The International Journal of Aviation Psychology

ISSN: 1050-8414 (Print) 1532-7108 (Online) Journal homepage: www.tandfonline.com/journals/hiap20

Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness

Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum & Michael G. Lilienthal

To cite this article: Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum & Michael G. Lilienthal (1993) Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness, The International Journal of Aviation Psychology, 3:3, 203-220, DOI: 10.1207/s15327108ijap0303_3

To link to this article: https://doi.org/10.1207/s15327108ijap0303_3

Published online: 13 Nov 2009.

THE INTERNATIONAL JOURNAL OF AVIATION PSYCHOLOGY, 3(3), 203–220

Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness

Robert S. Kennedy and Norman E. Lane
Essex Corporation
Orlando, Florida

Kevin S. Berbaum
The University of Iowa

Michael G. Lilienthal
United States Navy
Washington, DC

Simulator sickness (SS) in high-fidelity visual simulators is a byproduct of modern simulation technology. Although it involves symptoms similar to those of motion-induced sickness (MS), SS tends to be less severe, to be of lower incidence, and to originate from elements of visual display and visuo-vestibular interaction atypical of conditions that induce MS. Most studies of SS to date index severity with some variant of the Pensacola Motion Sickness Questionnaire (MSQ). The MSQ has several deficiencies as an instrument for measuring SS. Some symptoms included in the scoring of MS are irrelevant for SS, and several are misleading. Also, the configural approach of the MSQ is not readily adaptable to computer administration and scoring. This article describes the development of a Simulator Sickness Questionnaire (SSQ), derived from the MSQ using a series of factor analyses, and illustrates its use in monitoring simulator performance with data from a computerized SSQ survey of 3,691 simulator hops. The database used for development included more than 1,100 MSQs, representing data from 10 Navy simulators. The SSQ provides straightforward computer or manual scoring, increased power to identify "problem" simulators, and improved diagnostic capability.

Requests for reprints should be sent to Robert S. Kennedy, Essex Corporation, Suite 227, 1040 Woodcock Road, Orlando, FL 32803.

The Pensacola Motion Sickness Questionnaire (MSQ) was developed about 25 years ago (Kellogg, Kennedy, & Graybiel, 1965; Kennedy, Tolhurst, & Graybiel, 1965) to study motion sickness (MS). The MSQ was originally based on the rating schemes of Wendt and his colleagues (Alexander, Cotzin, Klee, & Wendt, 1947) and Hemingway (1942), and assigns numbers to the degree of MS severity when symptoms terminate short of actual vomiting (emesis). It consists of a list of (usually) 25 to 30 symptoms associated with or premonitory of MS onset; for each symptom, an individual indicates its presence and/or degree of severity. The symptoms are then converted to scores on a scale ranging from no symptoms (0) to confirmed emesis (highest score). Intermediate scores are assigned according to the number, type, and severity of symptoms.

The current MSQ scoring was developed over the course of numerous studies conducted in a variety of MS-inducing settings. In most of these studies, individuals were exposed to stimuli sufficiently severe to induce emesis or near-emesis, or to occasion them to request release from the experiment. The symptoms reported in the course of the experiments were compared to the severity of outcomes. Those symptoms that predicted various degrees of malaise were combined into clusters or patterns such that higher numbers on the MSQ scale were associated with greater likelihood of emesis as the stimulation continued. Clusters and scale points were constructed using a combination of quantitative prediction methods and clinical judgment.

THE NEED FOR A SIMULATOR SICKNESS QUESTIONNAIRE (SSQ)

The objective of more recent MSQ scoring research has been to provide a scale to indicate the onset of MS under less severe conditions of stimulation. As such, the more recent MSQ scoring approach was the obvious method of choice in the studies of simulator sickness initiated in the early 1980s (Frank, Kennedy, Kellogg, & McCauley, 1983). Symptoms of simulator sickness are often similar to those of motion sickness but affect a smaller proportion of the exposed population and are usually much less severe. This is not surprising because current flight simulation technology involves stimulus conditions that produce pseudo-Coriolis, visual distortion, and visual/motion transport delays and asynchronies that are ordinarily absent in environments that produce sea and air sickness. (The pseudo-Coriolis illusion is the experience of an anomalous angle of head tilt and tilting of the visual stimulus that occurs when an observer moves his head across the plane of rotation of an optokinetic stimulus that would otherwise produce vection; Dichgans & Brandt, 1973.) Simulator sickness itself can also produce different symptom clusters as a result of intersimulator differences (Gower, Lilienthal, Kennedy, & Fowlkes, 1987; Kennedy et al., 1987; Ungs, 1987). Moreover, fixed-base flight simulators differ in a critical respect from ships at sea: namely, when one closes one's eyes in a simulator, the stimulus stops.

There are also some other differences between MS and simulator-induced

sickness that make the MSQ a less-than-ideal index of SS. Some of the symptoms in the present questionnaire are almost never reported under simulator exposure, or, if noted, are indicated at a level that fails to exceed a base level. (For each symptom, there is a base frequency with which that symptom will be checked by normal, healthy subjects who are not being exposed to any unusual stimulation at all.) Some of the symptoms that are valid for scoring MS are not necessarily appropriate for SS assessment. Drowsiness, for example, is a key indicator of the onset of motion-induced sickness (Graybiel & Knepton, 1976). It is also, however, a frequently indicated symptom in simulators for which there are no other reported symptoms. Drowsiness reported alone may indicate simple sleepiness in response to a tiresome simulator exercise, but its meaning may be quite different when followed by other symptoms suggestive of a parasympathetic reaction to powerful motion stimulation. The configural approach of the MSQ is not readily adaptable to computer administration and scoring. This deficiency is particularly important because control of SS may require routine on-site monitoring of symptoms to detect progressive calibration error or mechanical problems. Continuous quality control tracking requires more powerful and convenient scoring methods than those provided by the MSQ.

We have three major objectives: (a) to provide a more valid index of overall *simulator* sickness severity as distinguished from *motion* sickness; (b) to provide subscale scores that are more diagnostic of the locus of simulator sickness in a particular simulator for which overall severity was shown to be a problem; and (c) to provide a scoring approach to make monitoring and cumulative tracking relatively straightforward.

DEVELOPMENT OF THE SSQ SCORING SYSTEMS

Method

The data set used for development of the new SSQ consisted of 1,119 pairs (pre- and postexposure) of MSQs collected previously during on-site studies of 10 simulator sites (Baltzley, Kennedy, Berbaum, Lilienthal, & Gower, 1989; Kennedy, Lilienthal, Berbaum, Baltzley, & McCauley, 1989). The MSQ version used in these studies contained the 28 symptoms shown in Table 1.

Identification and description of the simulators studied and the general properties and composition of the data set are described elsewhere (Kennedy et al., 1989). Because our purpose was to determine which symptoms showed systematic changes from pre-exposure to postexposure, symptoms selected too infrequently to be of value as statistical indicators (i.e., with less than 1% frequency) and symptoms that showed no change in frequency or severity were eliminated from further analyses. Vomiting, for example, is clearly an important sign of MS and SS, but occurred only twice in approximately 1,200 simulator exposures—too low a rate for correlations and other statistical data to be stable. Symptoms that might give misleading indications were also eliminated from

TABLE 1
Symptoms in MSQ and SSQ

MSQ Symptom	*Retained for SSQ*	*Eliminated for SSQ*
General discomfort	X	
Fatigue	X	
Boredom		X
Drowsiness		X
Headache	X	
Eyestrain	X	
Difficulty focusing	X	
Increased salivation	X	
Decreased salivation		X
Sweating	X	
Nausea	X	
Difficulty concentrating	X	
Depression		X
Fullness of head	X	
Blurred vision	X	
Dizziness (eyes open)	X	
Dizziness (eyes closed)	X	
Vertigo	X	
Visual flashbacks		X
Faintness		X
Awareness of breathing		X
Stomach awareness	X	
Decreased appetite		X
Increased appetite		X
Desire to move bowels		X
Confusion		X
Burping	X	
Vomiting		X

subsequent analyses. These symptoms (e.g., boredom) had their highest frequency of occurrence in simulators that had little or no other indicated symptomatology, and were rarely seen in simulators that had high frequency or severity on most other symptoms. Altogether, 12 of the 28 symptoms were eliminated; these are identified in Table 1.

Most studies involving the MSQ use differences between post and pre scores as the main indicator of problem severity. However, difference scores have poor reliability (Cronbach & Furby, 1970). It is also well known that illness increases MS susceptibility thresholds (DeWit, 1957; Kellogg et al., 1965). As part of MSQ administration, a pre-exposure checklist was employed in which respondents were asked if they were "sick" or in other than their "usual state of fitness." When respondents who gave positive answers to either of these two questions were dropped, there was very little variance remaining in the pre-exposure data. Records from subjects who reported themselves as "other than healthy" were excluded from analysis. The SSQ scoring system re-

ported on herein is intended only for application to postexposure symptoms, with the further precondition that a screening of "unhealthy" subjects is required.

Factor Analyses

Factor-analytic models of symptom subgroups within the questionnaire were studied to provide a basis for scoring. Simulator sickness may be due to various stimuli, including linear oscillation at 0.2 Hz, vection, visual distortion, flicker, conflict among oculomotor systems, and cue asynchrony. Not surprisingly, a variety of symptoms may result. Coincidence or clusters of symptoms can be identified by factor analysis. Two forms of factor analysis were used. The "conventional" method was a principal-factors analysis, iterated until communalities stabilized, followed by normalized varimax rotation. This approach produces factors that—although theoretically orthogonal (independent)—will be correlated whenever all symptoms share at least some variance in common. That is, there is a "general" factor present on which all variables have "real" nonzero loadings. For purposes of diagnostic use of a scale, it may be desirable to have subscales that are as independent (i.e., clearly a measure of a single component) as possible. To determine the presence and magnitude of the general factor, the hierarchical factor-analysis method (Wherry, 1984) was used. This method extends the analysis of the rotated-factor matrix to extract a general factor (if there is one) and two or more group factors. (Group factors are those on which some subset of variables will have large loadings whereas other variables will have loadings near zero.) Group factors obtained in this way are typically cleaner (i.e., much less correlated) than those obtained from varimax rotation. Results of these analyses are summarized hereafter.

Results

Principal factors analysis/varimax. Analyses were conducted extracting three-, four-, five-, and six-factor solutions from the 16 symptom variables. The three-factor solution was most readily interpretable. The three distinct symptom clusters were labeled *Oculomotor (O;* eyestrain, difficulty focusing, blurred vision, headache), *Disorientation (D;* dizziness, vertigo), and *Nausea (N;* nausea, stomach awareness, increased salivation, burping). The factor matrix is given in Table 2. Each of the three factors was used as the basis for an SSQ subscale.

Factor-analytic results using variants of the MSQ symptoms in related domains have yielded similar results. In a study of visual display unit (VDU) users after 3-hr sessions, Morrissey and Bittner (1986) found dimensions that correspond closely to ours. Bittner and Guignard (1988) produced similar symptom clusters in a study of MS symptoms at sea. In these studies, there were invariably a visual factor and a nausea factor, and usually a factor of dizziness, disorientation, blurred vision, and/or sweating. Discrepancies in factor patterns across these analyses are remarkably small given the differences in the stimulus domains.

TABLE 2
Varimax Factors From SSQ Symptoms

	Varimax Factors			
SSQ Symptom	*N*	*O*	*D*	η^2
General discomfort	.65	.40	.18	.62
Fatigue	.15	.54	−.04	.32
Headache	.22	.53	.15	.35
Eyestrain	.00	.74	.17	.58
Difficulty focusing	−.01	.61	.43	.56
Increased salivation	.53	.21	.13	.34
Sweating	.31	.24	.08	.16
Nausea	.75	.08	.30	.66
Difficulty concentrating	.32	.39	.27	.33
Fullness of head	.12	.17	.37	.18
Blurred vision	.01	.36	.40	.29
Dizzy (eyes open)	.17	.07	.76	.60
Dizzy (eyes closed)	.17	.09	.65	.46
Vertigo	.18	.08	.37	.17
Stomach awareness	.64	.03	.21	.45
Burping	.41	.04	.22	.22
Eigenvalue	2.21	2.11	1.98	
Percent of variance	14	13	12	

The three-factor solution suggested the existence of three (partially) independent symptom clusters, each reflecting the impact of simulator exposure on a different "target system" within the human. A given simulator may cause symptoms that fall into none, one or more, or all of these clusters, depending on the mechanism or mechanisms by which the human is affected. This target-system organization of symptoms has both theoretical and practical importance. It may eventually be useful in studying the physiological basis of the reported symptoms; it likewise simplifies the process of determining where and in what ways a simulator may be causing problems for the user.

The four-, five-, and six-factor solutions, although less useful for scoring, were also of some interest in that they suggested the capability of respondents to make more fine-grained distinctions among their feelings following simulator exposure. The four-factor solution split the Oculomotor factor into two separate factors—one concerned with the disturbance of visual processing during the simulation (blurred vision, difficulty focusing) and the other with the symptoms caused by that disturbance (headache, eyestrain, fatigue). The five-factor solution consisted of the same four factors with the addition of what appeared to be a "tired and hungry" factor (fatigue, difficulty concentrating), almost certainly an artifact created by the passage of time during the simulation. The six-factor solution split the Nausea factor into two parts—one reflecting the premonitory signs of nausea (increased salivation, burping) and the other reflecting the advanced stages of the process (nausea, sweating). These additional solutions are not nearly as well defined as the

three-factor solution, and, more important, there are too few simulator-relevant symptoms in the present MSQ/SSQ to provide adequate reliability for a subscale score based on the additional factors. The potential for an expanded SSQ (i.e., more symptom choices) is discussed later.

The 10 simulators on which data were collected (for a list, see the Interpreting the SSQ Scores section hereafter) divide themselves conveniently into two distinct groups—five simulators with little or no reported symptomatology and five with a markedly higher level of reported symptoms. The approximately 500 observations from the low-incidence group contributed only modestly to variance and even less to the covariance among variables, and thus acted to reduce the absolute level of correlations among the symptoms. Because factor patterns are more clearly defined for higher correlation values, analyses were repeated using only the approximately 600 observations from the five high-incidence simulators. Although there were no changes in factor structure and interpretation, there was noticeable improvement in factor definition within that structure (i.e., large loadings became larger and small loadings became smaller). Accordingly, data from the high-incidence simulators (e.g., the 2F64C) were used in subsequent analyses, and the rotated loadings reported in Table 2 (and in later tables) are based on that portion of the sample.

Hierarchical. Inspection of the varimax factor matrix in Table 2 shows that, for each factor, there are several variables with moderate to large loadings (.50 to .75) and about an equal number with loadings in the range of .15 to .35. In addition, there are a number of variables that have substantial loadings on at least two and sometimes all three of the factors. The ideal pattern for such a matrix would be one in which every factor has a few very large loadings, with most of the other loadings near zero, and every variable has a large loading on only one factor, with its loadings on the other factors near zero. The pattern in Table 2 departs sharply from the ideal; it is highly characteristic of a varimax-rotated factor structure in which every variable contains at least some variance in common with every other variable (i.e., there is some general factor underlying the structure that must be removed from each variable before the group factor structure can be cleanly determined). In addition to confounding the interpretation of the group factors, the general factor itself is quite likely to be of both theoretical and practical importance.

To examine the presence of a general factor in the SSQ symptom matrix, the varimax-rotated matrix in Table 2 was transformed to hierarchical structure using the technique described by Wherry (1984), as modified by Wherry (1986). Results are shown in Table 3. Note that there is now a set of loadings on a general factor in addition to those on the three group factors. Note also that the three group factors are recognizable as the "same" factors seen in the varimax solution.

Comparing Table 2 to Table 3, it can be seen that (a) about 50% of the variance in the varimax solution has been extracted by a substantial general factor, (b) all the variables have sizable loadings on the general—that is, it is a "real" general, and (c) the number of "insignificant" loadings (i.e., absolute value < .10) has sharply increased on the group factors, making factor interpretation somewhat simpler.

TABLE 3
Hierarchical Factors From SSQ Symptoms

SSQ Symptom	*Hierarchical Factors*				
	G[a]	*N*	*O*	*D*	η^2
General discomfort	.64	.38	.22	−.06	.62
Fatigue	.26	.05	.48	−.12	.32
Headache	.40	.07	.43	.02	.35
Eyestrain	.34	−.12	.67	.09	.58
Difficulty focusing	.43	−.15	.52	.32	.56
Increased salivation	.48	.33	.08	−.05	.34
Sweating	.32	.18	.15	−.03	.16
Nausea	.66	.46	−.11	−.04	.66
Difficulty concentrating	.47	.13	.27	.11	.33
Fullness of head	.32	−.01	.09	.26	.18
Blurred vision	.33	−.11	.29	.31	.29
Dizzy (eyes open)	.51	−.03	−.06	.58	.60
Dizzy (eyes closed)	.47	−.01	−.03	.49	.46
Vertigo	.32	.05	−.01	.25	.17
Stomach awareness	.52	.41	−.12	.00	.45
Burping	.39	.25	−.07	−.07	.22
Eigenvalue	3.15	0.81	1.41	0.95	
Percent of variance	20	5	9	6	

[a]General.

The importance of the general factor for the SSQ is twofold. It may reflect the overall extent of symptom severity, and as such provides the best index of whether a given simulator has a sickness problem. Also, measures based on the group factors with the general removed are likely to be purer indicators of what is causing the problem than those based on the less independent varimax factors.

The SSQ measures based on the hierarchical model should at least theoretically be diagnostically superior to those based on the varimax model. However, the varimax factors, though interpretively less clear, are better defined in the mathematical sense for the limited set of symptoms contained in the SSQ. Just as there are too few symptoms to support reliable use of the four-, five-, and six-factor varimax solutions, there are also too few to provide reliable computation of weighted scores from the group factors in the hierarchical solution. Thus, although the general/group scores are almost certainly more appropriate from a theoretical standpoint, there is a tradeoff between diagnosticity and scale robustness that makes the varimax-based scales the measures of choice for the present limited symptom set. The long-term solution is to expand the SSQ with simulator-relevant symptoms that are related to (or even redundant with) the 16 presently included, in order to improve the reliability of the factor scores.

Beyond the small number of symptoms per factor, a second basis for conservatism in the use of the more complex factor scores lies in the lack of independence among the observations on which correlations are based. Some of those observa-

tions represent second and sometimes third exposures for the same individual. This causes variance arising from differences between respondents to be mixed with variance within respondents, which is at least in part a violation of the formal model underlying factor analysis procedures. Although the inherent structures of the matrices are not likely to be affected by this partial contamination, the precise effects on the values of factor loadings are not known, and the structures should be viewed in the purest statistical sense as clusters of variables rather than as mathematically exact factor representations. This suggests a preference for scoring systems that are less dependent on specific values from the factor analysis and more reliant on the general nature of the clusters that the analysis defines.

Deriving the SSQ Measures

The SSQ scoring system was chosen because it is the least dependent on the precision of parameter values derived from the sample. The weighting system and the associated constants produce score distributions with two important properties: (a) For all subscales and all total scores, the lowest possible score (i.e., no reported symptoms at all) is zero, and (b) the standard deviation of the scaled scores is 15 for the total sample of about 1,200 observations. Because the number of observations was so large, the sample was treated as if it were a population that could be used as a baseline against which future simulator evaluation data could be compared.

Scores were obtained by simple addition of the unweighted values of the symptoms in each cluster as defined by the varimax model. Thus, each entry in the weighting vectors for N, O, and D was either a 1 (if the varimax loading was greater than .30) or a 0 (otherwise). A simple index of total severity (TS) was computed from the sum of the three subscale scores prior to conversion.

USING AND SCORING THE SSQ

Administration

In order to use the SSQ, it is necessary to administer either a form containing the 16 symptoms identified in Table 1 with the 4-point scale for all items, or a modified MSQ using the 4-point scale, scoring only the appropriate 16 symptoms. In either case, information should be solicited about subjects' present states of health, as noted previously. The scoring procedures outlined hereafter presume that all individuals in other than their usual state of fitness are eliminated from the sample, and that only postexposure data are scored.

Scoring

Table 4 contains the scoring procedures for the SSQ. The SSQ uses unit weights and is both simpler to use and more stable than scoring based on more precise weights defined by varimax factor weights. To compute the scale scores, each symptom variable score (0, 1, 2, 3) was multiplied by the appropriate weight, and the weighted values were summed down the column to obtain the weighted total. The N, O, and D scores are then calculated from the weighted totals using the conversion formulas given at the bottom of the table. The TS score is obtained by summing all the weighted totals and applying the TS conversion formula.

Interpreting the SSQ Scores

There is no particular interpretive meaning to the values in the conversion formulas; their function is only to produce scales with similar variabilities on which values

TABLE 4
Computation of SSQ Scores

	Weight		
SSQ Symptom[a]	*N*	*O*	*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[b]	[1]	[2]	[3]

Score
N = [1] × 9.54
O = [2] × 7.58
D = [3] × 13.92
TS[c] = [1] + [2] + [3] × 3.74

[a]Scored 0, 1, 2, 3. [b]Sum obtained by adding symptom scores. Omitted scores are zero. [c]Total Score.

can be more readily compared. It should be noted, however, that although variability on all scales is equated for the calibration sample of more than 1,100 observations, the midpoints are not equal.[1]

Another way of interpreting and anchoring the scale values is in comparison to those in Table 6, which gives the means and standard deviations on the SSQ scale for each of the simulators in the calibration group. Data on only nine simulators are given; one was omitted because the sample size was too small to be stable. Detailed descriptions of the visual and motion systems and other characteristics of these simulators is given in Kennedy et al. (1989). Means from new simulator evaluations should fall within the ranges of those in Tables 5 and 6, and an indication of the relative severity and nature of the SS problem in a new sample can be obtained by comparison. If the new means fall within the range of the upper three to four simulators on a given scale, closer examination of the data and of the simulator itself is probably warranted.

The understanding and interpretation of SSQ values for simulators other than those in the calibration sample are facilitated by comparison of obtained values to those in Tables 5 and 6. Table 5 contains the percentile points for each SSQ scale over the 1,100+ observations in the calibration sample.

Because the 10 simulators surveyed are probably representative of those in the training simulator community, the information in Table 5 allows the determination (for a new score value) of the number of observations in the population that were more or less extreme than that value. It should be noted from the table that for every scale, the 0-value (the zero point) contains at least 40%, and as much as 75%, of the observations. This points out that the modal position with respect to SS across simulators in general is no indicated symptomatology at all, and the sensitivity of the scales is largely at the upper extremes of the symptomatology range. Therefore, the scales do not distinguish among simulators that have no problems, but are rather intended to discriminate problem simulators from those with no indicated difficulties.

Note in the tables the additional diagnostic power of the separate subscales. The 2F87F, for example, is very high on the TS scale, roughly equivalent to a single overall index. As the subscales indicate, however, it is one of the lowest on the N scale, and one of the highest on O; its high score on TS is most likely due almost entirely to some property of the visual system, rather than to some generalized deficiency. Although the impression obtained from the overall index is neither incorrect nor misleading, the availability of the subscales can focus attention more quickly on the probable nature of a solution.

---

[1]There are three scale attributes that can be equated by conversion formulas: zero point, midpoint, and standard deviation. Only two of these can be manipulated in a given set of conversions. For example, once the midpoints are equated by adding or subtracting a constant, the zero points can only be equated by allowing the standard deviations to vary across scales. Similarly, fixing the midpoints and standard deviations automatically defines different zero points for each scale.

TABLE 5
Percentile Points and Descriptive Statistics for SSQ Scale in the Calibration Sample

Percentile Point	*SSQ Scale Value*			
	N	*O*	*D*	*TS*[a]
40	0.0	0.0	0.0	0.0
45	0.0	0.0	0.0	3.7
50	0.0	7.6	0.0	3.7
55	0.0	7.6	0.0	3.7
60	0.0	7.6	0.0	7.5
65	9.5	7.6	0.0	7.5
70	9.5	15.2	0.0	11.2
75	9.5	15.2	0.0	15.0
80	9.5	22.7	13.9	22.5
85	19.7	27.7	13.9	22.5
90	28.6	30.3	27.8	30.0
95	38.2	45.5	41.7	44.9
96	38.2	45.5	41.7	44.9
97	47.7	53.1	55.7	48.7
98	57.2	53.1	55.7	56.2
99	66.8	60.7	83.5	75.9
M	7.7	10.6	6.4	9.8
SD	15.0	15.0	15.0	15.0
Minimum	0.0	0.0	0.0	0.0
Maximum	124.0	90.9	97.4	108.6
n	1101	1111	1109	1099

[a]Total Severity.

TABLE 6
SSQ Scale Means by Simulator for the Calibration Sample

Simulator	*Aircraft*	*SSQ Scale M*			
		N	*O*	*D*	*TS*[a]
2F64C	SH-3	14.7	20.0	12.4	18.8
2F120	CH-53E	7.5	10.5	7.4	10.0
2F121	CH-53D	7.2	7.2	4.0	7.5
2F110	E-2C	7.1	13.1	6.8	10.3
2E7	F/A-18	6.1	5.1	6.2	6.8
2F117	CH-46E	5.4	7.8	4.5	7.0
2F87F	P-3C	4.5	15.2	4.3	10.5
2F132	F/A-18	2.7	6.1	0.6	4.2
2F112	F-14	1.7	1.8	0.0	1.5
M		7.7	10.6	6.4	9.8
SD		15.0	15.0	15.0	15.0

[a]Total Severity.

AN ILLUSTRATIVE EXAMPLE: FIELD TESTING A SEMIAUTOMATICALLY SCORED SSQ

Method

The SSQ was implemented on a portable computer and fielded at a number of simulator sites to provide a relatively inexpensive, maintenance-free method for SS data collection. The present example includes the results of a 20-month program of data acquisition. The system was fielded at each of two TH-57 trainers, Devices 4 (October 1988) and 2 (January 1989), in Naval Air Station, Whiting Field, Florida. The TH-57 is a primary helicopter flight trainer used in undergraduate training to transition pilots from fixed-wing to rotary-wing aircraft. Pilots completed the computerized survey immediately upon exiting the simulator. To date, data have been recorded from 3,691 hops. The data are time and date stamped, which provides information about symptomatology occurrence over time, hop number (for any given pilot), and days between hops. In addition, the data can be used to compare the incidence rates between the two units as well as with rates obtained from other devices.

The benefits of automated-SS data collection are twofold. First, it is crucial to identify and monitor the degree of severity and subsequent adaptation to simulator-induced illness and possible aftereffects. Self-scoring software can provide immediate feedback to increase the pilot's awareness of symptoms and resulting limitations so that necessary precautions can be taken to reduce risk in subsequent activities. Second, when monitored across time, the overall incidence rate as well as specific types of symptoms can serve as a baseline against which subsequent symptomatology data can be compared. The data can be used for intersimulator as well as intrasimulator comparison.

The baseline data can be used for several systems engineering applications. Increased report of symptoms may be indicative of a simulator malfunction, and the data may be used for troubleshooting purposes to diagnose the malfunction (e.g., system calibration that has exceeded the tolerance set forth in the device specification). Symptom profiles (e.g., excessive report of visual disturbance) may provide insight for the identification of specific simulator-engineering features that should be targeted for engineering efforts to alleviate the problem. Finally, automated monitoring of simulator-induced symptomatology can be used to assess the impact of engineering feature modifications as well as changes to the training syllabus (e.g., increased hop length).

Results

Figure 1 contains the distribution of scores that were obtained for 3,691 cases. It may be seen that essentially half the population reported no symptoms from their exposure to the simulator and half reported various amounts of sickness from mild to severe (as was found with the simulators reported in Table 6). Subsequent analyses were performed using arithmetic means and, because of the extreme

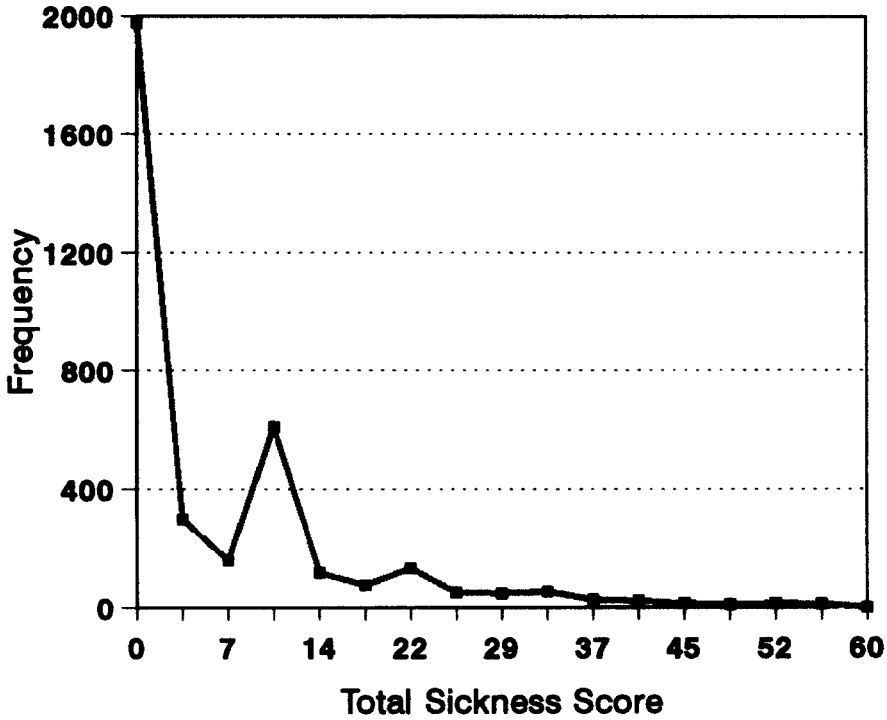

FIGURE 1 Frequency distribution of total sickness scores (N = 3,691).

skewness, 75th-percentile scores. The rationale for using the 75th percentile was that this is essentially the midpoint of the part of the population that was adversely affected by the exposure.

Figure 2 shows the 75th-percentile score for the period that the data were collected (November 1989–February 1991). It may be seen that after the simulators were installed, there was an initial settling-in period (reflected in relatively high symptomatology levels) followed by a relatively flat incidence level for the 75th-percentile score over a 12-month period. In June and July 1990, there was an increase in the sickness score. From discussions with personnel at the simulator site, there was a corresponding surge in simulator usage, to the extent that Saturday flying and two hops per day were accomplished. The data from Figure 2 coincides with the surge and, when the normal flight schedule was resumed, the incidence of sickness also went down. Because of the extreme stability of these measures, it is possible to use the score at the 75th percentile on a weekly basis in order to evaluate periods just before and just after maintenance to observe whether sickness rates rise before a maintenance period and/or fall subsequent to a maintenance period.

Figure 3 reveals the effects of repeated hops on adaptation to SS stress. Note that on Hop 1, the 75th percentile person exhibits a score of 15, which is comparable to the average score on the Navy's most sickness-inducing simulator (see Table 6).

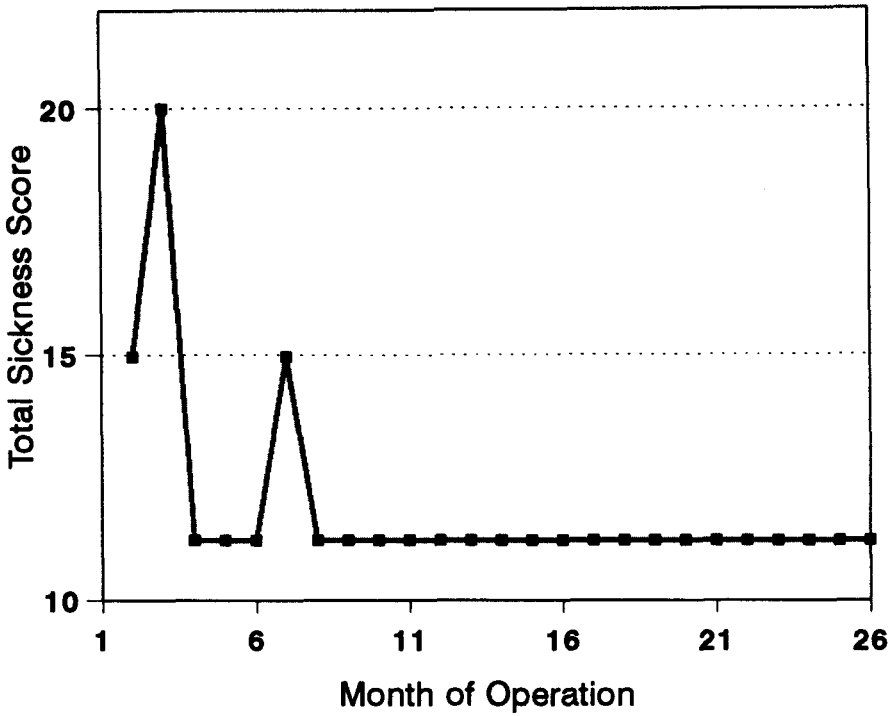

FIGURE 2 75th-percentile sickness scores across 26 months of data collection.

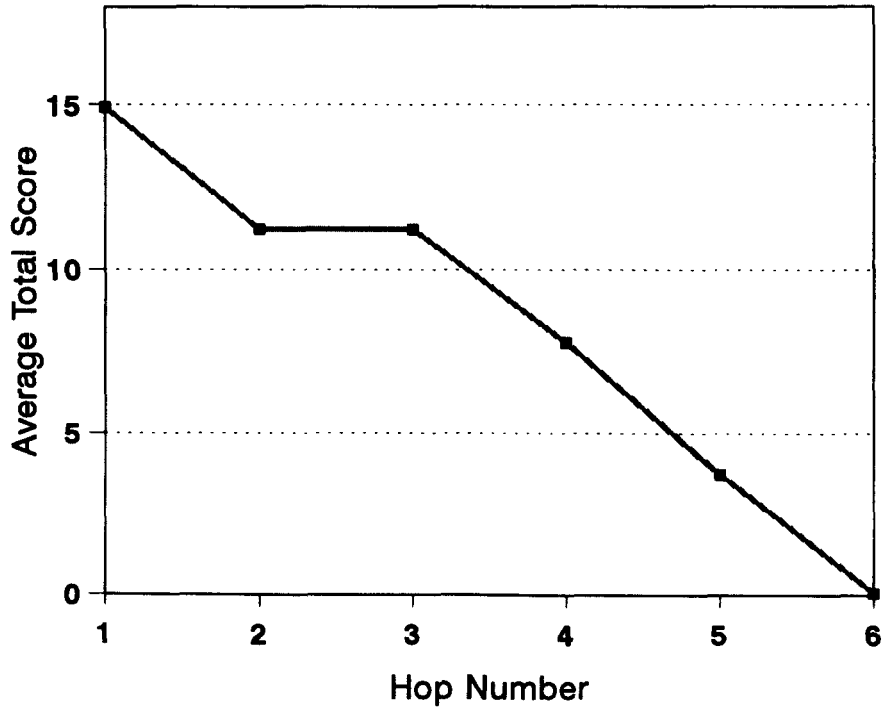

FIGURE 3 75th-percentile sickness scores as a function of hop number.

However, after 4 hops, most of the SS has subsided. Figure 4 shows the score obtained on each pilot's second exposure as a function of the separation between exposures. Note that sickness rates are highest when hops are on the same day or one day apart. Similarly, when hops are spaced more than 5 days apart, there is also little adaptation. The optimal spacing, in terms of controlling the incidence of SS in this simulator, appears to be 2 to 5 days between hops. Perhaps this interval is long enough to avoid carryover of symptoms from one exposure to the next and short enough to retain whatever adaptation to the simulator has been gained.

Discussion

Our illustrative example of the use of the SSQ as a semiautomatic reporting system included 3,691 exposures. Using the scoring method derived earlier, we found that adaptation occurs over hops, so that after four hops SS is very slight. The best regime for promoting this adaptation is to separate flights by 2 to 5 days. One of the interesting findings is that the 75th-percentile metric can be a very stable index of activity in simulators and perhaps has some utility for monitoring maintenance of equipment during in-service engineering. We recommend that baseline scores should be obtained and then compared to the incidences and symptom mixture after any engineering changes are made to simulator configurations.

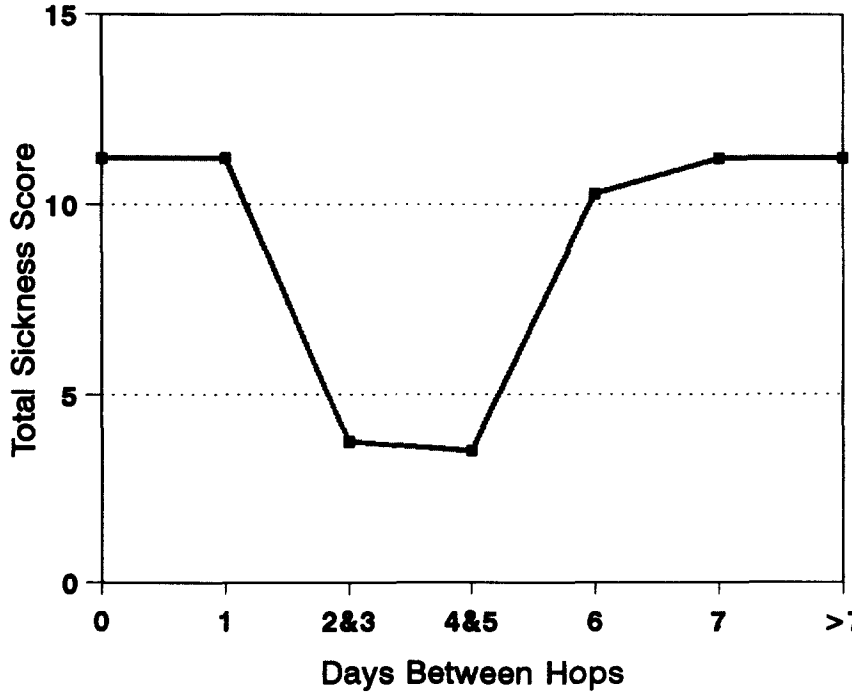

FIGURE 4 75th-percentile sickness scores as a function of days between hops.

CONCLUSIONS

Some summary statements follow from the foregoing. The patterns of symptom presence and severity associated with SS are sufficiently different from those of motion sickness to justify the use of separate measuring systems tailored to quantification of those specific patterns. In both the present analysis and related analyses, there seem to be at least three separate dimensions underlying MS and SS. Each of these dimensions operates through a different target system in the human organism to produce undesirable symptoms. The importance of identifying and understanding these dimensions is that the mechanisms for amelioration and control may be different for each affected target system. SSQ scoring, based on factor-analytic models, provides both good indications of overall SS severity and reasonably powerful subscale scores for diagnostic purposes. This simple method—using unit weights on variables identified by varimax rotation—will be adequate for most applications. A deficiency of this scoring system is that its subscales are more highly correlated than is optimal for diagnostic use. Scoring based on hierarchical factor rotation would produce subscales with much lower interdependence. However, the limited number of SS-relevant symptoms in the present analysis is not sufficient to properly define and anchor subscales produced by hierarchial rotation, and the reliability of the scales would be too low. Development of hierarchically based scaling would require larger symptom lists that incorporated planned redundancy with the present list.

ACKNOWLEDGMENTS

This research was supported by U.S. Navy Contract N61339-86-D-0026, granted to the Naval Training Systems Center; Subcontract 15X-64011V from Martin Marietta Energy Systems; and U.S. Navy Contract N61339-81-C-0105, granted to the Naval Training Equipment Center.

REFERENCES

Alexander, S. J., Cotzin, M., Klee, J. B., & Wendt, G. R. (1947). Studies of motion sickness: XVI. The effects upon sickness rates of waves of various frequencies but identical acceleration. *Journal of Experimental Psychology, 37,* 440–448.

Baltzley, D. R., Kennedy, R. S., Berbaum, K. S., Lilienthal, M. G., & Gower, D. W. (1989). The time course of post–flight-simulator sickness symptoms. *Aviation, Space, and Environmental Medicine, 60,* 1043–1048.

Bittner, A. C., Jr., & Guignard, J. C. (1988). Shipboard evaluation of motion sickness incidence. In F. Aghazadeh (Ed.), *Trends in ergonomics/human factors* (Vol. 5, pp. 529–541). New York: North Holland.

Cronbach, L. J., & Furby, L. (1970). How to measure change—or should we? *Psychological Bulletin, 74,* 68–70.

DeWit, G. (1957). Acquired sensitivity to seasickness after an influenza infection. *Practica Otorhinolaryngologica, 19,* 579–586.

Dichgans, J., & Brandt, T. (1973). Optokinetic motion sickness and pseudo-Coriolis effects induced by moving visual stimuli. *Acta Otolaryngologica, 76,* 339–348.

Frank, L. H., Kennedy, R. S., Kellogg, R. S., & McCauley, M. E. (1983). *Simulator sickness: Reaction to a transformed perceptual world: I. Scope of the problem* (NAVTRAEQUIPCEN TN-65). Orlando, FL: Naval Training Equipment Center.

Gower, D. W., Lilienthal, M. G., Kennedy, R. S., & Fowlkes, J. E. (1987). Simulator sickness in U.S. Army and Navy fixed- and rotary-wing flight simulators. In *Proceedings of the AGARD Medical Panel Symposium on Motion Cues in Flight Simulation and Simulator Induced Sickness* (pp. 8-1–8-20). Brussels, Belgium: North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development.

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

Hemingway, A. (1942). *Results of 500 swing tests for investigating motion sickness* (Project No. 31, Rep. No. 5). Randolf Field, TX: School of Aviation Medicine.

Kellogg, R. S., Kennedy, R. S., & Graybiel, A. (1965). Motion sickness symptomatology of labyrinthine defective and normal subjects during zero gravity maneuvers. *Aerospace Medicine, 36,* 315–318.

Kennedy, R. S., Berbaum, K. S., Allgood, G. O., Lane, N. E., Lilienthal, M. G., & Baltzley, D. R. (1987). Etiological significance of equipment features and pilot history in simulator sickness. *Proceedings of the AGARD Medical Panel Symposium on Motion Cues in Flight Simulation and Simulator Induced Sickness* (pp. 1-1–1-22). Brussels, Belgium: North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development.

Kennedy, R. S., Lilienthal, M. G., Berbaum, K. S., Baltzley, D. R., & McCauley, M. E. (1989). Simulator sickness in U.S. Navy flight simulators. *Aviation, Space, and Environmental Medicine, 60,* 10–16.

Kennedy, R. S., Tolhurst, G. C., & Graybiel, A. (1965). *The effects of visual deprivation on adaptation to a rotating environment* (Naval School of Aviation Medicine Tech. Rep. No. 918). Pensacola, FL: Naval School of Aerospace Medicine.

Morrissey, S. J., & Bittner, A. C., Jr. (1986). Vestibular, perceptual, and subjective changes with extended VDU use: A motion sickness syndrome? In W. Karkowski (Ed.), *Trends* in ergonomics/human factors (Vol. 3, pp. 259–265). New York: North Holland.

Ungs, T. J. (1987). Simulator induced syndrome: Evidence for long term simulator aftereffects. In *Proceedings of the Human Factors Society 31st Annual Meeting* (Vol. 1., pp. 505–509). Santa Monica, CA: Human Factors Society.

Wherry, R. J. (1984). *Contributions to correlational analysis.* Orlando: Academic.

Wherry, R. J., Jr. (1986). *Theoretical development for identifying underlying internal processes: Vol. 2. Modifications to hierarchical factor analysis: Positive manifold (POSMAN) rotations* (NAMRL SR 86-1/NADC Rep. 86105-60). Pensacola, FL: Naval Aerospace Medical Research Laboratories.