ERGONOMICS, 2001, VOL. 44, NO. 3, 229–243

Simplified subjective workload assessment technique

AMEERSING LUXIMON and RAVINDRA S. GOONETILLEKE*

Human Performance Laboratory, Department of Industrial Engineering and Engineering Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

Keywords: SWAT; mental workload; arithmetic task.

Although the subjective workload assessment technique (SWAT) has been widely used, it has two main problems: it is not very sensitive for low mental workloads and it requires a time-consuming card sorting pretask procedure. In this study are presented five variations of SWAT in an effort to overcome the limitations. Four of the variants used the continuous SWAT subscales while one used the discrete SWAT subscale. Fifteen subjects participated in the experiment. The scales were compared with the original SWAT scale in terms of sensitivity and pretask procedure completion time when performing arithmetic tasks. The results show that all four variants are more sensitive than the conventional SWAT scale and that the pairwise comparison procedure takes significantly less pretask completion time compared with the original SWAT scale. Thus, the conventional pretask procedure can be replaced by a simple unweighted averaging to yield a scale of high sensitivity.

1. Introduction

Owing to technological advancements and the wide use of computers, a good portion of work is now cognitive. Consequently, there is a constant need to assess the cognitive or the mental workload a system imposes on a person. Four techniques are commonly used to assess mental workload: physiological measures, subjective measures, secondary task measures and primary task measures (Meshkati and Loewenthal 1988, Meshkati *et al.* 1995).

These measures vary in terms of certain important criteria (table 1) that determine their usefulness for individual applications (Eggemeier 1988). Nevertheless, mental workload measures have been compared mostly in terms of their sensitivity (Wierwille and Casali 1983, Wierwille and Connor 1983, Hart and Staveland 1988, Hill *et al.* 1992), which is evaluated in numerous ways (Hendy *et al.* 1993). For example, Wierwille and Connor (1983) defined sensitivity as the ability to discriminate between different load conditions. Hart and Staveland (1988) on the other hand have been more concerned about the between-subject variability as an indication of sensitivity. Hill *et al.* (1992) computed sensitivity in terms of factor validity and so on.

*Author for correspondence. e-mail: ravindra@ust.hk

Ergonomics ISSN 0014-0139 print/ISSN 1366-5847 online © 2001 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/00140130010000901

Table 1. Criteria to rate mental workload measures.

Criteria	Reference	Comments
Validity	Meshkati (1988)	Should satisfy content, predicability and construct validity
Selectivity	Stassen *et al*. (1990)	Immunity to other variables or selectivity
Obtrusiveness	Meshkati (1988), Stassen *et al*. (1990), Eggemeier (1988)	Should not interfere with and cause degradation in ongoing primary task performance
Diagnosticity or relevance	Jex (1988), Meshkati (1988), Stassen *et al*. (1990)	Capability of a technique to discriminate the amount of workload imposed on different resources or capabilities of the human operator
Sensitivity	Jex (1988), Meshkati (1988), Stassen *et al*. (1990), Eggemeier (1988)	Monotonic trend with respect to mental workload Insensitive to other variables or ambient environment Capability of a technique to discriminate significant variations in workload levels imposed by a task or group of tasks
Repeatability or reliability	Jex (1988), Meshkati (1988), Stassen *et al*. (1990)	Proven test–retest repeatability Differential stability (parallel) trends) among subjects with practice on a task Validated means and variance statistics with norms for the population
Consistency or concordant	Jex (1988), Stassen *et al*. (1990)	Ubiquitous trends in target population
Bandwith	Stassen *et al*. (1990)	Track the mental workload variation with time
Convenient	Jex (1988)	Easy to learn and administer Portable for use in field trials and evaluations Low cost for a given level of measurement reliability

Even though much effort has been made to develop objective measures of workload, subjective workload assessment techniques continue to be popular due to their ease of use, general non-intrusiveness, low cost, high face validity and known sensitivity to workload variations (Reid and Nygren 1988). Subjective mental workload can be defined as the subject's direct estimate or comparative judgement of the mental or cognitive workload experienced at a given moment (Reid and Nygren 1988).

Several types of uni- and multidimensional subjective scales exist. Examples are the Cooper Harper scale (Wierwille and Casali 1983), direct scaling (Ghiaseddin 1995), the multidescriptor scale (Casali and Wierwille 1983), the workload–compensation–interference/technical effectiveness scale (Wierwille and Connor 1983), the overall workload scale (Hill *et al*. 1992), the consumer mental workload scale (Owen 1992), SWAT (Reid and Nygren 1988), and the National Aeronautics and Space Administration—Task Load Index (NASA-TLX) (Hart and Staveland 1988). However, results from various studies have shown that NASA-TLX and SWAT are very popular and are widely used (Hendy *et al*. 1993).

NASA-TLX is a multidimensional scale for which the overall mental workload is a function of mental demand, physical demand, temporal demand, performance, effort and frustration dimensions, with each of these dimensions on a continuum. SWAT is also a multidimensional scale, but its dimensions of time load (T), mental effort load (E), and psychological stress load (S) are at three discrete levels. Even though SWAT has been tested with a psychological model of human-perceived information processing demand, studies such as Hart and Staveland (1988) and Hill *et al*. (1992) have shown that NASA-TLX is superior to SWAT in terms of sensitivity especially for low mental workloads (Nygren 1991).

When using the SWAT scale, a participant is required to perform a card sorting (CS) pretask procedure followed by a task (or event) scoring procedure. During the pretask procedure, the participant ranks 27 SWAT cards, which are yielded from the combinations of the three discrete dimensions at three discrete levels (Reid and Nygren 1988). For each dimension, the levels have descriptors that represent the lowest mental workload (level 1) to the highest mental workload (level 3). In the CS procedure, the participants are required to rank the cards beginning with the card that represents the lowest mental workload and ending with the card that represents the highest workload (Reid and Nygren 1988).

The first step in analysing the sorted card data is to determine the level of agreement among the participants using Kendall's coefficient of concordance (W). A single scale is developed by averaging data if $W > 0.75$ (Reid and Nygren 1988). However, if the focus of the study pertains to individual differences, then scales for individual participants are developed. When $W < 0.75$, homogeneous subgroup scales are developed, if appropriate. Based on the relative importance of each dimension, six hypothetical orderings have been developed (table 2). For example, TES is the ordering when there is greatest emphasis on T, the second greatest on E and the third on S. Similarly, TSE, ETS, EST, STE and SET weighting schemes can be derived (Reid and Nygren 1988). The Spearman correlation between the participant's sorting and the hypothetical ordering is used to decide which one of the six subgroups is more suitable. Once the number of groups has been determined, conjoint analysis is performed to obtain a workload scale ranging between 0 and 100. As can be seen, this is a very tedious procedure for obtaining the workload ratings.

2. Study rationale

Even though SWAT has been widely used and appears to be more suitable than other mental workload techniques in terms of diagnosticity and content validity, it has constantly been criticized as having low sensitivity for low mental workloads. The sensitivity of the scale could be improved by adding more levels to each of the dimensions (Nygren 1991). However, adding more levels will create more combinations and thus the number of SWAT cards would increase drastically, making the CS task prohibitively difficult, error prone, and even more time-consuming than the already tedious procedure. Hence, an alternative way of analysing the SWAT data should be explored (Nygren 1991).

This study is an attempt to improve the sensitivity of the SWAT scale while also reducing the pretest completion time. It used simple arithmetic manipulations to test the modified forms of the scale since these could simulate low and medium mental workloads. These worked along the lines of Kahneman *et al*. (1969), for example, who showed that simple digit transformation tasks result in significant effects on pupil diameter and skin resistance, but marginally significant results for heart rate; or Humphrey and Kramer (1994), who showed that there is a significant change in the P300 component of Event-Related Potential (ERP) when doing mental arithmetic tasks with addition and multiplication operations of single digit numbers.

The objectives of the study were to investigate five different variations of the original SWAT scale in terms of their sensitivity and pretask procedural times. There

Table 2. Six hypothetical weighting schemes.

Rank order	Descriptor combination in the card is time (T), effort (E) and stress (S) respectively TES			TSE			ETS			EST			STE			SET		
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	2	1	2	1	1	1	2	2	1	1	1	2	1	2	1	1
3	1	1	3	1	3	1	1	1	3	3	1	1	1	3	1	3	1	1
4	1	2	1	1	1	2	2	1	1	1	1	2	2	1	1	1	2	1
5	1	2	2	1	2	2	2	1	2	2	1	2	2	2	1	2	2	1
6	1	2	3	1	3	2	2	1	3	3	1	2	2	3	1	3	2	1
7	1	3	1	1	1	3	3	1	1	1	1	3	3	1	1	1	3	1
8	1	3	2	1	2	3	3	1	2	2	1	3	3	2	1	2	3	1
9	1	3	3	1	3	3	3	1	3	3	1	3	3	3	1	3	3	1
10	2	1	1	2	1	1	1	2	1	1	2	1	1	1	2	1	1	2
11	2	1	2	2	2	1	1	2	2	2	2	1	1	2	2	2	1	2
12	2	1	3	2	3	1	1	2	3	3	2	1	1	3	2	3	1	2
13	2	2	1	2	1	2	2	2	1	1	2	2	2	1	2	1	2	2
14	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
15	2	2	3	2	3	2	2	2	3	3	2	2	2	3	2	3	2	2
16	2	3	1	2	1	3	3	2	1	1	2	3	3	1	2	1	3	2
17	2	3	2	2	2	3	3	2	2	2	2	3	3	2	2	2	3	2
18	2	3	3	2	3	3	3	2	3	3	2	3	3	3	2	3	3	2
19	3	1	1	3	1	1	1	3	1	1	3	1	1	1	3	1	1	3
20	3	1	2	3	2	1	1	3	2	2	3	1	1	2	3	2	1	3
21	3	1	3	3	3	1	1	3	3	3	3	1	1	3	3	3	1	3
22	3	2	1	3	1	2	2	3	1	1	3	2	2	1	3	1	2	3
23	3	2	2	3	2	2	2	3	2	2	3	2	2	2	3	2	2	3
24	3	2	3	3	3	2	2	3	3	3	3	2	2	3	3	3	2	3
25	3	3	1	3	1	3	3	3	1	1	3	3	3	1	3	1	3	3
26	3	3	2	3	2	3	3	3	2	2	3	3	3	2	3	2	3	3
27	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

were two main differences among these variations. First, each of the scales had one of three possibilities for the pretask procedure: CS, pairwise comparison (PWC) (figure 1) or none. Second, the scales of each dimension were chosen to be continuous or discrete. The five variations were:

1. Discrete SWAT dimensions (D_{SWAT}) was similar to the SWAT scale. In D_{SWAT}, a PWC procedure was used instead of a CS procedure. Also conjoint analysis was used for subgroups instead of using one analysis for each individual.
2. Continuous SWAT dimensions with minimum weight equal to zero (W_0) used a PWC procedure and continuous subscales. The overall mental workload was obtained by using a weighting scheme similar to that of the NASA-TLX scale (Hart and Staveland 1988). In this scheme, weights of 0, 1/3 or 2/3 were used.
3. Continuous SWAT dimensions with non-zero minimum weight (W_1) was similar to (W_0). However, to avoid the assignment of 0 weight to a dimension, weights of 1/6, 1/3 or 1/2 were used.
4. Continuous SWAT dimensions with equal weight (A_{SWAT}) did not require a pretask procedure. The overall mental workload was obtained using an unweighted average of the three SWAT dimensions (Biers and Masline 1987, Hendy *et al*. 1993).

5. Continuous SWAT dimensions with weight based on principal component analysis (PC_c) also did not require a pretask procedure. The overall mental workload was obtained by weighting the dimensions using the coefficient of the first principal component (Hendy *et al.* 1993).

SWAT and D_{SWAT} used discrete scales and conjoint analysis was used for scaling in order to obtain an overall mental workload scale. In D_{SWAT}, the PWC task was used to group the participants into one of the six hypothetical orderings based on their emphasis. If one participant selected time load twice, and psychological stress once, it implies that he/she gave more importance to T, then S, then E. As a result, the participant would be grouped into the TSE subgroup. If the participant gave equal weight to all three dimensions, then the values for the dimensions would be averaged to develop an overall mental workload scale. Once participants were grouped into one of the six subgroups, the hypothetical ordering for the subgroups shown in table 2 was used to perform conjoint analysis.

On the other hand, W_0, W_1, A_{SWAT} and PC_c made use of the continuous subscale and the results could be used to verify the effectiveness of different weighting schemes. A_{SWAT} and PC_c weighting schemes have been used by Hendy *et al.* (1993) when comparing different NASA-TLX weighting schemes. Furthermore, Biers and Masline (1987) used the A_{SWAT} weighting scheme to compare different discrete SWAT scales.

3. Methodology

3.1. *Participants*

Fifteen students aged between 20 and 30 years from the Hong Kong University of Science and Technology participated in the study.

3.2. *Procedure*

After completing an informed consent form, each participant was given information about the experimental objectives and procedures. For the CS procedure, the participants were asked to sort the 27 SWAT cards in increasing order of mental workload (Reid and Nygren 1988). In the PWC procedure (figure 1), the participants were asked to select one of the two dimensions of workload that they felt were more important to them. Each participant did both the CS and PWC procedures in random order and the time taken for each procedure was recorded.

After that, the participants performed three different types of arithmetic manipulations (addition, subtraction, and multiplication) at 10 difficulty levels. The values of the first and second operand were randomly generated to obtain the 10 levels (table 3). The type of operations and the task level determined the task difficulty (or mental workload). The tasks were presented to the participants through an interface designed in Visual C+ + (v.5) and were completely randomized and counter-balanced. After the participants completed each task, their estimate of the mental workload was assessed using the three dimensions of the continuous and discrete SWAT scales. Here again, the sequence was randomly assigned.

3.3. *Results and analysis*

All analyses were performed using the SAS package except for the Kendall coefficient of Concordance, which was computed using the SPSS statistical software.

Please tick one of the two dimensions of workload that you think is more important to you.

Mental effort load ❐/ time load ❐

Time load ❐/ psychological stress load ❐

Psychological stress load ❐/ mental effort load ❐

Figure 1. Pairwise comparison (PWC) procedure.

Table 3. Ten task levels and their first and second operand.

Level	First operand	Second operand
1	7	4
2	48	9
3	338	6
4	4452	7
5	87	93
6	742	64
7	9664	57
8	819	657
9	9178	197
10	3284	3457

The comparison between the two pretask procedural times using a *t*-test showed a significant difference ($t(14) = 9.097$, $p < 0.05$): the mean time for the PWC task was 22.08 compared with 476.49 s for the CS procedure.

The Kendall coefficient of concordance for the pretask procedure in this study was 0.002. Hence, an individual SWAT scale was developed instead of a group SWAT scale (Reid and Nygren 1988). The high Spearman Correlation coefficients between the assigned hypothetical subgroup and the actual card ranking for each participant are shown in table 4. In this study, none of the participants gave an equal weighting for all three dimensions, so it was easy to group the participants into one of the six hypothetical subgroups.

Table 5 shows the correlation results among the three dimensions of the discrete SWAT scale and the three dimensions of the continuous SWAT scale. Results showed a relatively low correlation among the discrete dimensions. Correlations between time load and mental effort, time load and psychological stress, and mental effort and psychological stress were 0.63, 0.59 and 0.54 respectively. However the correlations among the continuous SWAT scales were relatively high (min $R^2 = 0.59$).

Correlation results (table 6) showed that the variants of SWAT scales were quite related ($R^2 > 0.64$). The correlation analyses by operation (table 6) showed that the correlation between discrete scales (SWAT and D_{SWAT}) and continuous scales (W_0, W_1, A_{SWAT}, $\mathrm{PC_c}$) was relatively high for the multiplication task (min $R^2 = 0.69$) but relatively low for the addition (max $R^2 = 0.33$) and subtraction tasks (max $R^2 =$

Table 4. Spearman correlation coefficient between each subgroup and the actual card sorting for each subject.

Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Subgroup	SET	TSE	EST	TES	SET	SET	SET	ETS	EST	SET	SET	TES	ETS	SET	SET
Spearman coefficient	0.79	0.86	0.62	0.96	0.75	0.96	0.82	0.83	0.71	0.88	0.84	1.00	0.82	0.77	0.87

Table 5. Correlation results for discrete and continuous scales.

	Correlation results					
	Discrete			Continuous		
Variable	T	E	S	T	E	S
Discrete T	1	0.63	0.59	0.84	0.70	0.60
Discrete E		1	0.54	0.61	0.74	0.54
Discrete S			1	0.61	0.65	0.84
Continuous T				1	0.84	0.77
Continuous E					1	0.80
Continuous S						1

0.49). For all operations, the correlation between the discrete scales was relatively high (min $R^2 = 0.60$), and so was that between the continuous scales (min $R^2 = 0.85$).

A 3 (Operations)*10 (level) analysis of variance was performed with the W_0, W_1, A_{SWAT}, $\mathrm{PC_c}$, SWAT and D_{SWAT} scores as dependent variables. For all the dependent variables all main effects and the interaction 'Operation*level' were significant at $p = 0.0001$. The *post-hoc* Student–Newman–Keuls (SNK) test results are shown in figures 2 and 3. It could be seen that multiplication imposed a significantly heavier mental workload. Similar effects were seen for a higher task level, where the number of digits increased. One important observation was that SWAT could not discriminate between levels 9 and 10 whereas all the continuous scales could, indicating that the continuous scales were more sensitive (based on the definition given by Wierwille and Connor 1983).

Owing to the presence of a significant interaction, a simple effects ANOVA was performed. The ANOVA by operation (figure 4) shows that the discrete scales (SWAT and D_{SWAT}) could not discriminate between the different task levels for the addition task, while the continuous scales showed significant differences. Furthermore, figure 5 shows the *post-hoc* SNK analysis for each task level. Again, SWAT could not discriminate among the three operations for task levels 2 and 3 while most other variants of SWAT could. This shows that all the proposed SWAT variants were better than SWAT in terms of sensitivity. Since there was a very high correlation between the discrete scales, the overall workload of only one discrete scale (SWAT) is shown in figure 6. The results of one continuous scale (A_{SWAT}) are shown in figure 7.

Sensitivity of the scales was calculated using two methods. First, the factor loading sensitivity analysis was used (Hill *et al.* 1992). In this method, the loading of the different mental workload measures on the first factor was calculated. If the loading was high, the mental workload measure was said to be more sensitive (Hill *et al.* 1992). Second,

Measure	Operation
W_0, W_1, A_{SWAT}, PC_C, SWAT and D_{SWAT}	+ - x

Operations are shown in increasing level of mental workload.

Underlining denotes that the means are not significantly different at $p < 0.05$.

Figure 2. SNK result for the main effect, Operation.

Measure	level
W_0	1 2 3 4 5 6 8 7 9 10
W_1	1 2 3 4 5 6 8 7 9 10
A_{SWAT}	1 2 3 4 5 6 7 8 9 10
PC_C	1 2 3 4 5 6 7 8 9 10
SWAT	1 2 3 4 5 6 7 8 9 10
D_{SWAT}	1 2 3 4 5 6 7 8 9 10

Mental workload increases from left to right.

Underlining denotes that the means are not significantly different at $p < 0.05$.

Figure 3. SNK result for the main effect, Task Level.

Table 6. Correlation results for all operations. Those shown in brackets are for subtraction, curly brackets are for addition and parentheses for multiplication.

	W_0	W_1	A_{SWAT}	PC_C	SWAT	D_{SWAT}
W_0	1	0.992	0.97	0.97	0.814	0.798
	[1]	[0.994]	[0.974]	[0.973]	[0.657]	[0.672]
	{1}	{0.981}	{0.922}	{0.92}	{0.527}	{0.341}
	(1)	(0.993)	(0.973)	(0.973)	(0.848)	(0.84)
W_1		1	0.993	0.993	0.822	0.833
		[1]	[0.993]	[0.993]	[0.657]	[0.69]
		{1}	{0.98}	{0.979}	{0.563}	{0.446}
		(1)	(0.993)	(0.993)	(0.845)	(0.866)
A_{SWAT}			1	0.999	0.818	0.855
			[1]	[0.999]	[0.648]	[0.701]
			{1}	{0.999}	{0.576}	{0.535}
			(1)	(0.999)	(0.83)	(0.879)
PC_C				1	0.818	0.855
				[1]	[0.648]	[0.701]
				{1}	{0.576}	{0.536}
				(1)	(0.83)	(0.879)
SWAT					1	0.889
					[1]	[0.878]
					{1}	{0.775}
					(1)	(0.867)
D_{SWAT}						1
						[1]
						{1}
						(1)

	Operation		
	Subtraction (-)	Addition (+)	Multiplication (x)
W_0	2 1 3 8 4 5 6 7 9 10	1 2 4 6 5 3 8 9 10 7	1 2 3 5 4 6 7 8 9 10
W_1	2 1 3 8 4 5 6 7 9 10	1 2 4 5 6 3 8 9 10 7	1 2 3 4 5 6 7 8 9 10
A_{SWAT}	2 1 3 4 8 5 6 7 9 10	1 2 4 5 6 3 8 9 10 7	1 2 3 4 5 6 7 8 9 10
PC_C	2 1 3 4 8 5 6 7 9 10	1 2 4 5 6 3 8 9 10 7	1 2 4 3 6 5 7 8 9 10
SWAT	1 2 3 8 4 7 5 6 10 9	5 2 4 6 3 1 9 8 7 10	1 2 3 4 5 6 7 8 9 10
D_{SWAT}	1 2 3 8 4 7 5 6 10 9	2 1 4 5 6 3 9 10 8 7	1 2 3 4 5 6 7 8 9 10

Mental workload increases from left to right.

Underlining denotes that the means are not significantly different at $p < 0.05$.

Figure 4. SNK result for Task Level separated by the different Operations.

	Levels									
	1	2	3	4	5	6	7	8	9	10
W_1	*	c	*	*	*	b	*	a	b	b
W_2	*	a	c	*	*	b	a	a	b	b
A_{SWAT}	*	a	c	d	*	b	a	a	b	b
PC_c	*	a	c	d	*	b	a	a	b	b
SWAT	*	*	*	b	b	b	a	a	b	b
D_{SWAT}	*	a	a	b	b	b	a	a	b	b

* - + x or + - x.

a - + x.

b + - x.

c - + x .

d + - x .

Underlining denotes that the means are not significantly different at $p < 0.05$.

Operations are shown with mental workload increasing from left to right.

Figure 5. SNK result for operation separated by Task Level.

the means were compared (Wierwille and Connor 1983) using the SNK comparison of means test. Here, the scale is considered to be sensitive if it can discriminate between the task levels. The ANOVA and *post-hoc* SNK tests have shown objectively the sensitivity of different scales, however, the sensitivity was not quantified. It could be quantified using factor loadings. Therefore, a factor analysis was performed with W_0, W_1, A_{SWAT}, PC_c, SWAT and D_{SWAT} as dependent variables. This analysis showed the emergence of one factor that could explain 91% of the variation. The factor loading on W_0, W_1, A_{SWAT}, PC_c, SWAT and D_{SWAT} were 0.96938, 0.98469, 0.98496, 0.98492, 0.89673 and 0.90952 respectively. Based on Hill *et al.* (1992), the higher factor scores for A_{SWAT} and PC_c imply that these two scales are more sensitive than the other scales. The SWAT scale has the lowest sensitivity due to the low factor loading. Even D_{SWAT}, using conjoint analysis, had better factor loading than the SWAT scale.

4. Discussion

The variants that were tested are shown in figure 8. The PWC (22.08s) procedure clearly takes significantly less time compared with the CS (476.49 s) procedure. The

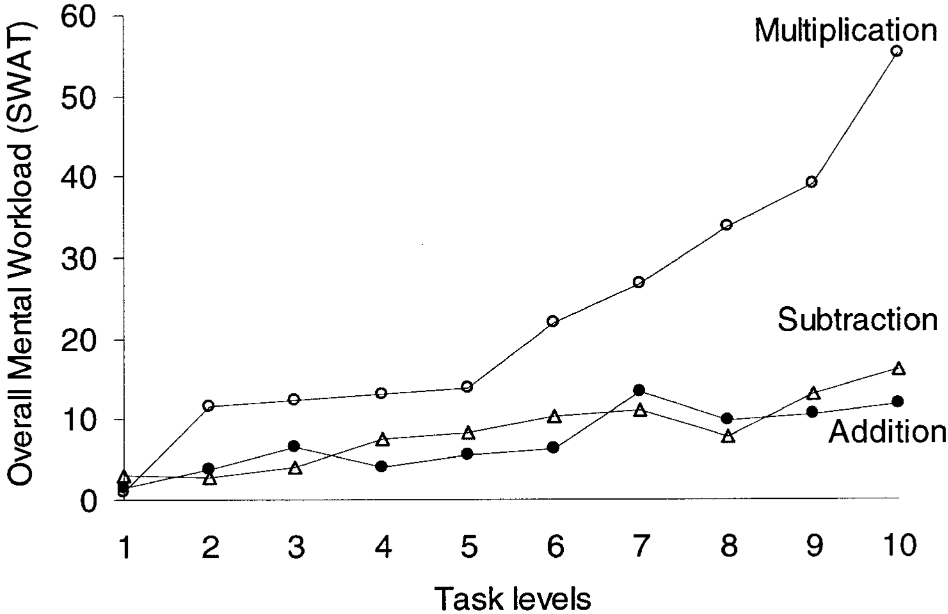

Figure 6. Overall mental workload for each Operation using the SWAT scale.

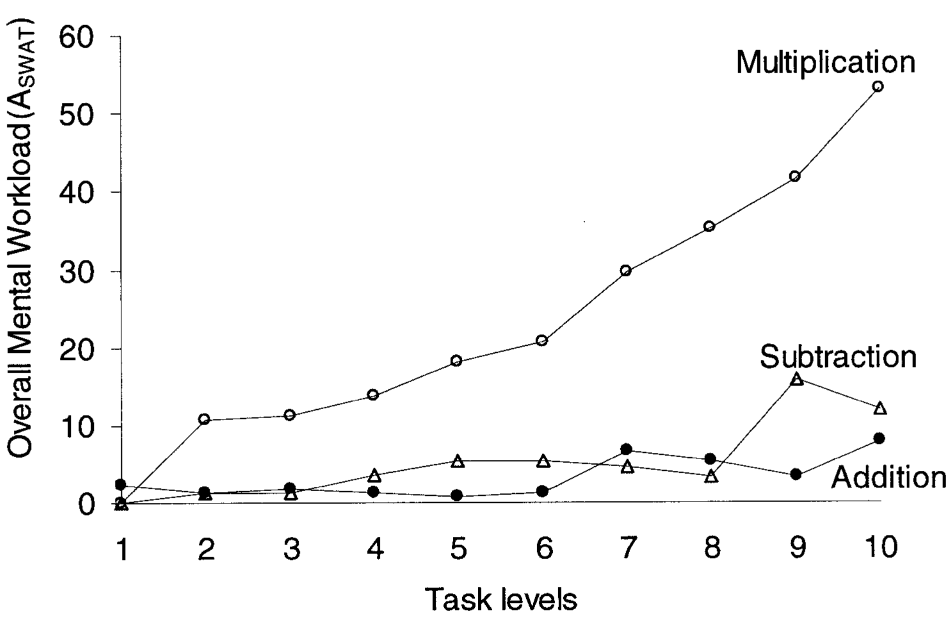

Figure 7. Overall mental workload score for each Operation using the unweighted averaging of continuous SWAT dimensions.

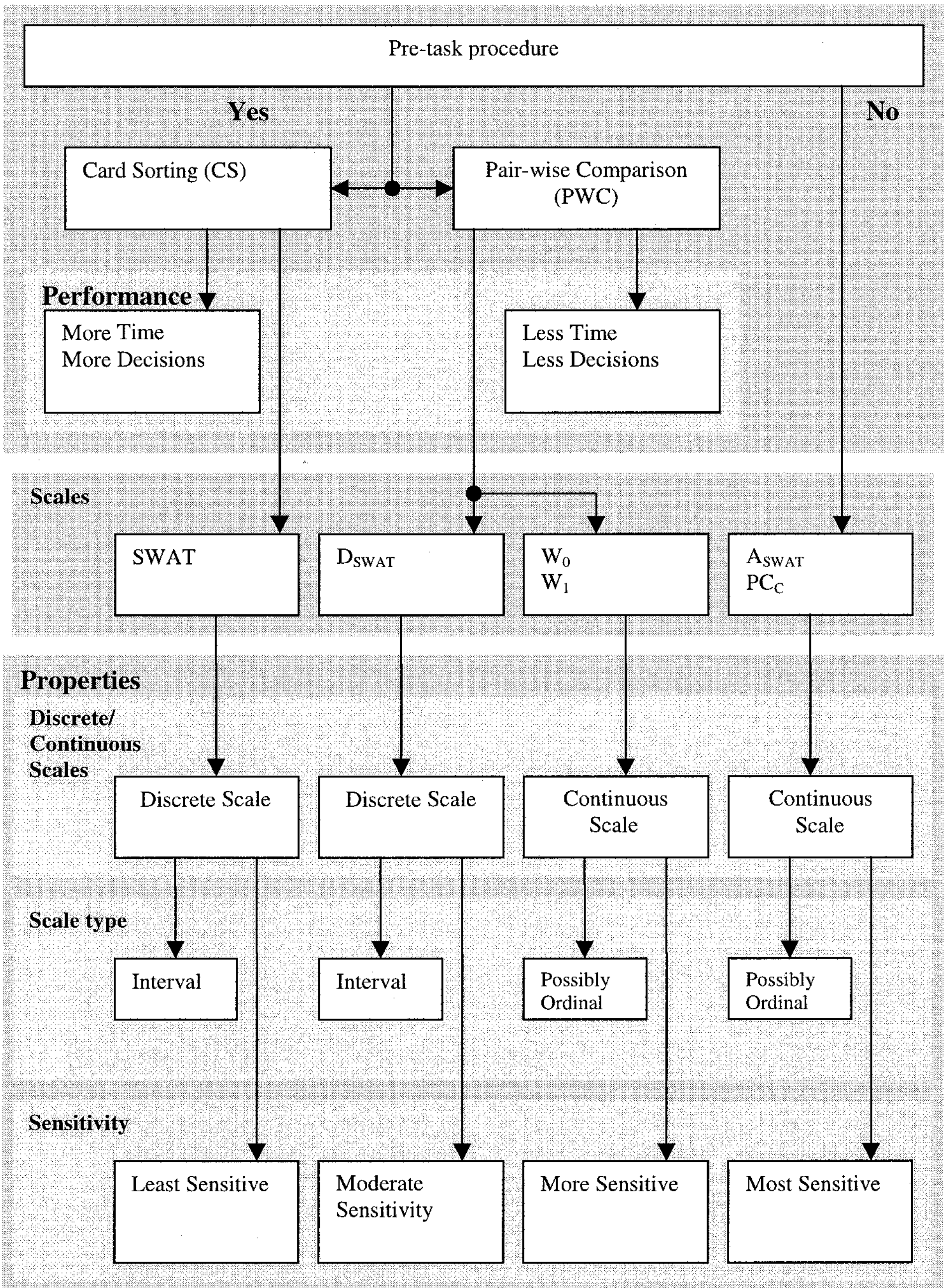

Figure 8. Properties of SWAT and its proposed variants.

CS procedure required more decisions to be made compared with the PWC procedure, and hence it may be error prone. The card sorting procedure however, even though time-consuming, could be used to develop individual workload scales. Otherwise a subgroup and group scale could be developed when a PWC or no pretask procedure was used respectively. When the PWC is used, conjoint analysis can generate the six hypothetical subgroups (TES, TSE, ETS, EST, STE, SET). Increasing the number of levels in the SWAT scale will generate more combinations of SWAT cards, making the CS more time-consuming and error prone (Nygren 1991). As the PWC procedure is simpler, it provides a further opportunity to increase the SWAT scale sensitivity by increasing the number of levels.

The SWAT scale dimensions are correlated as shown in table 5. For the discrete SWAT scale, these correlations were 0.63 between T and E, 0.59 between T and S, and 0.54 between S and E. These values are similar to those of Hart and Staveland (1988) who after several studies obtained intercorrelations of greater than 0.65, 0.5 and 0.45 between T and E, T and S, and S and E respectively. Thus, it appears that the conjoint analysis assumptions have been violated. However, these values have little or no bearing as pointed out by Reid and Nygren (1988) since it is perceptual independence (i.e. dimensions being perceived to be independent) that is important (rather than statistical independence) if an additive model of conjoint scaling is to be used. As indicated by Hart and Staveland (1988) the statistical correlation has a positive effect as it helps to enhance the overall workload.

In general, with the increasing difficulty of the task, the participant's estimate of the mental workload increased. Although, this is a general trend, continuous scales were more sensitive than the discrete scales. Arithmetic tasks used in various mental workload studies (Kahneman *et al.* 1969, Humphrey and Kramer 1994, Ghiaseddin 1995) were used to show this property. Since these studies have shown a change in mental workload (subjectively and objectively) when doing mental arithmetic tasks involving addition, subtraction or multiplication, it was likely that there were differences in workload among the different tasks. With a heavy mental workload (multiplication tasks), there was high correlation among all SWAT variants. This was not the case at low mental workloads (addition and subtraction tasks) where the discrete SWAT scales were less sensitive as shown in the literature.

The ability to discriminate the different task levels was used as one measure to check the sensitivity of the different scales. Figure 3 shows that SWAT is the most insensitive scale among the six variants. For example, it cannot discriminate task level 9 from level 10. Using similar reasoning, we could find the relative sensitivities of the different scales using figures 4 and 5. However, this method is quite cumbersome. Instead, an objective factor loading method was used to rank the different SWAT scales in terms of sensitivity. The rankings were SWAT, D_{SWAT}, W_0, W_1, PC_{C}, and A_{SWAT} from least to most sensitive. Even though A_{SWAT} and PC_{C} used simple scaling and no pretask procedure, their sensitivity was greater than W_0 and W_1, which required a pretask procedure and used a scaling procedure similar to the NASA-TLX scale. This is in agreement with previous literature. The weight generated using principal component and averaging is better than the weighting techniques used for NASA-TLX (Hendy *et al.* 1993). In fact, any set of weights could be used for scaling (Nygren 1991).

Although the sensitivity of discrete scales (SWAT and D_{SWAT}) is weaker than continuous scales, these scales possess an interval property due to the use of conjoint scaling. Furthermore, the sensitivity of D_{SWAT} could be improved by increasing the SWAT levels in each dimension. Depending on different task situations, different SWAT variants could be used. For example, if a sensitive scale is needed, then the pretask procedure can be skipped and the unweighted average of the three dimensions can be used to generate a measure of the overall mental workload.

5. Conclusion

Correlation and factor analysis results showed that all six SWAT scales were somewhat related especially for medium mental workload tasks (multiplication). For low mental workload tasks (addition and subtraction), sensitivity analysis using factor loading and comparison of means showed that the original SWAT scale is not the best. The A_{SWAT} with continuous dimensions had the highest sensitivity. Even though the D_{SWAT} scale was not as sensitive as the continuous scales, it may still have some advantages due to the interval property in the scale as a result of conjoint scaling. The D_{SWAT} scale however, was better than the original SWAT scale in terms of sensitivity and pretask procedural time. Furthermore, the sensitivity of the D_{SWAT} scale could be improved by increasing the number of levels in each of the SWAT dimensions.

In this experiment, simple arithmetic tasks were used to determine the differences among the variations of the SWAT scale. Further validation may be needed with tasks of varying workloads to confirm the above findings.

References

Biers, D. W. and Masline, P. J. 1987, Alternative approaches to analysing SWAT data. In *Proceedings of the 31st Annual Meeting of the Human Factors Society* (Santa Monica: Human Factors and Ergonomics Society), 63–66.

Biers, D. W. and McInerney, P. 1988, An alternative to measuring workload: use of SWAT without the card sort. In *Proceedings of the 32nd Annual Meeting of the Human Factors Society* (Santa Monica: Human Factors and Ergonomics Society), 1136–1139.

Casali, J. G. and Wierwille, W. W. 1983, A comparison of rating scale, secondary-task, physiological, and primary-task workload estimation techniques in a simulated flight task emphasising communications load, *Human Factors*, **25,** 623–641.

Eggemeier, F. T. 1988, Properties of workload assessment techniques. In P. A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Amsterdam: North-Holland), 41–62.

Ghiaseddin, R. R. 1995, The direct scaling of apparent mental workload and its comparison across different arithmetic tasks. Unpublished PhD thesis, University of Notre Dame, Indiana.

Hart, S. G. and Staveland, L. E. 1988, Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Amsterdam: North-Holland), 139–183.

Hendy, K. C., Hamilton, K. M. and Landry, L. N. 1993, Measuring subjective workload: When is a one scale better than many?, *Human Factors*, **35,** 579–601.

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklad, A. L. and Christ, R. E. 1992, Comparison of four subjective workload rating scales, *Human Factors*, **34,** 429–439.

Humphrey, D. G. and Kramer, A. F. 1994, Toward a psychophysiological assessment of dynamic changes in mental workload, *Human Factors*, **36,** 3–26.

Jex, H. R. 1988, Measuring mental workload: problems, progress, and promises. In P. A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Amsterdam: North-Holland), 5–39.

Kahneman, D., Tursky, B., Shapiro, D. and Crider, A. 1969, Pupillary, heart rate and skin resistance changes during a mental task, *Journal of Experimental Psychology*, **79,** 164–167.

McBurney, D. H. 1994, *Research Methods*, 3rd edn (Pacific Grove: Brooks/Cole).

Meshkati, N. 1988, Toward the development of a cohesive model of workload. In P. A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Amsterdam: North-Holland), 305–314.

Meshkati, N., Hancock, P. A., Rahimi, M. and Dawes, S. M. 1995, Techniques in mental workload assessment. In J. R. Wilson and E. N. Corlett (eds), *Evaluation of Human Work: A Practical Ergonomics Methodology*, 2nd edn (London: Taylor & Francis), 749–782.

Meshkati, N. and Loewenthal, A. 1988, An eclectic and critical review of four primary mental workload assessment methods: a guide for developing a comprehensive model. In P. A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Amsterdam: North-Holland), 251 – 268.

Nygren, T. E. 1991, Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload, *Human Factors*, **33,** 17 – 33.

Owen, R. S. 1992, Consumer mental workload: meaning and measurement, Unpublished PhD thesis, Ohio State University, Ames.

Reid, G. B. and Nygren, T. E. 1988, The subjective workload assessment technique: a scaling procedure for measuring mental workload. In P. A. Hancock and N. Meshkati (eds), *Human Mental Workload* (Amsterdam: North-Holland), 185 – 218.

Stassen, H. G., Johannsen, G. and Moray, N. 1990, Internal representation, human performance model and mental workload, *Automatica*, **26,** 811 – 820.

Wierwille, W. W. and Casali, J. G. 1983, A validated rating scale for global mental workload measurement applications. In *Proceedings of the 27th Annual Meeting of the Human Factors Society* (Santa Monica: Human Factors and Ergonomics Society), 129 – 133.

Wierwille, W. W. and Connor, S. A. 1983, Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator, *Human Factors*, **25,** 1 – 16.