DOCUMENT RESUME

ED 446 117		TM 031 878

AUTHOR	Yeung, Alexander Seeshing; Lee, Cynthia Fong King; Pena, Isabel Maria; Ryde, Jeni
TITLE	Toward a Subjective Mental Workload Measure.
PUB DATE	2000-01-00
NOTE	16p.; Paper presented at the International Congress for School Effectiveness and Improvement (Hong Kong, China, January 4-8, 2000).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Cognitive Processes; *College Students; *Construct Validity; Foreign Countries; Higher Education; *Prediction
IDENTIFIERS	Australia; *Cognitive Load; Confirmatory Factor Analysis; Hong Kong; Workload

ABSTRACT

The lack of a conceptually and psychometrically strong measure of human mental workload has undermined interpretations of research based on cognitive load theory. Applying a confirmatory factor analysis approach to construct validation, a Subjective Mental Workload Survey was developed with four distinct subscales: Difficulty, Incompetence, Affect, and Effort. Participants were 52 English-speaking students in Australia and 12 Chinese-speaking students in Hong Kong. For these participants, there were 383 records of scores. Although Difficulty typically used in previous cognitive load research, reasonably well reflected cognitive load, a Workload measure incorporating Difficulty, Incompetence, and Effort proved to be a better predictor of cognitive load. This Workload factor is applicable to analysis of group differences using statistical procedures such as analysis of variance, a finding that shows that tasks more likely to overload working memory score higher in Workload estimates. An appendix contains a list of variables used in the study and alpha reliability estimates. (Contains 50 references.) (Author/SLD)

Toward A Subjective Mental Workload Measure

Alexander Seeshing Yeung, University of Western Sydney at Macarthur, Australia
Cynthia Fong King Lee, Baptist University, Hong Kong
Isabel Maria Pena, and Jeni Ryde, University of Western Sydney at Macarthur, Australia

Paper presented at the International Congress for School Effectiveness and Improvement, 4-8 January 2000, Hong Kong. This research is financially supported by the Faculty of Education and Languages, University of Western Sydney at Macarthur, Australia.

Toward A Subjective Mental Workload Measure

Alexander Seeshing Yeung, University of Western Sydney at Macarthur, Australia
Cynthia Fong King Lee, Baptist University, Hong Kong
Isabel Maria Pena, and Jeni Ryde, University of Western Sydney at Macarthur, Australia

Paper presented at the International Congress for School Effectiveness and Improvement, 4-8 January 2000, Hong Kong. This research is financially supported by the Faculty of Education and Languages, University of Western Sydney at Macarthur, Australia.

Abstract

Recent research has shown the effects of working memory overload on complex mental tasks and the importance of considering cognitive load in instructional design. However, the lack of a conceptually and psychometrically strong measure of human mental workload has undermined interpretations of research based on cognitive load theory. Applying a confirmatory factor analysis approach to construct validation, a Subjective Mental Workload Survey is developed with four distinct scales: Difficulty, Incompetence, Affect and Effort. Although Difficulty typically used in previous cognitive load research reasonably well reflected cognitive load, a Workload measure incorporating Difficulty, Incompetence and Effort proved to be a better predictor of cognitive load. This Workload factor is applicable to analysis of group differences using statistical procedures such as analysis of variance which shows that tasks more likely to overload working memory score higher in Workload estimates.

Working memory limitations can have an important impact on learning in complex areas (e.g., Just & Carpenter, 1992; Paas & Van Merrienboer, 1994; Sweller & Chandler, 1994). Because human working memory is limited, when information input exceeds working memory capacity, the *cognitive load* imposed by such information is unduly high (Sweller, 1993). Cognitive load (sometimes known as mental workload) has been a serious concern in various fields where simultaneous processing of a large amount of information is inevitable. In education, cognitive load has been identified as an important factor to be considered in instructional design in science (e.g., Chandler & Sweller, 1991), geometry (e.g., Paas & Van Merrienboer, 1994; Mousavi, Low, & Sweller, 1995), technical instructions (e.g., Sweller, Chandler, Tierney, & Cooper, 1990; Chandler & Sweller, 1996), computer studies (Decroock, Van Merrienboer, & Paas, 1998; Mayer & Moreno, 1998), and statistics (Paas, 1992). Studies based on cognitive load theory (Sweller, 1993; Yeung, 1999; Yeung, Jin, & Sweller, 1998) have found that the effectiveness of instructional design is at least partly dependent on its ability to manage cognitive load. However, a major limitation in cognitive load research lies in the lack of a psychometrically sound measure of cognitive load. Thus interpretations of instructional effectiveness on the basis of cognitive load management are sometimes unclear. This study aims at establishing and validating a theoretically strong and psychometrically sound measure of cognitive load.

The establishment of a sound psychological measure of mental workload is important because it it is crucial for the understanding of cognitive load and its effects on learning. Experimental studies in education and other areas such as management and aviation have often interpreted changes in performance based on cognitive theories. However, without an accurate measure of cognitive load, such interpretations can be challenged. Thus performance changes may not be due to cognitive load management because any experimental result based on performance changes "does not preclude alternative interpretations of the results" (Yeung, 1999, p. 213). Methodological advances in the assessment of cognitive load is therefore seriously needed in various fields where human mental workload and working memory limitations form the basis of investigation.

Approaches to Mental Workload Measurement

An evaluation of any cognitive load intervention requires a strong measure of cognitive load. A strong measure of cognitive load should be a measure with proven validity and reliability. Support for the reliability of such a measure would require multiple indicators that are closely correlated. Support for the validity of such a measure would require a clear relationship between the psychological measure and a criterion measure of human working memory overload. Although various measures have been used in cognitive load research, such a strong psychological measure of cognitive load does not exist.

Despite the considerable impact of cognitive load on learning, the magnitude of cognitive load in specific tasks cannot be readily evaluated. Because mental resources are not directly measurable, instead, researchers have attempted to measure the mental effort a learner puts into a learning task. This may be measured in three major ways: (a) behavioral--measures of performance as a function of cognitive load, (b) subjective--conscious perception of cognitive load involved in a task, and (c) physiological--physiological changes in respect to cognitive load (Gopher, 1994). Physiological measures include physiological changes in the brain, metabolic and cardiorespiratory changes, heart rate and cortisol values (e.g., Backs & Seljos, 1974; Grasby et al., 1994; Humphrey & Kramer, 1994; Itagaki, Niwa, Itoh, & Momose, 1995; Lebedev, 1994; Paulesu, Frith, & Frackowiak, 1993; Sovcikova & Bronis, 1990; Veltman & Gaillard, 1993; Zipser, Kehoe, Littlewort, & Fuster, 1993), but they cannot be administered to a large group at the same time. Thus some other researchers suggested methods for assessing subjective mental effort (e.g., Bi & Salvendy, 1994; Hendy, Hamilton, & Landry, 1993; Paas & Van Merrienboer, 1993). For example, Moray (1979) suggested that mental effort may be inferred from responses to a Likert scale which probes the learners' perceived difficulty of a learning task. The advantage of a psychometric measure over a physiological measure is its cost-effectiveness, ease of administration, and feasibility in most settings (e.g., administering to a large class with little cost).

Psychological Mental Workload Measure

One popular example of subjective ratings for assessing mental effort is Paas and Van Merrienboer's (1993, 1994) evaluation of instructional efficiency. They used a rating scale to obtain the learners' perception of difficulty of a task and they measured their performance on it. They then used both the performance and perceived difficulty scores in their analysis of instructional efficiency. Paas and Van Merrienboer (1994) found that the learners' subjective rating of difficulty was a more sensitive and reliable technique to assess mental effort than the physiological measurement of changes in heart rate. Since then, many researchers have adopted this combined use of behavioral and subjective cognitive load measures.

However, as Paas and Van Merrienboer (1993) have pointed out, it is not always possible to accurately interpret the efficiency scores particularly when an analysis of variance does not show a statistically significant difference between the efficiency of different instructional methods. Most handicapping in their mental effort measure is that they used only one item, thus making it impossible to control for unreliability. Furthermore, it may not be clear whether Paas and Van Merrienboer's mental effort measure should be equivalent to a measure of cognitive load without controlling for the individual's motivational factors such as sustained effort and self-perceptions of competence. One may argue that cognitive load is not only reflected in the task difficulty or mental effort but also how much a learner attends to the task, feels competent and has positive perceptions about the task. Thus according to Leplat (1978), Tulga (1978), and Yoshitake (1971) a subjective measure of cognitive load should also consider factors such as personality, motivation, confidence, concentration and willingness to think in addition to perceptions of task difficulty (see Moray, 1982 for a review). Thus motivation (Ames, 1992; McInerney, Roche, McInerney, & Marsh, 1997) and self-concept (Marsh & Yeung, 1997a, b; Yeung & Lee, 1999) that may provide the major driving forces that lead to behavioral outcomes should be accounted for in the measurement of mental workload.

Conceptually, task difficulty per se may not really reflect the tendency of working memory overload in a task. Perhaps Paas and Van Merrienboer's (1993) use of terms such as difficulty and mental effort interchangeably may need some clarification. A difficult task does not necessarily lead to greater mental effort. Unless the individual is interested in the task and wants to do it well, task difficulty would lead the individual to give it up instead of trying to invest further effort. Thus conceptually, a task with high cognitive load should be a task that is perceived to be difficult and cannot be done well even though the individual likes the task and invests a high level of effort in it. In sum, a more accurate measure of cognitive load should include these crucial motivational factors examined with a strong construct validity approach such as confirmatory factor analysis which is used here.

Conceptualization and Operationalization of a Mental Workload Construct

To develop a psychological measure of human mental workload, a Subjective Mental Workload Survey (SMWS) instrument assessing four psychological factors was developed: perceived difficulty of task (Difficulty), perceived incompetence (Incompetence), negative affect toward the task (Affect), and lack of effort (Effort). For establishing the reliability of these factors, four items were used in each scale. In relating these psychological constructs to a criterion measure which reflects working memory overload, a simple copying task was devised. Specifically, participants were asked to copy a simple sentence from an overhead projection on the wall. They were timed and the number of gazes required to accomplish the task was counted. Because copying a string of characters requires temporary storage of only a manageable amount of characters or *chunks* of information (Miller, 1956) before reproducing them on paper, the total number of gazes required for copying the sentence would yield a performance score (Gaze) which reflects the extent to which working memory is overloaded. Confirmatory factor analysis was conducted with the survey scores to test the construct validity of the scales and to examine the relation between the constructs with the performance score. Structural equation modeling further examined which of the four factors would best predict the performance score. Whereas the inclusion of multiple indicators would help control for measurement errors in confirmatory factor analysis, great caution was taken here when using time on task as an indicator for the criterion measure of mental workload. In the present study, the relations between the psychological and criterion measures were first tested with the single-item Gaze factor and then with a factor inferred from both time and gaze. To the extent that the patterns of correlations are similar, subsequent analysis would use the two-indicator criterion measure; otherwise analyses would be based on the Gaze measure alone.

On the one hand, on the basis of the literature (e.g., Leplat, 1978; Tulga, 1978; Yoshitake, 1971), the four a priori factors should each predict the performance measure reasonably well. On the other hand, on the basis of Paas and Van Merrienboer's (1993) findings, Difficulty would be expected to be the strongest predictor among other factors. The present study proposes a more stringent test of these hypothesis. In essence, a reasonable reflection of human mental workload would require the consideration of a combination of psychological factors. Whereas a task which tends to overload working memory would be expected to be reflected through higher perceived difficulty (D), lower perception of competence (I), lower affect toward the task (A), and resistance to investing undue effort (E), cognitive load in a task essentially means incompetence in accomplishing the task which is perceived to be difficulty even when the individual is fond of the task--$(D\text{x}I)/A$--and is putting in a considerable amount of effort—$(D\text{x}I)/E$. The present design specifically scrutinizes this hypothesis by introducing an interaction term involving these four measures (called Workload hereafter) and examining the predictive power of this interaction term relative to the individual factors suggested in the existing literature.

From a practical perspective, potential contributions of the SMWS instrument depend on its applicability to analysis of experimental data using standard statistical procedures such as analysis of variance (ANOVA) to assess group differences. To scrutinize the applicability of the SMWS measures, ANOVAs were conducted to examine group differences. Support for the applicability of the instrument requires the results to be in line with cognitive load theory and working memory research.

Method

Participants

The participants were 52 English-speaking students in various courses of languages in a university in Sydney, Australia and 12 Chinese-speaking students in an English proficiency course in a university in Hong Kong. Although all the students speak English, the sample had a diversity of cultural backgrounds. They copied a sentence from each of six language versions and responded to a survey after each copying task, giving a total of 384 records of scores. Because there were only 64 students, the unit of analysis for the present study was based on the individual observations with complete data ($N = 383$), one record having missing data.

Measures

Psychological Mental Workload Measures

The items for each factor are listed in Appendix.

Difficulty. Extending Paas and Van Merrienboer's (1993) perceived difficulty measure, four items asked about perceived difficulty of the task. They were coded such that higher scores reflected greater difficulty.

Incompetence. Adapted from self-concept and self-efficacy instruments, four items asked whether students think they were competent in the task. The responses were coded such that higher scores reflected stronger perceptions of incompetence.

Affect. Adapted from self-concept measures, four items asked whether students like the task. High scores reflected a lack of interest in the task.

Effort. Adapted from measures in motivation research, four items asked whether students invested an effort in the task. Higher scores reflected a resistance to investing effort in the task.

The Copying Task

Each student was asked to copy a simple sentence in each of six language versions projected from an overhead transparency. The languages were German, Italian, English, Spanish and Chinese (simplified version typically used in China) adapted from a manual for an electrical appliance, and a reversed version of English in which the sequence of the same characters in the English sentence was reversed. The difference in number of characters and phonographic representations in various languages restricted direct comparisons between languages but did not affect the focus of the present study on correlations among measures. A simple English sentence was used for practice before the actual tests began.

Criterion Measures

Gaze. The number of gazes for completion of each copying task formed a measure of the extent of working memory overload. The administrator of the test recorded the total gazes required for each task on a record sheet. This score formed a single-item Gaze scale.

Time. The amount of time in seconds for completion of each copying task was recorded. The inclusion of time on task provided the possibility of multiple indicators for the criterion measure. However, the use of this indicator would depend on how well it correlates with the Gaze measure and the similarity of patterns of correlations with the other constructs.

Procedure

The procedure was explained to each student who signed an informed consent form before the tasks began. When the student was ready, the teacher turned on the overhead projector showing the sentence to be copied on a white screen placed such that the student had to look up to read it. The first trial was an English sentence but the number of gazes was not recorded. The student was reminded to look at his or her own writing and not the screen when copying the sentence on paper. After the practice, the copying tasks were administered in six languages in the following order: German, Italian, English, Spanish, English reversed and Chinese. The first time the student looked up to read the sentence was counted 0 and subsequent gazes were counted from 1. Upon completion of the copying task, the class teacher recorded the total number of gazes and the total time spent in seconds on a record sheet and the student completed a survey with the 16 response items. The same procedure was followed for all six languages.

Statistical Analysis

Preliminary Analysis

The correlation between a Difficulty item (Item 2 in the Difficulty scale in Appendix) and the criterion measure was examined. This correlation reflects how well the Paas and Van Merrienboer (1993) Difficulty measure using a single item is related to working memory load and provides a basis for comparison with the SMWS measures. In preliminary analysis the internal consistency of each SMWS measure was evaluated.

Table 1

Goodness-of-fit Summary of Models

Model	χ^2	df	TLI	RNI	Remark
A. Models based on 4 factors					
1. 1 factor	1175.87	104	.664	.709	16 variables form 1 factor
2. 4 SMWS factors	276.66	98	.941	.951	4-factor measurement model
3. 4 SMWS factors + Gaze	298.76	110	.941	.952	5-factor measurement model
4. 4 SMWS factors + Criterion	325.16	125	.943	.954	2 indicators for Criterion
5. 4 factors predict Criterion	325.16	125	.943	.954	D, I, A, E predict Criterion
6. Difficulty predicts Criterion	347.75	128	.939	.949	D alone predicts Criterion
7. 2 factors predict Criterion	345.62	127	.939	.949	D and I predict Criterion
8. 3 factors predict Criterion	325.20	126	.944	.954	D, I, and E predict Criterion
B. Models including subjective mental Workload measure					
9. 6 factors no correlated unique	1700.74	194	.753	.793	6-factor measurement model
10. 6 factors 16 correlated unique	493.02	178	.943	.957	22 variables form 6 factors
11. 5 factors predict Criterion	493.02	178	.943	.957	All factors predict Criterion
12. 4 factors predict Criterion	526.13	179	.938	.952	D, I ,E, A predict Criterion
13. 3 factors predict Criterion	526.38	180	.939	.952	D, I and A predict Criterion
14. Difficulty predicts Criterion	550.24	182	.936	.949	D alone predicts Criterion
15. Workload predicts Criterion	506.36	182	.943	.955	Workload alone as predictor

Note. $N = 383$. TLI = Tucker-Lewis index. RNI = Relative noncentrality index. Unique = Uniquenesses. The χ^2(df) value of the null model for Models 1 and 2 testing 4 factors--Difficulty (D), Incompetence (I), lack of Affect (A), and lack of Effort (E)--is 3803.16(120), that for Model 3 is 4103.39(136), that for Models 4 to 9 is 4460.33(153), and that for Models 10 to 16 is 7501.50(231). The Workload measure is operationalized as (DxI/A)+(DxI/E). All models converged to proper solutions.

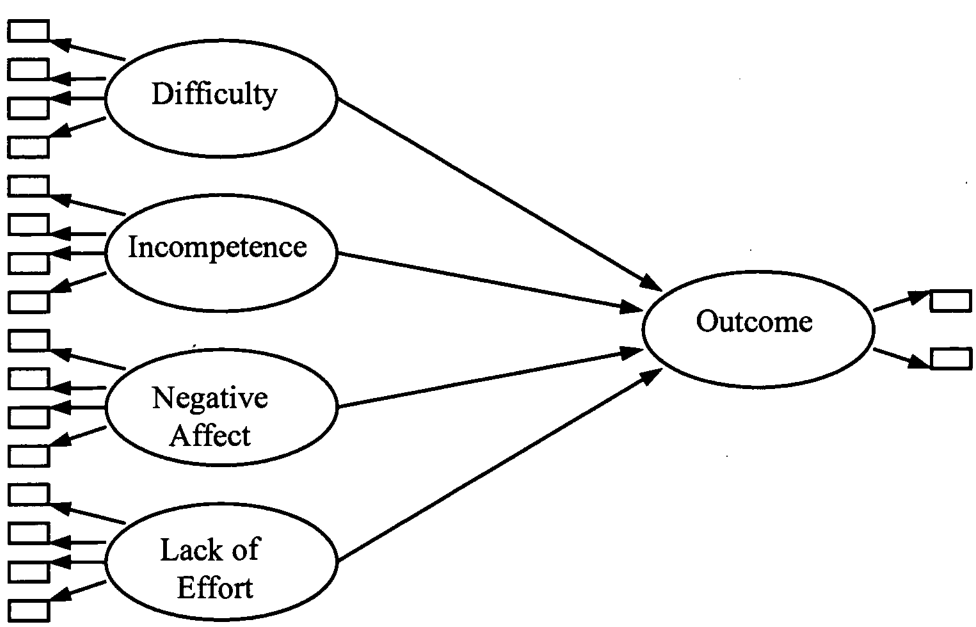

Figure 1. (a) Four SMWS factors predicting the outcome.

Confirmatory Factor Analysis and Structural Equation Models

Applying confirmatory factor analysis (CFA), each scale was first tested to fit a single-factor congeneric model (Joreskog & Sorbom, 1989). A series of CFA and structural equation models (SEM) were tested. They are presented in two sections in Table 1. In the first section (Section A), validity of the survey items was tested in a single-factor model (Model 1) and a four-factor model (Model 2). Model 3 included a single-item Gaze criterion variable to examine its relations with the four a priori psychological factors. Model 4 attempted to replicate Model 3 except that Model 4 used two indicators (gaze and time) for the criterion measure. To the extent that Model 4 provided a reasonable fit to the data and similar correlations among constructs as Model 3, subsequent structural equation models would be based on Model 4. Models 5 to 8 were a series of nested models that tested the relative ability of each of the factors in predicting the mental workload criterion as an outcome variable (Figure 1).

In the second section (Section B), a new factor was added to the previous models. The additional factor was derived from the four factors tested in Section A such that Workload = $(DxI)/A + (DxI)/E$, where D was Difficulty, I was Incompetence, A was negative Affect, and E was Lack of Effort. Thus higher Difficulty and Incompetence scores coupled with higher Effort scores, that is $(DxI)/E$, together with higher Difficulty and Incompetence coupled with lower Negative Affect, that is $(DxI)/A$, would result in a higher value presumably reflecting higher mental workload. Because there were four items for each of the Difficulty, Incompetence, Affect and Effort scales, the formation of four indicators for this new Workload scale was straight-forward (Yang Jonsson, 1998). For example, the first item from the Difficulty scale was multiplied by the first item from Incompetence divided by the first item from Effort, resulting in $(DxI)/$. Then the first item from the Difficulty scale was multiplied by the first item from Incompetence divided by the first item from Affect, resulting in $(DxI)/A$. Finally, $(DxI)/E$ and $(DxI)/A$ were added together to form the first Workload indicator. The same procedure was followed for the other three indicators for the Workload factor.

The previous four scales, the newly formed Workload scale, and the criterion measure were first tested in measurement models positing six factors (Models 9 and 10). A comparison between Models 9 and 10 tested whether correlated uniquenesses (correlations between the disturbance terms of items) needed to be included for model fit. Because the new Workload scale was derived from the other four scales, correlated uniquenesses were expected to be required between the Workload scale and each of the scales from which it was derived. To the extent that the six factors provided a reasonable fit to the data, then the relative ability of each of the five factors in predicting the criterion outcome was examined in subsequent SEM models (Models 11 to 15). Support for the SMWS scales requires good reliability and substantial factor coefficients for each of the scales and distinctiveness of each factor from other factors. A strong predictor would have a high correlation with and substantial path coefficient to the criterion measure.

The conduct of CFA and SEM has been described elsewhere (e.g., Bollen, 1989; Byrne, 1989, 1998; Joreskog & Sorborm, 1993; Marsh, 1994; Pedhazur & Schmelkin, 1991) and is not further detailed here. All analyses throughout this paper were conducted with the SPSS version of LISREL (Joreskog & Sorbom, 1989). The goodness of fit of models is evaluated based on suggestions of Marsh, Balla, and McDonald (1988) and Marsh, Balla, and Hau (1996) with an emphasis on the Tucker-Lewis index (TLI), but we present also the chi-square test statistic and the relative noncentrality index (RNI). A model is typically considered to fit the data when TLI > .9 but valuable information is often obtained by comparing competing models especially when the models are nested within each other. In comparing nested models, the more parsimonious model (that which has less estimated parameters) is typically favored if it fits as well as a competing model with more estimated parameters. This can be done by comparing their TLI values and by examining the difference between the χ^2 values of the models relative to the difference between their *df* values (McDonald & Marsh, 1990).

<u>Application in Mean Comparisons</u>

For application purposes, especially in experimental studies with small samples, it is not always possible to apply a CFA and SEM approach. Thus, it is important to demonstrate the applicability of the subjective mental workload measure in mean comparisons with small group sizes. Multivariate analysis of variance (ANOVA) was conducted to compare the differences in number of gazes, time, and the Workload measure between the copying task in Chinese by Chinese-speaking Hong Kong students and by non-Chinese-speaking Australian students. Because the between-group differences for the criterion measures of gaze and time are known (i.e., students unfamiliar with a non-alphabetic language would surely require more gazes and time for copying Chinese), the applicability of the Workload measure in differentiating the mental workload required for each condition could be tested. Support for the applicability of the Workload measure in analyzing experimental data requires the pattern of results in group differences should be similar across all three dependent variables. To form the Workload measure, the four items in each scale were averaged to obtain an unweighted mean for each scale. Then the values for D, I, A, and E were used to calculate the Workload value, $(D\text{x}I)/A + (D\text{x}I)/E$. Support for the applicability of the Workload measure requires the scores in gaze, time and Workload to be higher for the non-Chinese speaking students than the Chinese-speaking students when copying a Chinese sentence.

Results

Preliminary Analysis

The alpha reliability for each of the a priori scales was good (see Appendix). One-factor congeneric models for each scale all provided reasonable fit to the data (TLI > .9). These results provided a good foundation for subsequent CFA and SEM models. Before testing CFA models, the correlation between a Difficulty item (Item 2 in the Difficulty scale in Appendix) and the Gaze criterion measure was examined. This correlation (r = .58) partly supported the Paas and Van Merrienboer's (1993) use of a single-item Difficulty measure for working memory load but also showed that this measure could explain only about one-third of the variance. There is the need to develop a stronger mental workload measure.

Construct Validity of the Four Factors

Table 1 presents a summary of the goodness of fit of the models tested in the present study in two sections. In Section A, Model 1 positing a single factor inferred by 16 measured variables did not fit the data (TLI = .664). Model 2 testing the construct validity of a four-factor model with the 16 items provided a good fit to the data (TLI = .941). A comparison between Models 1 and 2 provided preliminary support for the four-factor structure of Model 2. In Model 2, all factor coefficients were statistically significant (from .56 to .85) and the correlations among the four factors were mostly moderate (.31 to .88). Thus the four a priori factors were reasonably distinct from each other. Model 3 including a single-indicator criterion measure of Gaze in a five-factor model showed similar relations among the SMWS factors. The fit of Model 3 was good (TLI = .941). In support of the ability of the SMWS factors to reflect mental workload, the correlations between Gaze and the SMWS factors were all statistically significant (.68, .68, .19, and .12, respectively for Difficulty, Incompetence, Affect, and Effort). In support of Paas and Van Merrienboer (1993), the high correlation between Difficulty and Gaze suggested that Difficulty is likely to reflect mental workload reasonably well. However, the results also showed that Incompetence may be another factor that is similarly able to reflect mental workload.

Model 4 was a replication of Model 3 except that Model 4 included the time for task completion as another indicator of the working memory criterion measure. Because the correlation between the Gaze measure and time on task was reasonably high (r = .76), it was possible to use both these measures to form a mental workload criterion variable. Model 4 provided a good fit (TLI = .943). Like the factor coefficients for the other scales, the factor coefficients for the criterion variable was good (.97 and .79, respectively for the indicators of gaze and time). Because the parameter estimates for Model 4 were very similar to those for Model 3--in particular the correlation between the criterion factor and each of the Difficulty, Incompetence, Affect and Effort factors were very similar (rs = .74, .70, .19, and .12, respectively)--subsequent models were based on Model 4 using two indicators for the criterion measure. It is important to note, however, that the patterns of results are very similar with one or two indicators for the criterion measure although only results based on two indicators are reported.

Which Factors Best Predict the Criterion Measure?

Model 5 posited paths from the four SMWS factors to the criterion measure (Figure 1). Because the number of estimated parameters was the same as that for Model 4, their χ^2 and *df* values, and hence their TLI values, were identical. The critical concern was the relative ability of each factor in predicting the outcome variable. Because the correlation between the criterion measure and each of the SMWS factors was statistically significant (*r*s = .74, .70, .19, and .12, respectively), each of the four SMWS factors were positively related to working memory load. However, the paths from Difficulty, Incompetence, and Effort were statistically significant (.46, .39, and -.22, respectively) whereas the path from Affect was nonsignificant (.01). These results showed that three of the factors were good predictors of working memory load whereas Affect was a relatively weaker predictor. Using Model 5 as a basis for comparison, a series of models were posited to examine the best combination of the factors in predicting the outcome measure. Table 1 presents only some of the critical models. Model 6 positing a path from only the Difficulty factor to the criterion measure while fixing all other possible paths to be zero did not fit as well as Model 5 (TLI values of .939 vs. .943). Model 7 positing paths from Difficulty and Incompetence did not do any better (TLI = .939). Model 8 (TLI = .944) positing paths from Difficulty, Incompetence and Effort provided a fit comparable to Model 5. Between Models 5 and 8, The difference in χ^2 (325.20 - 325.16 = 0.04) relative to their difference in *df* (126 - 125 = 1) was statistically nonsignificant ($p > .05$), thus favoring the more parsimonious Model 8. In Model 8, the paths from Difficulty, Incompetence, and Effort were all statistically significant (.47, .39, and -.22, respectively). These results showed that the psychological measurement of mental workload requires the consideration of more than just the Difficulty factor.

Construct Validity of the Workload Factor

Models 9 and 10 tested the ability of a six-factor model (4 SMWS factors; a Workload factor derived from Difficulty, Incompetence and Effort based on theory; and a criterion measure) to fit the data. Model 10 differed from Model 9 in that Model 10 included 16 correlated uniquenesses. Although the 16 correlated uniquenesses in Model 10 were posited a priori, we present also Model 9 for comparison. Model 10 (TLI = .943) provided a good fit to the data and was much better than Model 9 (TLI = .753). The factor coefficients in Model 10 ranged from .53 to .95. The factor correlations ranged from .01 to .87. In particular, the correlations between the criterion variable and the SMWS and Workload factors were all significant (.75, .71, .17, .12, and .81, respectively). Thus Model 10 formed the basis for subsequent SEM models.

Table 2.
Solution of Model 15

Variable	Difficulty	Incompetence	Affect	Effort	Workload	Criterion
Factor Coefficients						
Item 1	.86*	.79*	.78*	.85*	.87*	.91*
Item 2	.81*	.95*	.62*	.78*	.88*	.83*
Item 3	.71*	.85*	.52*	.72*	.82*	--
Item 4	.66*	.94*	.79*	.54*	.78*	--
Uniquenesses						
Item 1	.25*	.38*	.39*	.28*	.24*	.17*
Item 2	.34*	.11*	.62*	.39*	.23*	.31*
Item 3	.50*	.27*	.73*	.48*	.34*	--
Item 4	.57*	.12*	.38*	.71*	.39*	--
Path Coefficients (from Workload to Criterion)						
Workload	--	--	--	--	--	.83*
Factor Correlations						
Difficulty	--					
Incompetence	.87*	--				
Affect	.31*	.36*	--			
Effort	.35*	.48*	.50*	--		
Workload	.83*	.79*	.09*	.03	--	
Criterion	.69*	.66*	.08*	.02	.83*	--
Residuals	1	1	1	1	1	.31*

Note: $N = 383$. Items for each factor are listed in Appendix. Parameter estimates are completely standardized. The 5 SMWS factors were Difficulty (D), Incompetence (I), Negative Affect (A), Lack of Effort (E) and Workload (W) which was an interaction term, W = (DxI)/E + (DxI)/A. * $p < .05$.

Predictive Power of the Workload Factor

Models 11 to 15 presented in Table 1 are selected models that are critical to the present investigation. Model 11 positing paths from the five factors to the outcome measure served as a basis for comparison with competing nested SEM models. Because the number of estimated parameters in Model 11 was equivalent to Model 10, their TLI values (.943) were equivalent. Among the five paths to the outcome measure, the path from the Workload factor operationalized as *(DxI)/A* + *(DxI)/E* was the only statistically significant path to the outcome measure ($\beta = .64$). The paths from Difficulty, Incompetence, Affect and Effort were all statistically nonsignificant (βs = .19, .02, .04, and .02, respectively). Using Model 11 as a basis for comparison, a series of models were posited to examine the best combination of the factors in predicting the outcome. Model 12 positing paths from the four SMWS factors but not from the Workload factor did not fit as well as Model 11 (TLI values of .938 vs. .943). Model 13 positing paths from only the three strongest SMWS predictors previously found in Model 8 did not fit as well as Model 11 either (TLI = .938). Model 14 positing a path from only Difficulty to the outcome measure did even worse (TLI = .936). Model 15 (TLI = .943) positing a path from only the Workload factor--operationalized as *(DxI)/A* + *(DxI)/E*-- provided a fit comparable to Model 11. The solution of Model 15 is presented in Table 2. A comparison between Model 12 (TLI = .938) using the four SMWS factors to predict the outcome and the more parsimonious Model 15 (TLI = .943) using only the Workload factor to predict the outcome found that Model 15 using only Workload as the predictor provided a better fitting model. Thus the Workload measure derived from Difficulty, Incompetence, Negative Affect and Lack of Effort was able to predict the outcome measure as well as the four SMWS factors considered all together.

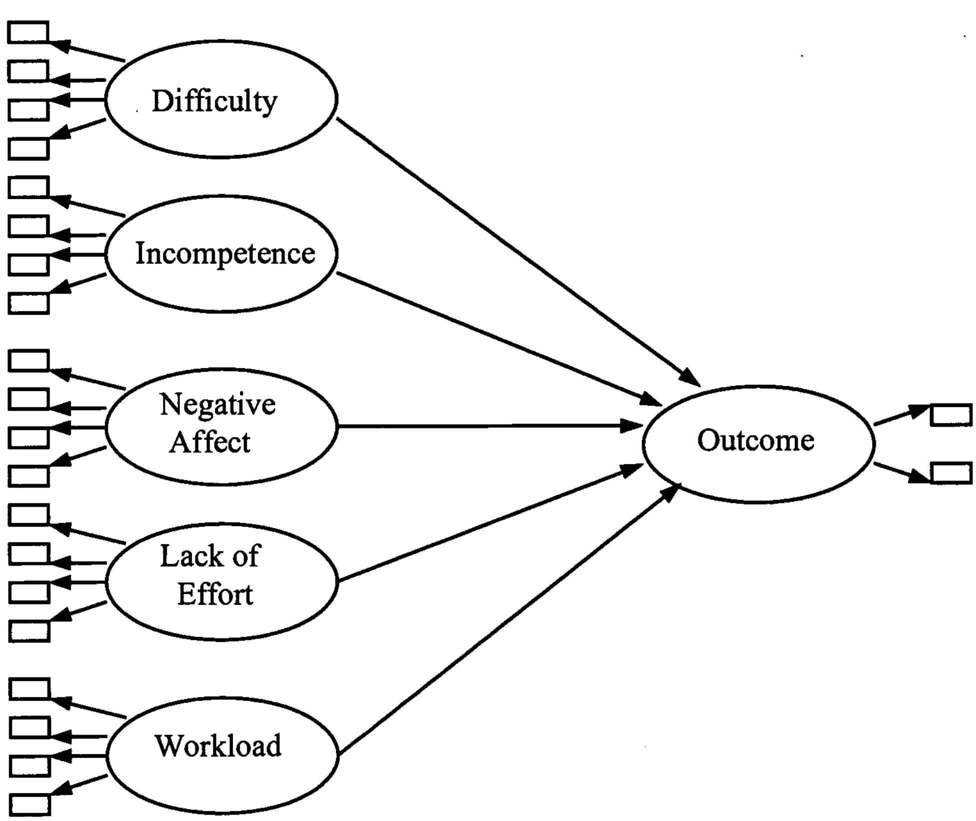

Figure 2. Four SMWS factors and Workload predicting the outcome.

Table 3

Means, Standard Deviations, and Between-group Comparisons

	Chinese (n=12)	Non-Chinese (n=52)	$F(1,62)$		
Gaze	3.67 (2.15)	36.69 (13.82)	67.29**	158.03	.52
Time	21.29 (5.15)	151.46 (87.92)	25.96**	6363.46	.30
Workload	3.82 (1.46)	21.20 (13.76)	18.86**	156.09	.23

Note. Scores were compared between Chinese-speaking students and non-Chinese speaking students. Univariate F-statistics from multivariate ANOVAs are presented.

Applicability of the Workload Measure

Chinese vs. Non-Chinese Speakers. The four items in each scale were averaged to obtain an unweighted mean for the *D*, *I*, *A* and *E* values which were used to calculate the Workload score, $(DxI)/A + (DxI)/E$. The means and standard deviations of gaze, time and Workload scores and univariate F-statistics are presented in Table 3. Chinese-speaking students required significantly fewer gazes and less time, and displayed lower mental workload than non-Chinese speakers when copying a Chinese sentence (all $ps < .001$). The results provided support for the applicability of the subjective mental workload measure in comparisons of group means.

Discussion

The present study attempted to examine an existing cognitive load measure and to develop a stronger measure of cognitive load (human mental workload). The results found that although Paas and Van Merrienboer's (1993) approach of using a task difficulty measure to infer cognitive load may provide a reasonably good estimate, a conceptually and psychometrically stronger measure incorporating measures of task difficulty, sense of incompetence in the task, and effort invested in the task may be a stronger predictor of cognitive load. The advantages of this new measure over the existing task difficulty measure include: (a) established reliability by using multiple indicators for each scale, (b) a conceptually sound reference to the working memory and cognitive load literature, (c) strong convergent and discriminant validity established through a strong construct validation approach and (d) proven relation of the measure to working memory load.

Confirmatory factor analysis supported the construct validity of the four SMWS scales. Each of the scales was significantly correlated with the criterion variable but they differed in their ability to predict it. The construct validation approach allowed a scrutiny of the ability of each factor and combinations of these factors in predicting the outcome measure of working memory load. Whereas three of the factors—Difficulty, Incompetence and Effort—predicted the outcome variable better than the Affect factor, an interaction term combining the four factors into a Workload construct was even better able to predict the outcome. The Workload construct was operationalized on the basis of the conceptual interpretation of human mental workload in that a task that is likely to overload working memory is probably a task perceived to be difficult and hard to accomplish despite considerable effort and despite favorable affect toward the task. With this interpretation, task difficulty (*D*) alone would not be able to reflect working memory load as well as the Workload construct; and neither would Incompetence (*I*), or Affect (*A*), or Effort (*E*), each on its own. Thus it was not surprising that the operationalization of Workload as $(DxI)/A + (DxI)/E$ predicted working memory load better than any of the factors by itself.

From a practical perspective, the advantage of using the Workload measure instead of a Difficulty measure or a combination of the four constructs described in the present investigation is not only that it is a stronger predictor of mental workload but also that it is easy to apply in experimental studies. Methodologically, in comparing the cognitive load involved in different experimental conditions, the same results should be obtained whether cognitive load is measured by considering a combination of Difficulty, Incompetence and Lack of Effort as distinct factors in a regression formula or by using the Workload = $(DxI)/A + (DxI)/E$ measure. However, because the path coefficients varied among the three factors when predicting mental workload, there may be complications as to whether the weighting of each predictor should be adjusted accordingly and whether such weighting would remain valid for different samples. Thus for practical applications, use of the Workload measure seems to have a definite advantage over the use of a regression formula involving the three distinct predictors.

Further research should examine whether the Workload measure introduced in the present study can be used in Paas and Van Merrienboer's (1993) elegant approach to estimating instructional efficiency. Paas and Van Merrienboer used the performance scores in a task and a perceived difficulty measure to estimate the efficiency of an instructional technique such that a high performance score coupled with a low difficulty score would yield a high instructional efficiency estimate. Conceptually, the Workload measure here should not contradict Paas and Van Merrienboer's approach to instructional efficiency assessment when used in the estimate.

Despite increasing support for the impact of cognitive load in instructional efficiency, the lack of a strong measure of cognitive load has caused major limitations in interpreting findings. These limitations have probably undermined the potentials of cognitive load theory in directing instructional design. The ANOVAs in the present study showed that the Workload measure not only was a promising measure reflecting cognitive load but was also applicable to mean comparisons in experimental studies. The measure allows an estimate of cognitive load associated with a mental task such that interpretations of positive effects in cognitive load management can be made when performance scores increase while Workload scores decrease.

In sum, interpretations of research on the effects of working memory overload on complex mental tasks and instructional designs based on cognitive load theory requires a psychometrically strong measure of human mental workload. Due to the lack of a very strong psychological measure of cognitive load, research results interpreted on the basis of performance scores are often undermined. The Subjective Mental Workload Survey may be a stronger alternative to an existing measure of task difficulty assumed to measure mental effort. In search of a good measure of cognitive load, researchers should consider the interaction among factors such as task difficulty, perceived incompetence, invested effort, and perhaps affects toward the task. Because a psychological measure of cognitive load is more cost-effective and more easily administered to large groups compared to physiological measures, further rigorous search for a strong cognitive load measure is worth pursuing. Such measures should be validated though strong methodologies and their applicability should be tested in various complex mental tasks and settings.

References

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. Journal of Educational Psychology, 84, 261-271.

Backs, R. W., & Seljos, K. A. (1974). Metabolic and cardiorespiratory measures of mental effort: The effects of level of difficulty in a working memory task. International Journal of Psychophysiology, 16, 57-68.

Bi, S. X., & Salvendy, G. (1994). A proposed methodology for the prediction of mental workload, based on engineering system parameters. Work & Stress, 8, 355-371.

Bollen, K. A. (1989). Structural equations with latent variables. New York: Wiley.

Byrne, B. M. (1989). A primer of LISREL: Basic applications and programming for confirmatory factor analytic models. New York: Springer Verlag.

Byrne, B. M. (1998). Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming. Mahwah, NJ: Erlbaum.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. Cognition and Instruction, 8, 293-332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. British Journal of Educational Psychology, 62, 233-246.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. Applied Cognitive Psychology, 10, 151-170.

Decroock, M. B. M., Van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). High versus low contextual interference in simulation-based training of troubleshooting skills—Effects on transfer performance and invested mental effort. Computers in Human Behavior, 14, 249-267.

Gopher, D. (1994). Analysis and measurement of cognitive load. In G. d'Ydewalle, P. Eelen, & P. Bertelson (Eds.), International perspectives on psychological science, Vol. 2: The state of the art (pp. 265-291). England: Erlbaum.

Grasby, P. M., Frith, C. D., Friston, K. J., Simpson, J., Fletcher, P. C., Frackowiak, R. S. J., & Dolan, R. J. (1994). A graded task approach to the functional mapping of brain areas implicated in auditory verbal memory. Brain, 117, 1271-1282.

Hendy, K. C., Hamilton, K. M., & Landry, L. N. (1993). Measuring subjective workload: When is one scale better than many? Human Factors, 35, 579-601.

Humphrey, D. G., & Kramer. A. F. (1994). Toward a psychophysiological assessment of dynamic changes in mental workload. Human Factors, 36, 3-26.

Itagaki, F., Niwa, S., Itoh, K., & Momose, T. (1995). Random number generation and the frontal cortex. International Journal of Psychophysiology, 19, 79-80.

Joreskog, K. G., & Sorbom, D. (1989). LISREL 7: A guide to the program and applications. Chicago: SPSS.

Joreskog, K. G., & Sorbom, D. (1993). LISREL 8: Structural equation modeling with the SIMPLIS command language. Chicago: Scientific Software International.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. Psychological Review, 99, 122-149.

Lebedev, A. N. (1994). The neurophysiological parameters of human memory. Neuroscience & Behavioral Physiology, 24, 254-259.

Lee, J. H. (1999). Test anxiety and working memory. Journal of Experimental Education, 67, 218-240.

Leplat, J. (1978). Factors determining workload. Ergonomics, 21, 143-149.

Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. Structural Equation Modeling, 1, 5-14.

Marsh, H. W., Balla, J. R., & Hau, K. T. (1996). An evaluation of incremental fit indices: A clarification of mathematical and empirical processes. In G. A. Marcoulides & R. E. Schumacker (Ed.), Advanced structural equation modeling techniques (pp. 315-353). Hillsdale, NJ: Erlbaum.

Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 102, 391-410.

Marsh, H. W., & Yeung, A. S. (1997a). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. Journal of Educational Psychology, 89, 41-54.

Marsh, H. W., & Yeung, A. S. (1997b). Coursework selection: The effects of self-concept and achievement. American Educational Research Journal, 34, 691-720.

Mayer, R. E., & Moreno, R. (1998). Split-attention effect in multimedia learning—Evidence for dual processing systems in working memory. Journal of Educational Psychology, 90, 312-320.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. Psychological Bulletin, 107, 247-255.

McInerney, D. M., Roche, L. A., McInerney, V., & Marsh, H. W. (1997). Cultural perspectives on school motivation: The relevance and application of goal theory. American Educational Research Journal, 34, 207-236.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.

Moray, N. (1979). Mental workload: Its theory and application. New York: Plenum.

Moray, N. (1982). Subjective mental workload. Human Factors, 24, 25-40.

Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. Journal of Educational Psychology, 87, 319-334.

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. Journal of Educational Psychology, 84, 429-434.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. Human Factors, 35, 737-743.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. Journal of Educational Psychology, 86, 122-133.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. Nature, 362, 342-345.

Pedhazur, E. J. & Schmelkin, L. P. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Erlbaum.

Sovcikova, E., & Bronis, M. (1990). Short-term memory and attention changes related to biorhythm and microclimatic conditions. 25th Conference on Higher Nervous Functions (1989, Olomouc, Czechoslovakia). Activitas Nervosa Superior, 32, 204.

Sweller, J. (1993). Some cognitive processes and their consequences for the organisation and presentation of information. Australian Journal of Psychology, 45, 1-8.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. Cognition and Instruction, 12, 185-233.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load and selective attention as factors in the structuring of technical material. Journal of Experimental Psychology: General, 119, 176-192.

Tulga, M. K. (1978). Dynamic decision making in multitask supervisory control: Comparison of an optimal algorithm to human behavior. Cambridge, MA: MIT Man-Machine Systems Laboratory.

Veltman, J. A., & Gaillard, A. W. (1993). Indices of mental workload in a complex task environment. International Symposium: Psychobiology: Psychophysiological and psychohumoral processes combined (1992, Giessen, Germany). Neuropsychobiology, 28, 72-75.

Yang Jonsson, F. (1998). Modeling interaction and nonlinear effects: A step-by-step LISREL example. In R. E. Schumacker & G. A. Marcoulides (Eds.), Interaction and nonlinear effects in structural equation modeling (pp. 17-42). Mahwah, NJ: Erlbaum.

Yeung, A. S. (1999). Cognitive load and learner expertise: Split-attention and redundancy effects in reading comprehension tasks with vocabulary definitions. Journal of Experimental Education, 67, 197-217.

Yeung, A. S., Jin, P., & Sweller, J. (1998). Cognitive load and learner expertise: Split-attention and redundancy effects in reading with explanatory notes. Contemporary Educational Psychology, 23, 1-21.

Yeung, A. S., & Lee, F. L. (1999). Self-concept of high school students in China: Confirmatory factor analysis of longitudinal data. Educational and Psychological Measurement, 59, 431-450.

Yoshitake, H. (1971). Relation between the symptoms and the feelings of fatigue. Ergonomics, 14, 175-186.

Zipser, D., Kehoe, B., Littlewort, F., & Fuster, J. M. (1993). A spiking network model of short-term active memory. Journal of Neuroscience, 13, 3406-3420.

Appendix

Variables in the Study and Alpha Reliability Estimates

Difficulty $\alpha = .84$

1.@ I think the exercise is easy enough.
2. The exercise was too hard for me.
3.@ I had no problems doing this exercise.
4. I got into trouble when I did the exercise.

Incompetence $\alpha = .93$

1. I did the exercise too badly.
2.@ I think I got everything right.
3.@ I think I did the exercise very well.
4.@ I think I did not make any mistakes.

Negative Affect $\alpha = .78$

1.@ I like this kind of exercise.
2.@ I am very interested in the exercise.
3. I hate doing the same thing again.
4.@ I do not mind doing something similar again.

Lack of Effort $\alpha = .80$

1.@ I paid attention throughout the exercise.
2.@ I did the exercise seriously.
3.@ I tried to get everything right in the exercise.
4.@ I worked hard to do the exercise.

Criterion Measure

Gaze Number of gazes

Time Time in seconds taken for completion of task

Note: @ These items were reverse coded. The items were in a random order in the survey.

2 copies included

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM031878

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: TOWARDS A SUBJECTIVE MENTAL WORKLOAD MEASURE

Author(s): Alexander Seeshing Yeung, Cynthia Fong King Lee, Isabel Maria Pena, Jeni Ryde

Corporate Source:

Publication Date: January 2000

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: M. Isabel Pena

Printed Name/Position/Title: Mrs Isabel Maria PENA - Lecturer

Organization/Address: UNIVERSITY OF WESTERN SYDNEY MACARTHUR
P.O. Box 555, CAMPBELLTOWN, NSW 2560
AUSTRALIA

Telephone: 61 2 9772 6304

FAX: 61 2 9772 1565

E-Mail Address: ~~I.~~ I.Pena@uws.edu.au

Date: 6/1/00