 2006.

[7] S. Bouchard, G. Robillard, and P. Renaud. Revising the factor structure of the Simulator Sickness Questionnaire. *Annual review of cybertherapy and telemedicine*, 5(Summer):128–137, 2007.

[8] S. Bruck and P. A. Watters. The factor structure of cybersickness. *Displays*, 32(4):153–158, 2011.

[9] J.-R. Chardonnet, M. A. Mirzaei, and F. Mérienne. Features of the Postural Sway Signal as Indicators to Estimate and Predict Visually Induced Motion Sickness in Virtual Reality. *International Journal of Human–Computer Interaction*, 33(10):771–785, 2017. doi: 10.1080/10447318.2017.1286767

[10] W. Chen, N. Ladeveze, C. Clavel, D. Mestre, and P. Bourdot. User cohabitation in multi-stereoscopic immersive virtual environment for individual navigation tasks. In *2015 IEEE Virtual Reality (VR)*, pp. 47–54. IEEE, 2015.

[11] M. Dennison and M. D'Zmura. Effects of unexpected visual motion on postural sway and motion sickness. *Applied ergonomics*, 71:9–16, 2018.

[12] M. S. Dennison and D. M. Krum. Unifying Research to Address Motion Sickness. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1858–1859. IEEE, 2019.

[13] A. S. Fernandes and S. K. Feiner. Combating VR sickness through subtle dynamic field-of-view modification. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 201–210, March 2016. doi: 10.1109/3DUI.2016.7460053

[14] S. Freitag, B. Weyers, and T. W. Kuhlen. Interactive exploration assistance for immersive virtual environments based on object visibility and viewpoint quality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 355–362. IEEE, 2018.

[15] D. H. Jeong, C. S. Lee, G. B. Jeon, C. G. Song, S. Babu, and L. Hodges. Differentiation on information gathering ability in real and virtual world. In *Proceedings of Pacific Graphics*, pp. 157–159, 2005.

[16] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993. doi: 10.1207/s15327108ijap0303_3

[17] B. Keshavarz and H. Hecht. Validating an Efficient Method to Quantify Motion Sickness. *Human Factors*, 53(4):415–426, 2011. doi: 10.1177/0018720811403736

[18] J. R. Lackner. Motion sickness: more than nausea and vomiting. *Experimental brain research*, 232(8):2493–2510, 2014.

[19] C. S. Maraj, K. A. Badillo-Urquiola, S. G. Martinez, J. A. Stevens, and D. B. Maxwell. Exploring the impact of simulator sickness on the virtual world experience. In *Advances in human factors, business management, training and education*, pp. 635–643. Springer, 2017.

[20] S. Mirhosseini, I. Gutenko, S. Ojal, J. Marino, and A. E. Kaufman. Automatic speed and direction control along constrained navigation paths. In *2017 IEEE Virtual Reality (VR)*, pp. 29–36. IEEE, 2017.

[21] K. R. Moghadam, C. Banigan, and E. D. Ragan. Scene Transitions and Teleportation in Virtual Reality and the Implications for Spatial Awareness and Sickness. *IEEE transactions on visualization and computer graphics*, 2018.

[22] J. Munafo, M. Diedrick, and T. A. Stoffregen. The virtual reality head-mounted display Oculus Rift induces motion sickness and is sexist in its effects. *Experimental Brain Research*, 235(3):889–901, 2017.

[23] R. Paris, J. Klag, P. Rajan, L. Buck, T. P. McNamara, and B. Bodenheimer. How Video Game Locomotion Methods Affect Navigation in Virtual Environments. In *ACM Symposium on Applied Perception 2019*, SAP '19. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3343036.3343131

[24] E. D. Ragan, S. Scerbo, F. Bacim, and D. A. Bowman. Amplified head rotation in virtual reality and the effects on 3d search, training transfer, and spatial orientation. *IEEE transactions on visualization and computer graphics*, 23(8):1880–1895, 2016.

[25] L. Rebenitsch and C. Owen. Individual Variation in Susceptibility to Cybersickness. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, p. 309–317, 2014. doi: 10.1145/2642918.2647394

[26] L. Rebenitsch and C. Owen. Review on cybersickness in applications and visual displays. *Virtual Reality*, 20(2):101–125, Jun 2016. doi: 10.1007/s10055-016-0285-9

[27] K. M. Stanney, R. S. Kennedy, and J. M. Drexler. Cybersickness is Not Simulator Sickness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 41(2):1138–1142, 1997. doi: 10.1177/107118139704100292

[28] T. A. Stoffregen and L. Smart. Postural instability precedes motion sickness. *Brain Research Bulletin*, 47(5):437 – 448, 1998. doi: 10.1016/S0361-9230(98)00102-6

[29] E. Suma, S. Finkelstein, M. Reid, S. Babu, A. Ulinski, and L. F. Hodges. Evaluation of the Cognitive Effects of Travel Technique in Complex Real and Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, 16(4):690–702, July 2010. doi: 10.1109/TVCG.2009.93

[30] S. Weech, S. Kenny, and M. Barnett-Cowan. Presence and cybersickness in virtual reality are negatively related: a review. *Frontiers in psychology*, 10:158, 2019.

[31] B. C. Witmer, J. H. Bailey, and B. W. Knerr. Training Dismounted Soldiers in Virtual Environments: Route Learning and Transfer. Technical report, Army Research Institute for the Behavioral and Social Sciences, Orlando, FL, 1995.

[32] S. D. Young, B. D. Adelstein, and S. R. Ellis. Demand Characteristics in Assessing Motion Sickness in a Virtual Environment: Or Does Taking a Motion Sickness Questionnaire Make You Sick? *IEEE Transactions on Visualization and Computer Graphics*, 13(3):422–428, May 2007. doi: 10.1109/TVCG.2007.1029

[33] C. Zanbaka, S. Babu, D. Xiao, A. Ulinski, L. Hodges, and B. Lok. Effects of travel technique on cognition in virtual environments. In *IEEE Virtual Reality 2004*, pp. 149–286. IEEE, 2004.

[34] C. A. Zanbaka, B. C. Lok, S. V. Babu, A. C. Ulinski, and L. F. Hodges. Comparison of path visualizations and cognitive measures relative to travel technique in a virtual environment. *IEEE Transactions on Visualization and Computer Graphics*, 11(6):694–705, 2005.