Validating an Efficient Method to Quantify Motion Sickness

Behrang Keshavarz and Heiko Hecht

Human Factors: The Journal of the Human Factors and Ergonomics Society 2011 53: 415 originally published online 11 April 2011

DOI: 10.1177/0018720811403736

Validating an Efficient Method to Quantify Motion Sickness

Behrang Keshavarz and **Heiko Hecht,** Johannes Gutenberg-Universität Mainz, Germany

Objective: Motion sickness (MS) can be a debilitating side effect associated with motion in real or virtual environments. We analyzed the effect of expectancy on MS and propose and validate a fast and simple MS measure.

Background: Several questionnaires measure MS before or after stimulus presentation, but no satisfactory tool has been established to quickly capture MS data during exposure. To fill this gap, we introduce the Fast MS Scale (FMS), a verbal rating scale ranging from zero (*no sickness at all*) to 20 (*frank sickness*). Also, little is known about the role of expectancy effects in MS studies. We conducted an experiment that addressed this issue.

Method: For this study, 126 volunteers participated in two experiments. During stimulus presentation, participants had to verbally rate the severity of MS every minute before filling in the Simulator Sickness Questionnaire (SSQ). To measure expectancy effects, participants were separated into three groups with either positive, negative, or neutral expectations.

Results: We compared the verbal ratings with the SSQ scores. Pearson correlations were high for both the SSQ total score ($r = .785$) and the nausea subscore ($r = .828$). No expectancy effects were found.

Conclusion: The FMS is a fast and valid method to obtain MS data. It offers the possibility to record MS during stimulus presentation and to capture its time course. We found expectancy not to play a crucial role in MS. However, the FMS has some limitations.

Application: The FMS offers improved MS measurement. It is fast and efficient and can be performed online in environments such as virtual reality.

Keywords: sensory conflict, visually induced motion sickness, rating scales, expectancy effects

Address correspondence to Behrang Keshavarz, Department of Experimental Psychology, Johannes Gutenberg-University, Wallstraße 3, 55099 Mainz, Germany; e-mail: behrang.keshavarz@uni-mainz.de

HUMAN FACTORS
Vol. 53, No. 4, August 2011, pp. 415–426.
DOI: 10.1177/0018720811403736.

INTRODUCTION

Motion sickness (MS) is a well-known nuisance when traveling by ship, car, train, or airplane. Almost every traveler is familiar with the displeasing feelings of nausea, pallor, cold sweating, or vomiting and has experienced such symptoms at least once in his or her lifetime. Because of the progress in virtual technologies, MS is no longer limited to situations involving real body movement. Equally strong MS-like symptoms can occur in virtual environments (simulator sickness [SS]), such as in flight or driving simulators (e.g., Frank, Casali, & Wierwille, 1988; Stoffregen, Hettinger, Haas, Roe, & Smart, 2000), while playing console video games (Merhi, Faugloire, Flanagan, & Stoffregen, 2007), or while watching a movie on a large screen (cinerama sickness). These sickness types are commonly referred to as visually induced MS (Bos, Bles, & Groen, 2008; Kennedy, Drexler, & Kennedy, 2010; Sharples, Cobb, Moody, & Wilson, 2008).

The sensory conflict theory (Oman, 1990; Reason, 1978; Reason & Brand, 1975) is one possible explanation for MS and SS. Basically it states that a conflict between the different sensory modalities is the cause of the symptoms. It also implies that the stronger the conflict, the more severe the symptoms. Current body position and movements are permanently conveyed to the central nervous system by the visual, vestibular, and proprioceptive organs. Thus, MS can arise if any of these channels are at variance with each other or even inconsistent within themselves (e.g., Brown, Hecht, & Young, 2002). For example, visually induced MS is thought to be caused by a visual-vestibular conflict: Vision indicates self-movement (referred to as *vection* when experienced; Flanagan, May, & Dobie, 2004), whereas the vestibular organs indicate a resting position within Earth's gravitational field. However, the sensory conflict theory is not without controversy, and alternative

theories of the origin of MS do exist, but none of them can explain MS in a definitive way (Bles, Bos, de Graaf, Groen, & Wertheim, 1998; Warwick-Evans, Symons, Fitch, & Burrows, 1998).

One reason for the unsatisfactory theoretical state of MS is the vagueness of symptoms and the inconsistent ways it has been measured. MS per se is a state of reduced well-being, and it is very difficult—if not impossible—to operationalize it in a satisfactory way. The aim of our study was to introduce a simple and fast method to measure MS repeatedly. Before introducing our measure and our studies, we give a brief overview of various existing measurements of MS with a focus on self-reports on the severity of MS symptoms. Objective measures, such as pallor of the participant's face, skin conductance, heart rate variability, hormonal changes, and so on correlate with MS to some extent, but there is no psychophysiological parameter that can satisfactorily capture the severity and time course of MS (see Shupak & Gordon, 2006).

Several questionnaires have been published in the past years with the goal to deliver a reliable measurement of MS severity. The Nausea Profile (Muth, Stern, Thayer, & Koch, 1996) is a multidimensional questionnaire for measuring nausea in general. The Nausea Profile contains three subscales: Somatic Distress (e.g., feeling fatigued, weak, hot), Gastrointestinal Distress (e.g., feeling sick, stomach awareness, ill), and Emotional Distress (e.g., nervous, scared, worried). A total nausea score as well as a score for every subscale can be calculated. The Nausea Profile provides detailed information about nausea but does not record further MS-typical symptoms, such as oculomotor problems. The Nausea Profile could function as a supplemental tool in addition to other methods of measuring MS.

Gianaros, Muth, Mordkoff, Levine, and Stern (2001) developed the Motion Sickness Assessment Questionnaire to measure multiple dimensions of MS. The Motion Sickness Assessment Questionnaire measures not just nausea but also symptoms germane to MS. The questionnaire consists of 16 items total, which have been related to four subscales on the basis of a factor analysis. The subscales include Gastrointestinal Feelings (e.g., sick to the stomach, queasy, nauseated), Central Issues (e.g., ligh-theaded, disorientated), Peripheral Issues (e.g., sweaty, clammy or with cold sweat), and Sopite-Related Issues (e.g., tired or fatigued, drowsy). The Motion Sickness Assessment Questionnaire can be applied to several types of motion sickness, including MS in the presence or absence of true motion.

A more specific tool for measuring visually induced MS is represented by the Simulator Sickness Questionnaire (SSQ) by Kennedy, Lane, Berbaum, and Lilienthal (1993). Because the experience of simulator sickness varies somewhat compared with classical MS in the presence of actual motion, the SSQ symptom list contains some variations as well. The SSQ appears to be the single most important questionnaire regarding simulator, cyber, or virtual reality sickness. The SSQ contains 16 items, which have to be rated by the participants on a 4-level Likert-type scale. They are divided into the three subscales: Disorientation (e.g., vertigo, concentration issues), Nausea (e.g., general discomfort, nausea), and Oculomotor Issues (e.g., eyestrain, focusing issues). A total SSQ score and scores for every subscale can be calculated.

However, questionnaires suffer from the structural disadvantage that they cannot be administered during stimulus presentation. It is rarely possible to collect questionnaire data while running an experiment because of the length of the questionnaire itself or because the main task is too distracting. Therefore, MS data are typically not being collected during stimulus presentation. Consequently, questionnaires preclude the analysis of chronological sequences during the period of stimulus presentation.

To fill this gap, we introduce a continuous and simple way to generate MS data during stimulus presentation with the use of verbal judgments. We have successfully collected experience with a fast and frugal MS scale in the domain of centrifuge experiments (e.g., Young, Sienko, Lyne, Hecht, & Natapoff, 2003) and propose to use this scale for MS assessment at large. Participants provide verbal ratings of MS on a 20-point scale, one rating per minute. On the basis of classical magnitude estimation methods, we felt that a 20-point scale would be optimal, given the trade-off between levels of gradation and ease in representation of the scale (Stevens, 1951).

TABLE 1: Number of Participants, Mean Age, and Standard Deviation by Gender for Both Experiments

	Male		Female		Total	
Group	*N*	M_{age} (*SD*)	*N*	M_{age} (*SD*)	*N*	M_{age} (*SD*)
Experiment 1	31	23.81 (3.09)	34	22.35 (2.75)	65	23.05 (2.99)
Experiment 2	14	28.29 (5.28)	47	23.33 (4.55)	61	24.48 (5.14)
Total	45	25.20 (4.38)	81	22.91 (3.90)	126	23.47 (4.21)

We conducted one experiment that addressed the role of expectancy on MS. We also included data from another experiment to have a broader basis for validating and comparing the results obtained by the FMS (Fast MS Scale) with those obtained by the traditional SSQ scale. Beyond anecdotal reports, little is known about how expectations influence or even distort the severity of MS. We had observers in our laboratory who turned pale and opted out of the experiment after several mentions of the term *motion sickness*. Young, Adelstein, and Ellis (2007) addressed the question whether an MS questionnaire itself influences perceived MS. The authors compared two groups: Group 1 had to fill in the SSQ both before (pretest) and after (posttest) a virtual reality stimulus presentation. Group 2 was given the posttest only. Results indicated higher SSQ posttest scores for the group who filled in the SSQ twice, once prior to the virtual immersion and once after. The authors reason that expectancy played a crucial role in experiments dealing with virtual reality and MS and that this aspect should be taken into account when designing such studies.

From these results, we hypothesized that MS can be affected by instructions that suggest different levels of potential MS. In Experiment 1, we amplified Young et al.'s (2007) method and used written instructions for the participants to induce expectancy effects instead of using sickness questionnaires. Observers were randomly assigned to one of three groups, one receiving a description of potential negative side effects (sickness), one suggestive of positive effects (happiness), and a neutral instruction (control group). All groups then saw the identical footage taken from a race car. In accordance with Young et al., we hypothesized that the negatively instructed group should show higher MS scores compared with the positive and control group. We measured MS using two measures, the minute-by-minute FMS and the SSQ after stimulus presentation.

To validate our data on the use of two different MS scales, we also reanalyzed data from an unrelated experiment (Experiment 2) that was designed to test the role of axis rotation in simulated rollercoaster rides. Three different groups were created, containing either pitch, pitch and roll, or pitch, roll, and yaw motion. A detailed discussion on the effect of rotation axis on MS is not part of this article.

METHODS

Participants

For the first experiment, 65 mostly undergraduate psychology students participated and were randomly assigned to one of the three expectancy conditions (positive, $n = 23$; negative, $n = 21$; neutral, $n = 21$). The additional data set (Experiment 2) contained another 61 observers, to complete a total number of $N = 126$ volunteers in both experiments. Table 1 shows the distribution by gender for both experiments. Participants were naïve with respect to the purposes of the study and received either partial course credit or were rewarded with 5. The stimuli were administered in accordance with the Declaration of Helsinki, which represents a set of principles to ensure research ethics in human experimentation.

Apparatus and Stimuli

In Experiment 1, the stimulus material consisted of a video taken with a camera mounted to the dashboard of a car at the approximate location of the front passenger's eye position. One complete lap on a racetrack (Nürburgring's Nordschleife, Hunsrück-Eifel, Germany) was recorded and repeated once without any break,

resulting in a total video length of 18 min 42 s with a resolution of 720 × 576 pixels. The sound was switched off during the presentation. All stimuli were presented on a large rear-projection screen (260 cm wide and 195 cm high) using a JVC D-ILA DLA SX21 projector. Participants took a seat in a stationary, height-adjustable chair, and a chinrest for head support was placed 300 cm in front of the screen. The display on the projection screen subtended a visual angle of 48° × 36°. For stimulus presentation, Windows Media Player 9 was used at a resolution setting of 1,280 × 720 pixels.

Participants were randomly assigned to one of three groups (positive, negative, or neutral expectation group). To induce different expectations, we varied the instructions, which had to be read carefully before the experiment. Participants in the positive group were told that stimulus presentation should lead to feelings of happiness, joy, and fun. For the negative group, the briefing included expectations of nausea, general discomfort, and even vomiting. The neutral group served as control group and received a neutral instruction without explicit expectations regarding the experiment.

The additional data set contained stimulus material consisting of three videos, each showing a computer-simulated roller-coaster ride. The movies varied in the roller coaster's combination of turns in the yaw, pitch, and roll axes. Only one of these videos was presented in each experimental condition. One round on the rollercoaster lasted between 44 s and 58 s and was repeated without break 16 to 21 times until a total video duration of approximately 15 min was reached. The resolution was 800 × 600 pixels, and stimuli were presented on a 163-cm (width) by 121-cm (height) large projection screen. A chinrest was placed 135 cm in front of the screen. Participants sat in a height-adjustable chair such that their eye height corresponded to the center of the screen. The display on the projection screen subtended a visual angle of 62° × 48°. For stimulus presentation, Windows Media Player 9 was used.

Procedure and Response Measures

The procedures and the response measures were almost identical for both experiments. Before starting the experiment, participants read the particular written instructions to familiarize themselves with the experimental procedure and the response measures. During stimulus presentation, observers gave verbal ratings of experienced MS on a 20-point scale (FMS) ranging from 0 (*no sickness at all*) to 20 (*frank sickness*) every minute. Participants were informed to focus on nausea, general discomfort, and stomach problems and to take these parameters into account when making their judgments. They were asked to ignore other possible distorting effects, such as nervousness, boredom, or fatigue. We are aware that participants might have had difficulties in following the instruction to ignore the irrelevant sensations, and we cannot be entirely certain that the FMS measure is free of symptoms related to but different from nausea proper. However, we believe that the participants were instructed efficiently enough to provide the necessary focus on the relevant symptoms.

As a feature only for Experiment 1, we added a verbal rating of how happy the observers felt. To ascertain that the expectancy manipulation had worked, observers rated their level of happiness also on a scale from 0 (*no happiness at all*) to 20 (*very happy*). Asking solely for MS could have nulled the intended expectancy effect. Note that the happiness scale was not of interest with regard to content and was not part of the data analysis.

The experimenter made sure that the rating procedure was understood before the experiment began. A single rating on the FMS that acted as a baseline measurement was made before stimulus presentation started. While participants watched the video footage, FMS scores were verbally requested and noted by the experimenter once a minute. The very last score was reported at the moment of video fade-out or just before the experiment was prematurely aborted. The latter could occur at the participant's request or if the experimenter considered the MS score to be indicative of impending frank sickness. When a score of 15 had been reached, the experimenter deliberately asked the participant if he or she wanted to abort the experiment. For a total of 6 participants (2 male, 4 female; 2 in the negative, 3 in the positive, and 1 in the neutral group), the experiment was aborted before stimulus

TABLE 2: Mean Peak FMS Scores and SSQ Values (Including SSQ Total Score and Subscores) for Experiments 1 and 2

Group	Peak FMS	SSQ-TS	SSQ-N	SSQ-D	SSQ-O
Experiment 1	7.51 (7.13)	54.55 (36.75)	61.20 (48.96)	93.37 (70.58)	53.41 (34.25)
Experiment 2	4.99 (3.90)	26.67 (21.90)	28.46 (29.24)	40.62 (44.91)	34.30 (22.97)
Total	5.81 (6.04)	41.05 (33.43)	45.35 (43.69)	67.83 (64.96)	44.16 (30.76)

Note. Standard deviations shown in parentheses. FMS = Fast Motion Sickness Scale; SSQ = Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993); TS = total score; N = Nausea subscale; D = Disorientation subscale; O = Oculomotor Issues subscale.

presentation had been finished because of frank sickness for 5 and other reasons for 1.

Immediately after the stimulus presentation had finished, participants filled in the SSQ. They were asked to complete the questionnaire with reference to the moment when they were exposed to the stimulus for the very last time. We thereby tried to ensure that the SSQ measured sickness for the time of stimulus exposure and remained minimally influenced by the recovery process after stimulus offset.

RESULTS

The peak FMS scores reported by the participants and the values for the SSQ, including the total score (SSQ-TS) and the subscores for Nausea (SSQ-N), Disorientation (SSQ-D), and Oculomotor Issues (SSQ-O), were analyzed. Table 2 illustrates the peak FMS scores and SSQ values separately for the main experiment and the additional data set.

Expectancy Effect

A one-factorial ANOVA with group (positive, negative, neutral) as between-subjects factor was calculated. No significant group effects were found for the FMS scores, $F(2, 62) = 0.139$, $p = .871$; the SSQ-TS, $F(2, 62) = 0.145$, $p = .865$; and the SSQ subscores, SSQ-N, $F(2, 62) = 0.336$, $p = .716$; SSQ-O, $F(2, 62) = 0.465$, $p = .630$; and SSQ-D, $F(2, 62) = 0.104$, $p = .910$.

To analyze the time course of MS, a repeated-measures ANOVA with the factor time was calculated. Group was added as between-subjects variable. The 6 participants who aborted the experiment prior to stimulus offset had to be eliminated from this analysis. Huyn-Feldt corrected ($\varepsilon = 0.121$) results showed a significant effect of time, indicating that sickness scores increased with the duration of stimulus presentation, $F(18, 1006) = 22.691$, $p < .001$, $\eta^2 = .288$. Figure 1 shows the time course for the three groups. Again, no expectancy effect was found, $F(2, 56) = 0.471$, $p = .627$, $\eta^2 = .017$.

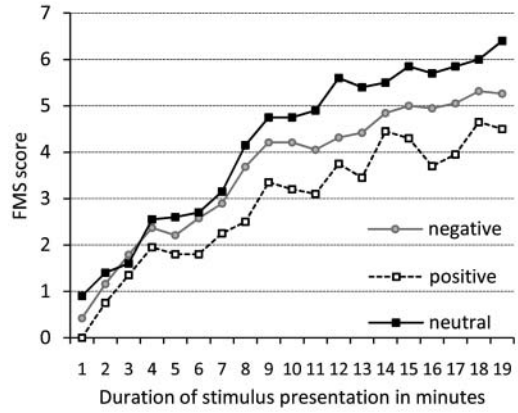

Figure 1. Time courses displaying the Fast Motion Sickness Scale (FMS) scores measured for the three groups (positive, negative, and neutral expectancy) in Experiment 1.

By way of a median split, participants were separated post hoc into a low sickness (FMS lower than 6, $n = 34$) and high sickness (FMS 6 or higher, $n = 31$) group, regardless of the expectation condition. Figure 2 illustrates the different time courses for both groups. The curves for the participants who aborted the experiment before stimulus offset because of frank sickness are shown in Figure 3.

Scale Validation

Figures 4 and 5 display a scatter plot for the peak FMS score and the SSQ-N and SSQ-TS scores for the current experiment and the additional data set. Bivariate correlations were calculated

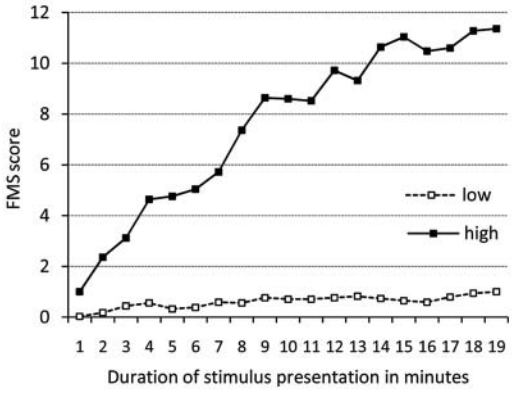

Figure 2. Time courses of the Fast Motion Sickness Scale (FMS) scores for participants separated by median split (low-susceptibility FMS scores below 6, $n = 34$; high-susceptibility FMS scores 6 or higher, $n = 31$) measured in Experiment 1.

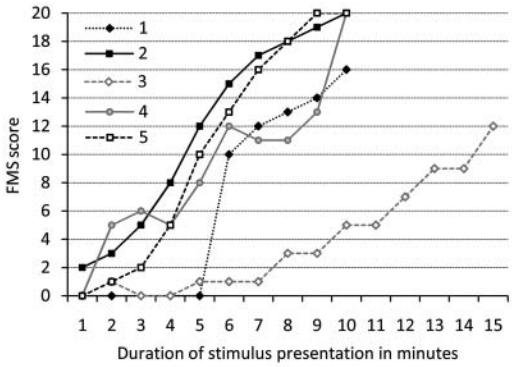

Figure 3. Time courses of 5 individual participants who aborted Experiment 1 prior to stimulus offset because of severe nausea.

for peak FMS scores and SSQ values, including the SSQ-TS and the subscores SSQ-N, SSQ-D, and SSQ-O. Peak FMS scores were highly correlated with the SSQ-TS as well as with all SSQ subscores ($p < .001$). Table 3 lists the correlations not only for the peak FMS scores but also for the very last FMS scores measured immediately at the end of the stimulus presentation.

DISCUSSION

One aim of our study was to introduce and evaluate a simple and effective measure of MS that could be taken repeatedly throughout the

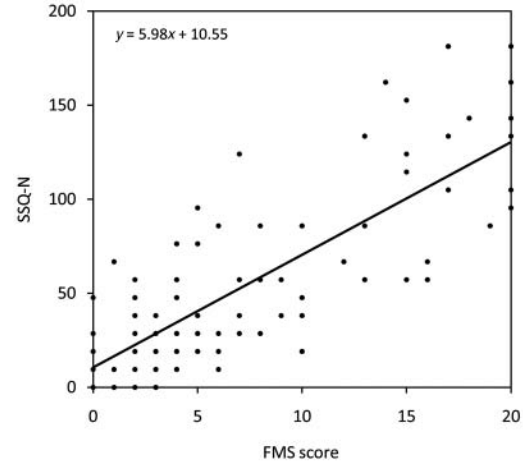

Figure 4. Scatter plot that shows the distribution of the peak Fast Motion Sickness Scale (FMS) score and the Simulator Sickness Questionnaire Nausea subscore (SSQ-N) for all participants. The regression line is included.

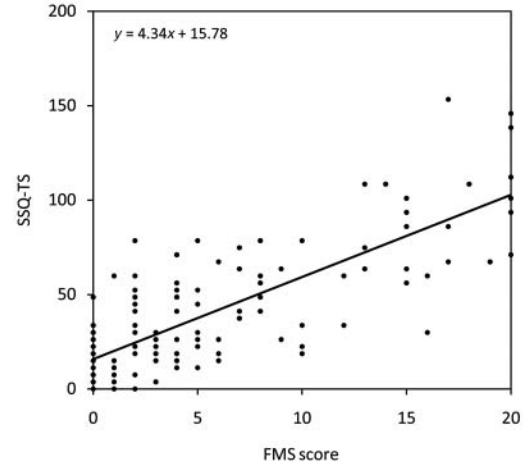

Figure 5. Scatter plot that shows the distribution of the peak Fast Motion Sickness Scale (FMS) score and the Simulator Sickness Questionnaire total score (SSQ-TS) for all participants. The regression line is included.

ongoing exposure to a provocative stimulus. If successful, such a measure would have many advantages to the existing questionnaires that quantify MS after stimulus presentation. On the basis of the rating scale used by Young et al. (2003), who used a 20-point scale ranging from 0 (*no sickness at all*) to 20 (*frank sickness*), we

TABLE 3: Bivariate Correlations Between Reported Peak FMS Scores and Very Last FMS Score Measured During Stimulus Exposure and SSQ Values

	SSQ-TS	SSQ-N	SSQ-D	SSQ-O
Peak FMS score	.785	.828	.795	.609
Last FMS score	.765	.815	.770	.578

Note. FMS = Fast Motion Sickness Scale; SSQ = Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993); TS = total score; N = Nausea subscale; D = Disorientation subscale; O = Oculomotor Issues subscale.

had observers verbally judge the severity of MS once per minute. For validation purposes, we compared this FMS to one of the most common MS questionnaires, the SSQ (Kennedy et al., 1993). Our second aim was to explore the influence of potential expectancy effects on the reliability of MS measurement. On the basis of previous findings (Young et al., 2007), we tested the assumption that simply expecting sickness to occur during stimulus presentation might lead to stronger MS symptoms. In the following, we discuss issues of validation, expectancy, and finally, the strengths and weaknesses of the FMS.

Validation of the FMS

The FMS proved to be an easy-to-administer tool. Our results revealed significant and high correlations between the verbal FMS ratings and the SSQ scores. The highest correlation was found between the peak FMS rating and the score on the SSQ subscale of Nausea ($r = .828$), which is in perfect concordance with our study's goal to find a new way to measure sickness in terms of nausea, discomfort, and stomach awareness. The total SSQ score reached a correlation of $r = .785$ with the FMS scores, which proves the strong relation between these two parameters. Somewhat lower correlations were found between FMS ratings and the SSQ subscales Disorientation ($r = .795$) and Oculomotor Issues ($r = .608$). This result was to be expected because we explicitly asked our participants to rate their sickness and disregard other adverse symptoms, such as nervousness, boredom, or fatigue.

Given the simplicity and robustness of the online ratings, it seems surprising in retrospect that such scales have not received more attention in the long tradition of MS research. Only a few studies have made use of verbal MS ratings during stimulus exposure, and hardly any of them have provided a validation of the verbal scale with other MS measurement tools. Unfortunately, the particular rating scales that have been used varied dramatically from study to study. Stoffregen, Faugloire, Yoshida, Flanagan, and Merhi (2008) simply asked their participants whether they felt sick or not using a yes-no dichotomization. They decided to classify their participants into a sick and a not-sick group to satisfy their particular needs of statistical analysis and thus did not collect quantitative ratings.

A 4-point MS scale (1 = *no symptoms*, 2 = *mild symptoms but no nausea*, 3 = *mild nausea*, 4 = *moderate nausea*) was first used by Bagshaw and Stott (1985) to analyze MS severity of aircrew members. Draper, Viirre, Furness, and Gawron (2001) used a similar 4-point scale (0 = *no discomfort*, 1 = *slight but noticeable discomfort*, 2 = *moderate discomfort*, 3 = *very strong discomfort*) to measure SS during stimulus exposure. These 4-point rating scales have proven to be very useful in monitoring MS. However, frank nausea may not be adequately represented. In another study, a 5-point rating scale was implemented to measure MS at short time intervals (McCauley, Royal, Wylie, O'Hanlon, & Mackie, 1976). Although this coarse scale can differentiate between mild, moderate, and severe MS, its resolution restricts the ability for more fine-grained assessments of MS.

A similar approach was chosen by Griffin and Newman (2004). They adopted a 7-point scale starting with 0 (*no symptoms*) and ending with the score 6 (*moderate nausea but want to stop*). In another study, Bos, de Vries, van Emmerik, and Groen (2010) introduced the so-called Misery Scale Index. This scale ranges from 0 (*no problems*) to 10 (*vomiting*), and every gradation on it represents the presence of particular MS symptoms. The scale starts with *uneasiness* (score 1); continues with *dizziness*,

warmth, *headache*, and *stomach awareness* (*vague*, *slight*, *fairly*, or *severe*, corresponding to scores 2 to 5); and then reaches *nausea* (*slight*, *fairly*, or *severe*, corresponding to scores 6 to 8); and finally ending in *retching* (score 9) and *vomiting* (score 10).

In our eyes, two things make this scale unfit for a general-purpose MS scale. First, training is required to ensure that participants apply the scores correctly. Second, symptom-based criteria are conflated with magnitudes. For instance, extreme dizziness would receive a relatively low score. And slight nausea does not necessarily include feelings of dizziness, headache, or warmth, as it is suggested by the scale. The explicit attachment of scores to symptoms makes it a special-purpose scale.

Another method for quantifying MS during immersion has been introduced by Nichols, Cobb, and Wilson (1997) under the label of Short Symptom Checklist (SSC). The SSC represents a short version of the SSQ by Kennedy et al. (1993) and contains six items in total (two for each subscale). The rating scale—in contrast to one used in the SSQ—ranges in five steps from *not at all* to *severe*. The authors state that the SSC can be adopted during virtual reality exposure and is able to reflect the participant's sickness profile during stimulus presentation. Nichols et al. required their observers to fill out the SSC every 5 min to map the time course of MS.

We believe that although the FMS varies little from the SSC, it does so on a few crucial points. First, the FMS is more finely graded: It offers a 20-step scale instead of the 5-point range. Second, the FMS is faster and more convenient because participants are simply asked for a single value instead of required to fill in a short questionnaire. The FMS can thus sample MS more often and without interference. Third, and most importantly, the SSC scores reveal lower correlations with the SSQ values gathered after immersion (similar to the procedure we used to validate the FMS). Nichols et al.'s values correlated between $r = .61$ and $r = .71$; the correlations found in our experiments between the FMS and the SSQ were higher ($r = .785$ and above)—except for the subscale Oculomotor Issue ($r = .61$).

Notwithstanding the success of the validation, there are some caveats associated with the validation method we have used. As previously described, we correlated the participants' peak verbal ratings during stimulus presentation with the SSQ subscores and the SSQ total score after stimulus offset. Strictly speaking, our procedure precluded a minute-by-minute validation. This short time interval prohibits the assessment of SSQ values at the same sampling rate as the verbal score. Instead, we validated the verbal score using the peak FMS score obtained in the SSQ. In our eyes, it would not have made sense to extend the sampling rate and to present a questionnaire, such as the SSQ, every few minutes, because this procedure would distract the participants too much from their task to focus on the stimuli. Additionally, this would lead to a disruptive stimulus presentation and would likely interfere with the genesis of MS in the first place.

An alternative validation process using psychophysiological online measurements, such as electrodermal activity or heart rate variability, is in principle possible and desirable. However, there is no single valid and reliable psychophysiological parameter that represents nausea or MS in an extensive way (Shupak & Gordon, 2006). In other words, the existing physiological measures only partially represent MS correlates. Although some very promising approaches have been started in the past years, further research is necessary before a measure can emerge as a suitable tool for a physiological validation method.

Another problematic issue is that we validated the peak FMS score obtained during the stimulus exposure using the SSQ score obtained after stimulus exposure. This temporal lag might have caused some memory distortions, although the long time constant of symptom decay makes this less likely. On the basis of our data, we believe that the time-lag aspect is indeed negligible with respect to the current study for two reasons. First, the time courses (see Figure 3 and Figure 4) indicate that the peak MS scores were obtained toward the end of the experiment or—after the FMS scores had reached their peak—remained stable until stimulus exposure had ended (except for the few cases in which the experiment was aborted).

Second, we met the concerns of memory effects by additionally correlating the very last FMS score (instead of the peak FMS score) with the SSQ ratings. Doing so, we found comparably high correlations, similar to the ones achieved with the peak FMS score (see Table 2). We therefore believe that the validation with the retrospective SSQ measurement is unlikely to be influenced by memory effects.

Expectancy effects. Young and colleagues (2007) focused on the role of demand characteristics, such as expectancy effects, in MS research. On the basis of their findings, the authors assumed that the very fact of handing out the SSQ to the participants as a pretest could have increased the sickness scores reported after stimulus offset. In this case, the SSQ should not be used before and after the stimulus exposure. Kennedy et al. (1993) did in fact advise against using the SSQ both as a pre- and postmeasurement tool within the same experiment. This advice has often gone unheeded.

We did not find any effects of expectancy in our study. Instructing observers with positive, negative, or neutral expectations did not significantly change FMS ratings. Neither did the groups in SSQ scores obtained after exposure. How can this positive but unexpected result be explained? One possibility may lie in method limitations. Simply briefing the participants might not have been powerful enough to induce the desired expectations in our participants. We do not believe that this was the case, because we addressed the participants in a very direct manner: We presented written instructions to the observers in each group and told them explicitly to expect either positive (happiness, joy) or negative (nausea, discomfort) feelings from the stimuli. Ultimately, replication studies are desirable that maximize expectation by wearing lab coats, distributing nausea bags, and so on.

The expectancy effects found by Young et al. (2007) might have been caused by administering the SSQ as a pretest. We agree with the authors that the SSQ might sensitize the participants toward the relevant symptoms during immersion and therefore might lead to higher ratings in the posttest. Or participants may have thought that they were supposed to report stronger feelings of MS after stimulus presentation than before because of social desirability. However, a clear reason for the discrepancy between Young et al.'s results and ours cannot be given. Further research is required to draw a final conclusion on this issue.

Strengths and Weaknesses of the FMS

An easy-to-administer rating scale has many advantages and some disadvantages compared with more detailed post hoc questionnaires, such as the SSQ. A unitary rating scale enhances the comparability of results found in different studies and would enable direct comparisons of MS severity and time courses throughout many researches. We presented the FMS as a quick and simple tool to measure nausea and general discomfort in terms of MS.

The FMS is easy to administer and is thus ideally suited to capture the time course of MS, providing information about the onset, course, severity, and the trend of MS, which are not measureable easily with questionnaires, such as the SSQ. Furthermore, the current experiment shows that the FMS is robust against expectancy effects and can be assessed straight from the beginning of the stimulus exposure.

On the basis of our previous experience (e.g., Young et al., 2003), we decided to set the rating scale's span to 20 points. Compared with scales of narrower ranges below 10 points, the FMS is able to detect slight changes in the severity of MS while the scale continues to be manageable for the observers. Note that there is a long tradition in the search for the optimal number of response alternatives in questionnaires (see Cox, 1980). To discuss this issue at large would go beyond the scope of this article, because the question about the optimal number of alternatives is still unsolved in general terms, and several studies have revealed incoherent results (see Preston & Coleman, 2000, for an overview).

In sum, Cox (1980) reported that seven response categories seem to be the optimal number of choices for many rating scales. However, there is also evidence that rating scales with more gradations provide an increase in information gain (Garner, 1960). Preston and Coleman (2000) systematically analyzed different rating scales (including scales with 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, and 101 response alternatives). Test-retest

reliability was comparable for scales with more than 5 points (with a slight decrease in the 101-point scale), and Cronbach's α increased as the number of response alternatives increased. Importantly, the participants' ability to express their feelings increased with rating scales with larger ranges.

In our study, we chose a 20-point scale for the FMS for several reasons: Our premier intent was to use an MS scale that is able to differentiate among lower degrees of MS and thus to capture different states of well-being as well as states of sickness. We believe that the low resolution of a classical 5- or 7-point scale would not offer this possibility. With regard to Preston and Coleman's (2000) arguments, a shorter rating scale is indeed easier to use than a longer one, but participants can express their feelings more precisely when they are offered more than seven response alternatives. In another approach, Bock and Oman (1982) tried to apply a ratio scale in the style of Stevens' (1951) magnitude estimation in psychophysics to measure MS. They asked their participants to label the mean severity of sickness with the score 10 and to respond to this anchor in the subsequent part of the experiment. Doing so, Bock and Oman showed that participants were able to judge sickness quite accurately using this ratio scale.

Although the FMS varies slightly from Bock and Oman's (1982) scale (the FMS uses two fixed anchors and therefore does not necessarily represent a ratio scale), we followed the authors' approach and chose the score 10 as the midpoint for the FMS. Furthermore, the feedback regarding the usability and practicability of the FMS was almost consistently positive in our experiments. With few exceptions, participants felt comfortable and competent to report nausea values using the FMS, as they stated in personal notes after the experiment was terminated. It is nonetheless conceivable that a narrowed rating scale would have shown similar performance as the 20-point scale, but our results indicate that the scale we chose met the requirements. Further research should attempt to optimize the FMS with regard to its anchoring and the number of response alternatives.

Besides the benefits of the FMS, our rating scale has some general limitations. First, MS is a genuinely subjective phenomenon that has no perfect physiological correlate but is nonetheless associated with oculomotor changes, pallor, body temperature, or even postural instability (Kennedy & Stanney, 1996; Riccio & Stoffregen, 1991). FMS is indifferent to the physiological correlates or causes of MS. It is also unable to differentiate between feelings of nausea and its precursors, such as drowsiness, which may be a syndrome in its own right (see, e.g., Graybiel & Knepton, 1976, on the sopite syndrome). Thus, the FMS reaches its limits whenever particular physiological symptoms become the focus of interest.

The FMS was designed to quantify the subjective aspects of nausea and general discomfort in MS. We hold that the latter are captured in a very meaningful and precise way by a unidimensional scale. Although our results indicate high correlations of FMS scores with the SSQ subscales and even the SSQ total score, we suggest further studies to finally validate the FMS in a more extensive manner. Given that the FMS is a practicable and easy-to-administer tool, prospective studies could add the FMS measure to any existing experimental design with minimal cost.

We have applied the FMS in studies dealing with visually induced MS. Because we have previously used the FMS successfully to measure vestibular MS (in observers making head turns while being centrifuged in darkness), we believe that the FMS is applicable to more varieties of MS, including SS, cybersickness, or seasickness. The FMS measures two of the cardinal symptoms in MS, namely, nausea and general discomfort. Although the symptomatology slightly varies between the subcategories of MS (Kennedy, Stanney, & Drexler, 1997), nausea is represented in all types of MS. However, further data in the manifold types of MS are essential to confirm and reinforce this assumption. We hope that researchers will use the FMS to verify this prediction.

Taken together, our results indicate that the solution to the problem of designing a general-purpose MS scale could lie in a frugal one-dimensional self-assessment. To our knowledge, there is no other unidimensional MS scale that basically follows the same recipe and that has

been validated by comparing its results with additional measurements, such as the SSQ. A continued validation of the scale is desirable. The FMS scoring system, which solicits quick self-assessments on a 0-to-20 scale, showed high correlations with the SSQ total score obtained after stimulus exposure and particularly high correlations with the SSQ Nausea subscale. Thus, we recommend the use of this or a similar rating scale as a general and fast tool to assess MS.

ACKNOWLEDGMENTS

We thank Chuck Oman for many inspiring discussions about motion sickness. Agnes Münch and Markus Homberg advised on technical matters and stimulus material, and Lisa Zschutschke helped collecting the data.

KEY POINTS

- We introduced a simple and easy way to measure motion sickness: the FMS (Fast Motion Sickness Scale).
- The FMS was validated by correlating the results of both the FMS and the Simulator Sickness Questionnaire (SSQ).
- 2 experiments with a total $N = 126$ were run.
- Analyses revealed high correlations (r ranging from .785 to .828) between the FMS and the SSQ.
- The FMS is a valid and reliable tool for measuring motion sickness.

REFERENCES

Bagshaw, M., & Stott, J. R. (1985). The desensitization of chronically motion sick aircrew in the Royal Air Force. *Aviation, Space, and Environmental Medicine*, *56*, 1144–1151.

Bles, W., Bos, J. E., Graaf, B. de, Groen, E., & Wertheim, A. H. (1998). Motion sickness: Only one provocative conflict? *Brain Research Bulletin*, *47*, 481–487.

Bock, O. L., & Oman, C. M. (1982). Dynamics of subjective discomfort in motion sickness as measured with a magnitude estimation method. *Aviation, Space, and Environmental Medicine*, *53*, 773–777.

Bos, J. E., Bles, W., & Groen, E. L. (2008). A theory on visually induced motion sickness. *Displays*, *29*, 47–57.

Bos, J. E., Vries, S. C. de, van Emmerik, M. L., & Groen, E. L. (2010). The effect of internal and external fields of view on visually induced motion sickness. *Applied Ergonomics*, *41*(4), 516–521.

Brown, E. L., Hecht, H., & Young, L. R. (2002). Sensorimotor aspects of high-speed artificial gravity: 1. Sensory conflict in vestibular adaptation. *Journal of Vestibular Research—Equilibrium and Orientation*, *12*, 271–282.

Cox, E. P. (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research*, *17*, 407–422.

Draper, M. H., Viirre, E. S., Furness, T. A., & Gawron, V. J. (2001). Effects of image scale and system time delay on simulator sickness within head-coupled virtual environments. *Human Factors*, *43*, 129-146.

Flanagan, M. B., May, J. G., & Dobie, T. G. (2004). The role of vection, eye movements and postural instability in the etiology of motion sickness. *Journal of Vestibular Research—Equilibrium and Orientation*, *14*, 335–346.

Frank, L. H., Casali, J. G., & Wierwille, W. W. (1988). Effects of visual display and motion system delays on operator performance and uneasiness in a driving simulator. *Human Factors*, *30*, 201–217.

Garner, W. R. (1960). Rating scales, discriminability and information transmission. *Psychological Review*, *67*, 343–352.

Gianaros, P. J., Muth, E. R., Mordkoff, J. T., Levine, M. E., & Stern, R. M. (2001). A questionnaire for the assessment of the multiple dimensions of motion sickness. *Aviation, Space, and Environmental Medicine*, *72*, 115–119.

Graybiel, A., & Knepton, J. (1976). Sopite syndrome: A sometimes sole manifestation of motion sickness. *Aviation, Space, and Environmental Medicine*, *47*, 873–882.

Griffin, M. J., & Newman, M. M. (2004). Visual field effects on motion sickness in cars. *Aviation Space and Environmental Medicine*, *75*, 739–748.

Kennedy, R. S., Drexler, J. M., & Kennedy, R. C. (2010). Research in visually induced motion sickness. *Applied Ergonomics*, *41*, 494–503.

Kennedy, R. S., Lane, N. E., Berbaum, K. S, & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *International Journal of Aviation and Space Psychology*, *3*, 203–220.

Kennedy, R. S., & Stanney, K. M. (1996). Postural instability induced by virtual reality exposure: Development of a certification protocol. *International Journal of Human-Computer Interaction*, *8*, 24–47.

Kennedy, R. S., Stanney, K. M., & Drexler, J. M. (1997). Cybersickness is not simulator sickness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 1138–1142). Santa Monica, CA: Human Factors and Ergonomics Society.

McCauley, M.E., Royal, J. W., Wylie, C. D., O'Hanlon, J. F., & Mackie, R. R. (1976). *Motion sickness incidence: Exploratory studies of habituation, pitch and roll, and the refinement of a mathematical model*. (Tech. Rep. HFR 1733-2). Santa Barbara, CA: Human Factors Research.

Merhi, O., Faugloire, E., Flanagan, M., & Stoffregen, T. A. (2007). Motion sickness, console video games, and head-mounted displays. *Human Factors*, *49*, 920–934.

Muth, E. R., Stern, R. M., Thayer, J. F., & Koch, K. L. (1996). Assessment of the multiple dimensions of nausea: The Nausea Profile (NP). *Journal of Psychosomatic Research*, *40*, 511–520.

Nichols, S., Cobb , S., Wilson, J. R. (1997). Health and safety implications of virtual environments: Measurement issues. *Presence*, *6*, 667–675.

Oman, C. M. (1990). Motion sickness: A synthesis and evaluation of the sensory conflict theory. *Canadian Journal of Physiology and Pharmacology*, *68*, 294–303.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*, 1–15.

Reason, J. T. (1978). Motion sickness adaptation: Neural mismatch model. *Journal of the Royal Society of Medicine*, *71*, 819–829.

Reason, J. T., & Brand, J. J. (1975). *Motion sickness*. London, UK: Academic Press.

Riccio, G. E., & Stoffregen, T. A. (1991). An ecological theory of motion sickness and postural instability. *Ecological Psychology*, *3*, 195–240.

Sharples, S., Cobb, S., Moody, A., & Wilson, J. R. (2008). Virtual reality induced symptoms and effects (VRISE): Comparison of head mounted display (HMD), desktop and projection display systems. *Displays*, *29*, 58–69.

Shupak, A., & Gordon, C. R. (2006). Motion sickness: Advances in pathogenesis, prediction, prevention, and treatment. *Aviation, Space, and Environmental Medicine*, *77*, 1213–1223.

Stevens, S. S. (1951). *Handbook of experimental psychology.* New York, NY: Wiley.

Stoffregen, T. A., Faugloire, E., Yoshida, K., Flanagan, M. B., & Merhi, O. (2008). Motion sickness and postural sway in console video games. *Human Factors*, *50*, 322–331.

Stoffregen, T. A., Hettinger, L. J., Haas, M. W., Roe, M. M., & Smart, L. J. (2000). Postural instability and motion sickness in a fixed-base flight simulator. *Human Factors*, *42*, 458–469.

Warwick-Evans, L. A., Symons, N., Fitch, T., & Burrows, L. (1998). Evaluating sensory conflict and postural instability: Theories of motion sickness. *Brain Research Bulletin*, *47*, 465–469.

Young, L. R., Sienko, K. H., Lyne, L.E., Hecht, H., & Natapoff, A. (2003). Adaptation of the vestibulo-ocular reflex, subjective tilt, and motion sickness to head movements during short-radius centrifugation. *Journal of Vestibular Research-Equilibrium and Orientation*, *13*, 65–77.

Young, S. D., Adelstein, B. D., & Ellis, S. R. (2007). 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*, 422–428.

Behrang Keshavarz received a master's degree in cognitive psychology from the Johannes Gutenberg-Universität of Mainz, Germany, where he is currently pursuing a PhD degree. His dissertation is concerned with visually induced motion sickness. In addition, he is involved in research on distance and space perception as well as time-to-contact estimation.

Heiko Hecht studied psychology and philosophy at the Universities of Trier, Germany, and Virginia, United States. After his dissertation (experimental psychology), he worked as a research scientist and lecturer at the Max-Planck Institut für Psychologische Forschung, NASA Ames Research Center, Universität Bielefeld, and the Massachusetts Institute of Technology. Since 2002, Heiko Hecht holds the chair of experimental psychology at the Johannes Gutenberg-Universität of Mainz, Germany. He has published in the domains of picture perception and virtual reality, artificial gravity, time-to-contact estimation, and intuitive physics.

Date received: July 12, 2010
Date accepted: February 15, 2011