Ergonomics
Vol. 50, No. 4, April 2007, 514–519

Taylor & Francis
Taylor & Francis Group

A self-analysis of the NASA-TLX workload measure

JAN M. NOYES* and DANIEL P. J. BRUNEAU

Department of Experimental Psychology, University of Bristol,
8 Woodland Road, Bristol BS8 1TN, UK

Computer use and, more specifically, the administration of tests and materials online continue to proliferate. A number of subjective, self-report workload measures exist, but the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) is probably the most well known and used. The aim of this paper is to consider the workload costs associated with the computer-based and paper versions of the NASA-TLX measure. It was found that there is a significant difference between the workload scores for the two media, with the computer version of the NASA-TLX incurring more workload. This has implications for the practical use of the NASA-TLX as well as for other computer-based workload measures.

Keywords: Subjective workload; NASA-TLX workload measure; Computer vs. paper administration; Absolute vs. relative workload measures

1. Introduction

Since its launch in 1978, the personal computer has become a familiar fixture in the workplace and in homes. Consequently, an increasing number of paper-based tasks are now being transferred onto computers. In this respect, the paper-based National Aeronautics and Space Administration-Task Load Index (NASA-TLX) is no exception and a computer-based version has been made available.

NASA-TLX (Hart and Staveland 1988) is one of the most widely known tools for assessing subjective workload. Since its inception, it has been extensively tested and frequently used in human performance studies (Jorgensen *et al.* 1999); further, it is considered to be a robust measure of subjective workload (Moroney *et al.* 1995).

Comparing tasks carried out on computers and paper has been well researched (see Mead and Drasgow 1993 for a meta-analysis) and there is some evidence to suggest that greater processing is required when information is presented on a computer screen. For example, Wästlund *et al.* (2005) in a comparison of computers and paper found that

*Corresponding author. Email: j.noyes@bristol.ac.uk

Ergonomics
ISSN 0014-0139 print/ISSN 1366-5847 online © 2007 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00140130701235232

consumption of information (measured in terms of reading comprehension) and production of information were both superior in the paper-based condition. They attributed these effects to an increase in cognitive demands, i.e. the computer task was more tiring and more stressful than the paper-based task and this led to 'a greater mobilization of both perceptual and executive cognitive resources'. Although cognitive workload as such was not measured, Wästlund *et al.* concluded that the computer-based tasks were resulting in a higher cognitive workload.

Hence, it was hypothesized that in a self-analysis of the NASA-TLX measure, more workload would be reported for completion of the computer- than the paper-based version.

This study has important implications for the use of computer-based measures in general and, more specifically, for the practical use of the NASA-TLX. If there is a difference between the workload incurred during completion of the two forms of the NASA-TLX, this is likely to affect how participants rate the subjective workload associated with a particular task.

2. Method

2.1. *Experiment 1*

2.1.1. Design and participants. A within-participants design was employed. The comprehension task and the NASA-TLX (either presented via the computer or in paper form) were counterbalanced. Half the participants completed the first comprehension task on the computer followed by two computer versions of the NASA-TLX; the first administration of the NASA-TLX assessed the workload of the comprehension task and the second assessed the workload associated with filling in the NASA-TLX measure. For this group of participants, the second comprehension task was presented on paper followed by two paper versions of the NASA-TLX. The other half of the participants completed the experiment in reverse order, i.e. a paper-based comprehension task and two paper versions of the NASA-TLX, followed by a computer-based comprehension task and two computer versions of the NASA-TLX. Allocation of equal numbers of participants to each of the two conditions was randomized.

A total of 30 undergraduates from the University of Bristol carried out the experiment. All were volunteers who participated as part of a course requirement. The 12 males and 18 females had an age range of 18.12 to 21.10 years (mean 19.50, SD 0.89). None was known or observed to have any visual or other impairment that may have affected their ability to complete the tasks. All participants had similar levels of academic ability as indicated by comparable course entry requirements.

2.1.2. Materials. The comprehension task comprised two texts taken directly from Robert Louis Stevenson's *Essays of Travel*. They were chosen on the basis that their contents were not emotionally charged. Both texts were in the English language and were matched for a word frequency of 1600 words.

The NASA-TLX, both computer and paper versions, were original and as supplied by the Human Systems Information Analysis Center, Wright-Patterson AFB, OH (Version 1.0; NASA AMES Research Center, Moffett Field, CA, USA). No attempt was made to modify these measures; this was deliberate.

A post-experiment questionnaire was administered. This included questions on task difficulty, positive and negative aspects of the experiment, comparison of the two modes

of presentation of the NASA-TLX and reasons for preferring either the computer or paper version.

2.1.3. Procedure. The comprehension task required participants to read through the chosen texts at their own pace. They were then asked ten questions about the content of the material they had just read. The text was then removed from sight and participants were asked to complete either a computer version or a paper version of the NASA-TLX in order to rate the workload of the comprehension task. They were then asked to rate the workload associated with filling in the NASA-TLX.

The administration of the NASA-TLX is a two-part procedure. First, the rater evaluates the contribution of each factor (its weight) to the workload of the task. There are 15 possible pair-wise comparisons of the six scales.

In the computer version of the NASA-TLX, participants view a window in which they use the mouse to slide along each of the six workload scales. A mouse click marks the desired point on each scale. A second window is opened and participants give weightings to the scales by selecting which member of each pair contributed more to the workload. This is done by the participant clicking twice on the right mouse button to select the first factor of the pair or by clicking twice on the left mouse button to select the second factor of the pair.

In the paper version of the NASA-TLX, participants circle the member of each pair that contributed more to the workload of that task. At the end of the session, the number of times that each factor is selected is tallied, which can range from 0 (not relevant) to 5 (more important than any other factor).

The second part of the workload assessment involves the rater providing numerical ratings for each of the six scales, which reflect the magnitude of that factor in the task. Each scale is presented as a 12 cm line divided into 20 equal intervals between bipolar descriptors. These are 'low' and 'high' in all cases except the scale 'performance', which extends from 'good' to 'poor' (and is reverse scored).

The overall workload score is calculated by multiplying each raw rating by the weight given to that factor by the participant. The sum of the weighted ratings is then divided by 15 (the sum of the weights) to give an absolute workload score, which will lie between 0 and 100.

At the end of the experiment, participants completed the post-experiment questionnaire. They were then thanked and debriefed.

2.2. *Experiment 2*

Experiment 2 differed from experiment 1 in terms of design (assessment of the NASA-TLX always took place using the other form, and at the end of the experiment participants were required to generate absolute and relative workload scores (see Leggatt and Noyes 1996), materials (shorter and simpler task) and procedure (the introduction of 'focusing' questions).

2.2.1. Design and participants. A within-participants design was employed with similar counterbalancing of participants and conditions, but the design was modified to include the collection of absolute workload scores (as in experiment 1) and relative workload judgement scores.

A total of 30 undergraduates from the University of Bristol carried out the experiment. None had taken part in experiment 1. All were volunteers; they participated as part of a

course requirement and reported no visual or other impairment that may have affected their ability to do the experiment. In contrast to experiment 1, all participants were female. Analysis of the data from experiment 1 had indicated no sex differences for either computer ($t(28) = 0.60$, $p = 0.56$) or paper ($t(28) = 0.70$, $p = 0.49$). The 30 participants had an age range of 18 to 40 years (mean 22.43, SD 6.10).

2.2.2. Materials. The comprehension task in experiment 1 was replaced by a word search task. This comprised two simple word searches, which took about 5 min to complete. Given their simplicity, it was assumed the word searches were equivalent in terms of difficulty. This task replaced the comprehension task because it was easier and faster for participants to carry out. Unlike experiment 1, the word search task was only presented on paper. The NASA-TLX computer and paper presentations remained the same, although participants completed an 'extra' paper questionnaire at the end of the experiment, where they provided relative judgements of the workload of the two conditions.

2.2.3. Procedure. As well as the relative workload assessment questionnaire, a further modification was made to the experiment 1 procedure with the addition of three questions pertaining to participants' impressions of the NASA TLX. The questions were: What are your initial thoughts of the questionnaire that you have just filled in? Was it: (1) informative; (2) confusing; (3) neither confusing nor informative? What was your impression of the questionnaire? Was it: (1) easy; (2) moderately difficult; (3) difficult; (4) very difficult? If given the opportunity, would you use the questionnaire on a regular basis for personal use or experimental research? – the answer to this question was either YES or NO. If No was answered, participants had to explain briefly the reasons why they had chosen the NO option. Participants completed these questions on paper.

3. Results

It is evident from table 1 that no significant differences were found between the computer and paper conditions in experiment 1. However, a significant difference was found in experiment 2. This related to both the absolute measures (i.e. the workload measures per se) and the relative measures (i.e. workload ratings of the computer- and paper-based formats on the same scale). No significant differences were found between the weightings of the dimensions themselves for the two conditions.

Table 1. Measures of workload (SD) for the computer- and paper-based forms of the NASA-TLX.

	Computer	Paper	t	p*
Absolute measures of workload				
Experiment 1	42.00 (19.24)	39.07 (18.16)	1.14	NS
Experiment 2	52.53 (13.11)	44.27 (19.84)	1.85	<0.05
Relative measures of workload				
Experiment 2	46.73 (21.14)	36.20 (16.20)	3.31	<0.005

*One-tailed p values are given since the direction of the difference was predicted.
NASA-TLX = National Aeronautics and Space Administration-Task Load Index.

Note: Correlations between the raw (unweighted) and the weighted ratings for the relative scores were 0.97 (for the computer version) and 0.91 (for the paper version), $p < 0.01$. These are similar to previous correlations (Byers *et al.* 1989, Nygren 1991, Moroney *et al.* 1995).

4. Discussion

The primary aim of this study was to assess the workload costs associated with completing the NASA-TLX measure in its computer- and paper-based forms. It was hypothesized that the computer-based version would demonstrate a significantly higher workload. This was based on the evidence in the literature that there is a greater cost associated with processing information on a computer than on paper, and in experiment 2 this was found to be the case.

Building on the work of Leggatt and Noyes (1996), a difference was found between the absolute and relative ratings of workload across the two media. The concept of workload is still the topic of much debate in the literature and participants likewise will have their own thoughts about what comprises workload. Accommodating this subjective element is important both in experimental and operational settings. Gopher and Donchin (1986) referred to humans as not being 'workload meters' who could be wired up to provide workload readings. Consequently, they need a context in which to make their judgements. This may help to explain the difference between the absolute and relative workload scores. The aim of this study was to compare the workload associated with the two presentation media. Participants could not do this until they had experienced both conditions. At the end of experiment 2, asking them to compare the computer and paper-based versions using the same measure provided them with a stronger context in which to make their judgements. Hence, it is suggested that workload measures in a comparative situation are more meaningful when considered as a relative concept. Leggatt and Noyes provided some support for this when they concluded that the main benefit of relative workload judgements emanated from the experience of using the different interface designs and not the relative judgements on the scale itself.

In methodological terms, this self-assessment of the NASA-TLX was not straight forward and there are a number of considerations. The first concerns the repetition of questionnaire completion and the extent to which the monotony of having to complete the workload measure on a number of successive occasions affects the results. Every attempt was made to counteract this, but the very nature of the study makes repetition unavoidable. One solution might be to introduce another workload measure in order to inject some variety, but in a repeated measures design, in order to allow relative assessments, a minimum of four questionnaire completions will still be necessary.

Second, in experiment 1, it was observed that there was a tendency for participants after one or two completions of the NASA-TLX to mark the rating scales in a fairly automatic way with apparently little thought. (Relative workload measures might help combat this as respondents need to give more thought to making the comparison.) In order to try and reduce this, experiment 2 was modified so that the analysis of the measure was carried out using a version that was new (initially anyway) to the participants. It was hoped that this unfamiliarity might result in participants giving more thought to their workload ratings. However, the instructions that accompany the NASA-TLX suggest participants need to have a few attempts initially at completing the measure in order to familiarize themselves with the scales. Further changes introduced in experiment 2 to help focus participants on the workload assessment

included asking them questions about the NASA-TLX and replacing the comprehension task with a shorter and easier, word search task. Subsequent analysis of the comprehension task indicated it had a greater workload associated with it (although this was not known at the time the experiments were devised).

In conclusion, the computer-based version of the NASA-TLX was found to incur more cognitive workload and this was especially the case when relative as opposed to absolute judgements were made. This has implications for the administration of self-report workload questionnaires in terms of whether they are computer- or paper-based and the collection of absolute vs. relative workload assessments.

References

Byers, J.C., Bittner, A.C. and Hill, S.G., 1989, Traditional and raw load index (TLX) correlations: are paired comparisons necessary? In *Advances in Industrial Ergonomics and Safety*, A. Mital (Ed.), volume I, pp. 481–485 (London: Taylor & Francis).

Gopher, D. and Donchin, E., 1986, Workload – An examination of the concept. In *Handbook of Perception and Human Performance*, K. Boff, L. Kaufman and J. Thomas (Eds.) (New York: Wiley).

Hart, S.G. and Staveland, L.E., 1988, Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In *Human Mental Workload*, P.A. Hancock and N. Meshkati (Eds.), pp. 5–39 (New York: Elsevier).

Jorgensen, A.H., Garde, A.H., Laursen, B. and Jensen, B.R., 1999, Applying the concept of workload to IT work. In *Virtual Proceedings of CybErg 1999: The Second International Virtual Conference on Ergonomics*, L. Straker and C. Pollock (Eds.) (Perth, Australia: International Ergonomics Association/Curtin University of Technology).

Leggatt, A.P. and Noyes, J.M., 1996, Workload measures: Absolute versus relative judgements. In *Virtual Proceedings of CybErg 1996: The First International Virtual Conference on Ergonomics*, L. Straker and C. Pollock (Eds.) (Perth, Australia: International Ergonomics Association/Curtin University of Technology).

Mead, A.D. and Drasgow, F., 1993, Equivalence of computerized and paper and pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, **114,** 449–458.

Moroney, W.F., Biers, D.W. and Eggemeier, F.T., 1995, Some measurement and methodological considerations in the application of subjective workload and measurement techniques. *International Journal of Aviation Psychology*, **5,** 87–106.

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.

Wästlund, E., Reinikka, H., Norlander, T. and Archer, T., 2005, Effects of VDT and paper presentation on consumption and production of information: Psychological and physiological factors. *Computers in Human Behavior*, **21,** 377–394.