Behavior Research Methods
2009, 41 (1), 113-117
doi:10.3758/BRM.41.1.113

NASA TLX: Software for assessing subjective mental workload

Alex Cao, Keshav K. Chintamani, Abhilash K. Pandya, and R. Darin Ellis
Wayne State University, Detroit, Michigan

The NASA Task Load Index (TLX) is a popular technique for measuring subjective mental workload. It relies on a multidimensional construct to derive an overall workload score based on a weighted average of ratings on six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration level. A program for implementing a computerized version of the NASA TLX is described. The software version assists in simplifying collection, postprocessing, and storage of raw data. The program collects raw data from the subject and calculates the weighted (or unweighted) workload score, which is output to a text file. The program can also be tailored to a specific experiment using a simple input text file, if desired. The program was designed in Visual Studio 2005 and is capable of running on a Pocket PC with Windows CE or on a PC with Windows 2000 or higher. The NASA TLX program is available for free download.

Obtaining mental workload levels during task performance is a difficult procedure. The workload level experienced by an operator can affect task performance. This effect can be caused by either excessive or reduced mental workload. Thus, estimating workload levels can help isolate sources that affect performance. Several techniques can be used to measure operator workload. Psychophysiological measurements using electroencephalograms (EEGs) have been used with much success. However, EEG machines are not the most portable, practical, or ubiquitous devices, which constrains their usefulness in experiments. Obtaining physiological mental workload data is a nontrivial task, and there has been a large amount of work on designing lightweight and portable workload measuring devices (Kramer, 1991).

Workload assessment techniques should possess the following properties: sensitivity, diagnostic capabilities, selectivity, low intrusiveness, reliability, and ease of implementation (Eggemeier, 1988). The NASA Task Load Index (TLX) has been shown to meet these criteria (Rubio, Díaz, Martín, & Puente, 2004). It is a subjective workload assessment technique that relies on a multidimensional construct to derive an overall workload score based on a weighted average of ratings on six subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration level. The use of these six subscales to compute an overall workload score has been found to reduce variability among subjects, relative to a unidimensional workload rating, while providing diagnostic information about workload sources (Hart & Staveland, 1988). Specific sources of mental workload imposed by different tasks are an important determinant of workload experiences. Three of the subscales relate to the demands imposed on the subject (mental, physical, and temporal demand), whereas the other subscales focus on the interaction of the subject with the task (performance, effort, and frustration level). Descriptions of the six subscales are shown in the Appendix. Thus, the NASA TLX combines subscale ratings that are weighted according to their subjective importance to subjects for a specific task.

The NASA TLX has been tested in experimental tasks (e.g., the Fittsberg task, choice reaction time, *same/different* judgments, mental arithmetic, mental rotation, etc.), task supervisory control simulations, and flight simulators. The results of the first validation studies are summarized in Hart and Staveland (1988).

The NASA TLX has been used in a variety of fields (Hart, 2006). It has been used in studies involving the evaluation of visual and/or auditory displays, vocal and/or manual input devices, and virtual/augmented vision. Studies have explored the relationship between NASA TLX ratings and other performance factors, such as fatigue (Baulk et al., 2007), stress (Reilly, Grasha, Matthews, & Schafer, 2003), trust (Turner, Safar, & Ramaswamy, 2006), experience (Greenwood-Ericksen, Oron-Gilad, Szalma, Stafford, & Hancock, 2006), and situational awareness (Kaber, Onal, & Endsley, 2000). Other NASA TLX studies have involved measures of physiological (e.g., cardiovascular, muscular, and skin- or brain-related) function thought to index different aspects of workload (Miyake, 2001; Ryu & Myung, 2005).

Other studies have used modified versions of the original NASA TLX. The use of an unweighted or raw TLX (RTLX) is the most common, because high correlations have been shown between the weighted and unweighted scores (Byers, Bittner, & Hill, 1989; Moroney, Biers,

A. Cao, av9852@wayne.edu

Eggemeier, & Mitchell, 1992). Park and Cha (1998) modified the NASA TLX subscales in their study of vehicle navigation systems. They replaced four of the original subscales and tailored the descriptions to the specific task of driving while using the navigation system. Another modification in the procedure is the delayed reporting of workload ratings. Although workload ratings are usually recorded immediately after the task, Moroney et al. (1992) showed that a 15-min delay does not significantly interfere with the recall of ratings.

Studies have shown the NASA TLX to be favored most by subjects, when compared with other subjective workload assessment techniques (e.g., SWAT, the Cooper–Harper scale), and also to be highly correlated with other measures of workload (Battiste & Bortolussi, 1988; Hill et al., 1992). NASA TLX reliability for repeated measures has shown correlations of .77 (Battiste & Bortolussi, 1988).

A principal reason for the popularity of the NASA TLX among researchers is its ease of implementation. The multidimensionality of the NASA TLX allows for a more detailed analysis of the workload source relative to other techniques that are based primarily on rankings of mental workload sources. The NASA TLX is also very portable and can be used in operational experiments. In studies involving a large number of observations, using a paper version of the NASA TLX entails unnecessary postprocessing time spent manually transferring the data to a spreadsheet, where transcription errors can occur. With the ubiquitous nature of personal computers (PCs) and personal digital assistants (PDAs), a software version of the NASA TLX would greatly assist in simplifying the collection, postprocessing, and storage of raw data.

An extremely high correlation of .94 has been reported between the ratings obtained by using either a paper-and-pencil or computer method (NASA, 1986) of administering the NASA TLX in a study involving three tasks (target acquisition, grammatical reasoning, and unstable tracking) at two levels of difficulty (easy and hard). Ratings obtained by using the computer version were found to be significantly higher than those obtained by using the paper-and-pencil version. However, the absolute difference in the ratings was not as important as the fact that the patterns of the magnitudes of the ratings were extremely consistent for all tasks (Hart & Staveland, 1988). Other studies have confirmed this medium effect with the NASA TLX (Noyes & Bruneau, 2007) and with other tests in general (Mead & Drasgow, 1993). In this article, we describe an implementation of a software version of the NASA TLX.

NASA TLX PROCEDURE

The NASA TLX consists of two parts: ratings and weights. Ratings for each of the six subscales are obtained from the subjects following the completion of a task. A numerical rating ranging from 0 to 100 (*least* to *most taxing*) is assigned to each scale. Weights are determined by the subjects' choices of the subscale most relevant to workload for them from a pair of choices. The weights are calculated from the tally of these choices from 15 combinatorial pairs created from the six subscales. The weights range from 0 to 5 (*least* to *most relevant*). The ratings and weights are then combined to calculate a weighted average for an overall workload score.

NASA TLX SOFTWARE

Installation Instructions

The program is available for download at ece.eng.wayne.edu/~apandya/Software/NASA_TLX. The Web page contains the executable file; an example input file for modifying the text; and the source code, including the Visual Studio 2005 project files. Both PC and Pocket PC (Windows CE) versions are available. Any problems with the Web page or software can be reported by contacting the authors. Known problems with using the program and their solutions will be posted on the Web page.

Program Operation

The NASA TLX program is a dialog-based Microsoft Foundation Class application. Upon execution, the program brings up the INITIAL dialog box, asking the experimenter for the subject number and the name of the output file (which will contain the raw data) to be created (Figure 1). The user can choose to create a new output file (default option) or to append the data to an existing file. Appending data to an existing file would be used in scenarios in which subjects participate in multiple data collection runs over a long period of time (e.g., several days) without the need to have the program running the entire time. The user can also choose whether to use the weighted (default option) or unweighted method for calculating the workload score. The program does not allow the experimenter to enter any letters in the subject box, preventing any direct identification of the subject. It is left up to the experimenter to document offline the link between subjects and their data.

Hitting "ok" brings the user to the MENU dialog box (Figure 2), where the experimenter specifies the condition and trial number for the set of workload parameters that will be collected next. The experimenter can then select whether to collect ratings or weights from the subject. If the unweighted method is selected, the "Weights" button will be disabled, since it is not needed.

The RATINGS dialog box displays the subscales, along with the corresponding description that the subject will need to provide a rating for (Figure 3). Each subscale contains bipolar descriptors of the two endpoints. The subject is required to make a mark by clicking the left button and dragging across the horizontal line to indicate the appropriate response. A circle is drawn at the recorded value once a mark has been made. If the subject attempts to draw more than one mark, only the last mark is shown and recorded. If the mark crosses the horizontal line at more than one location, only the first crossing will be recorded. If the mark does not cross the horizontal line, no measurement is taken. If the mark falls outside the ends of the line, the nearest endpoint is recorded. The program will not move on to the next subscale until a proper response is recorded (i.e., it eliminates the possibility of a nonscore). The order of presentation of the six subscales is random-

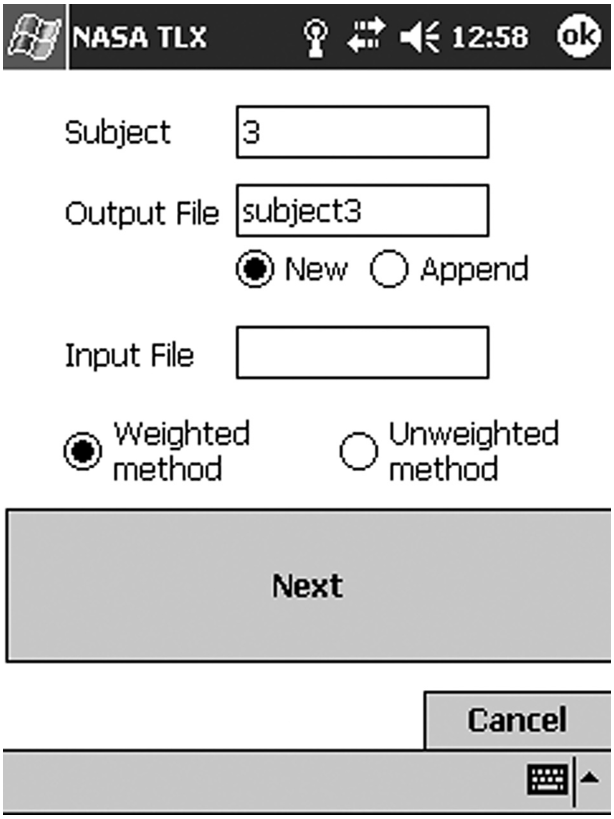

Figure 1. Initial dialog box, asking for subject ID, output filename, and whether to create a new file or append to an existing file. The experimenter can also enter an input file in order to tailor the subscales for a specific task. The experimenter also chooses whether to use the weighted or unweighted method for calculating workload scores.

ized. Once a valid answer is recorded, the mark along the horizontal scale is normalized to a number between 0 and 100, rounded to the nearest integer of 5, as recommended by the NASA TLX.

The WEIGHTS dialog box displays the combinatorial pair of subscales that the subject will choose from (Figure 4). Subjects are required to choose one of the two subscales via a radio button. The program will not allow the subject to move on to the next pairing until one of the two subscales is picked. The order of presentation of the 15 combinatorial pairs of subscales is presented in random order. In addition, the order within the pairing is also randomized.

At any time, the user can exit the RATINGS or WEIGHTS dialog box. A message box will ask for confirmation to exit or to continue. Exiting the dialog box will negate all ratings or weights recorded for the current condition or trial and bring the user back to the MENU dialog box. Thus, the program will only write to the output file once all six ratings or 15 comparisons have been completed.

The experimenter can choose to collect ratings and/or weights indefinitely. Hitting "Exit Program and Calculate Scores" in the MENU dialog box exits the program and calculates the weighted workload score for each condition or trial combination.

The program was written and compiled in Visual Studio 2005. The executable file can run on a PC (with Windows 2000 or later) without the need for the user to install any software. We have also ported the program to be compatible with smart devices, such as a PDA running Windows CE. The main difference between the programs is that a stylus is used in the Windows CE version to mark the ratings instead of a left buttonclick and drag.

Output Files

Two output files are created when the program is exited. The first file contains the raw data, with each completed set of ratings or weights in the order in which they were done. The file contains the following column headers: *Type*, *Subject*, *Condition*, *Trial*, *MD*, *PD*, *TD*, *PF*, *EF*, and *FR*. *Type* refers to whether the line is a set of ratings or weights. It is also relatively easy to differentiate between the two, since the subscale columns (i.e., the last six) will range from 0 to 100 for ratings and from 0 to 5 for weights.

The second output file (with the extension "_scores") contains the computed weighted workload score, along with the corresponding ratings and weights. The file also lists any collected ratings and weights that were not used in calculations because there were no corresponding weights or ratings. If the unweighted method is chosen, the second output

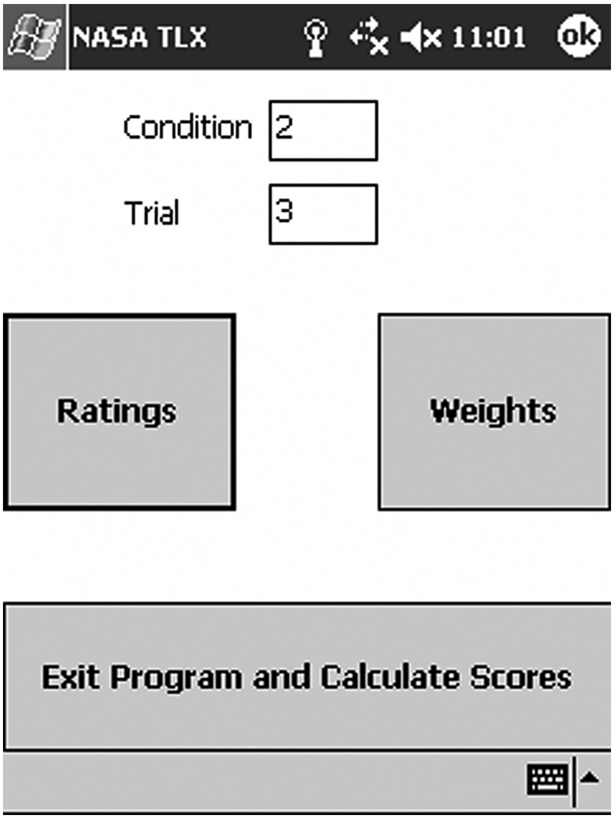

Figure 2. Menu dialog box, in which the condition and trial number are entered. The experimenter can choose to collect ratings or weights or to exit the program. The user will be returned to this dialog box after finishing one set of workload ratings or weights.

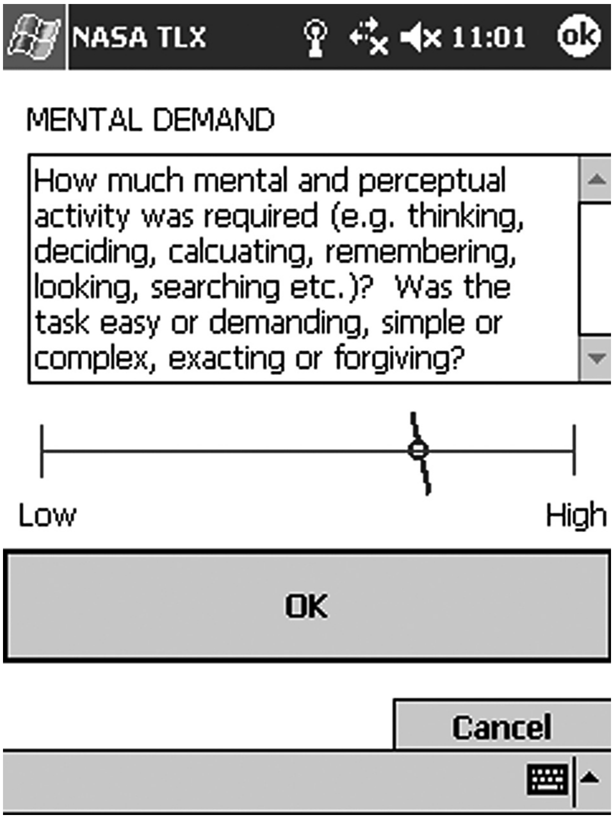

Figure 3. RATINGS dialog box, in which ratings for the six subscales are entered.

file is not created. Instead, the unweighted workload score is appended to the end of each line in the raw data file.

File columns are separated by a single space for easy import into other programs for analysis.

Modifying the Program for General Purpose Use

The descriptive text within the program can be modified for a specific experiment, if desired. This allows the experimenter to tailor the questions to a specific task, which can be accomplished by specifying the input text file containing the modified text in the INITIAL dialog box beside the "Input File" field. The user can choose to change the name of the subscales, the subscale descriptions, the subscale bipolar descriptors, and/or the number of subscales. Any text editor (e.g., WordPad) can be used to create or modify the text file. An example input file is included on the Web site.

One caveat related to changing the NASA TLX subscales and descriptions for a specific task is the need to establish the validity, sensitivity, and reliability of the new instrument before use (Cha & Park, 1997; Park & Cha, 1998). Another caveat is that the modified instrument can no longer be referred to as the NASA TLX (since that refers to the original set of subscales and descriptions, which no longer apply).

Other NASA TLX Programs

There are two other NASA TLX programs currently available for download. One is from the U.S. Navy Research Laboratory Web site: www.nrl.navy.mil/aic/ide/NASATLX.php. A drawback of this software is that only one set of ratings and weights is allowed to be entered before the program automatically exits. Also, the software only allows the subject ID to be entered and has no inputs for condition or trial number. There are no subscale descriptions in the ratings dialog box for the subject, which would be useful in case the subject needs to refer back to what the subscale is asking.

The second program is from the NASA Ames Research Center site: humansystems.arc.nasa.gov/groups/TLX/computer.php. The page requires that the user register before downloading the installation files. The drawbacks with this software include having to exit the program in order to switch between ratings and weights. In addition, the ratings and weights are not saved to the same file but, rather, to two separate files. Also, all the data are saved in a single column (as opposed to multiple columns) with no headers, making the data harder to use in a spreadsheet program without rearranging them. The program also does not calculate the overall workload score.

In addition, neither of these programs can be modified by the experimenter to change the subscales for a specific task/experiment. Thus, we feel that our program is better in terms of usability and flexibility, because it addresses each and every one of these drawbacks. This program is an

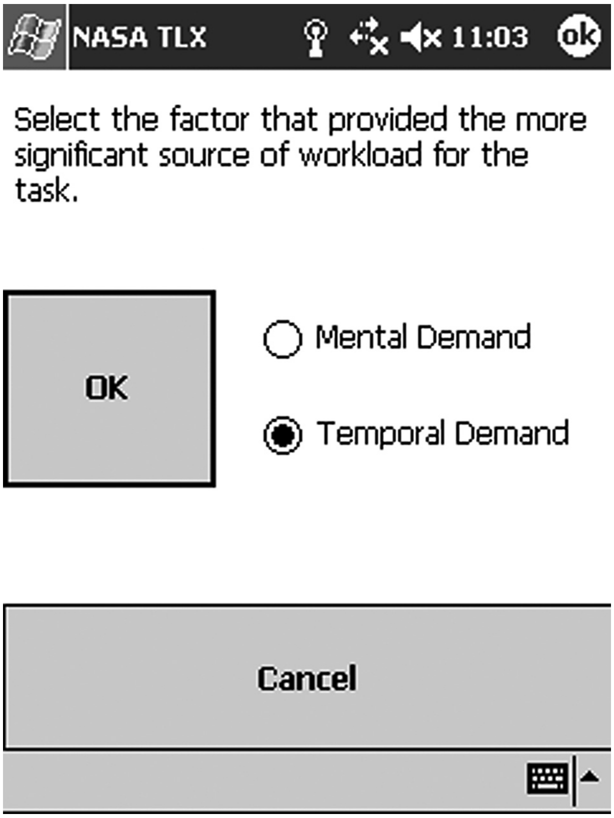

Figure 4. WEIGHTS dialog box, in which weights are determined from the tally of subscales from 15 combinatorial pairings.

easy-to-use tool for researchers who want to study subjective mental workload using the NASA TLX.

AUTHOR NOTE

Correspondence concerning this article should be sent to A. Cao, Wayne State University, Department of Electrical and Computer Engineering, Rm. 3160, 5050 Anthony Wayne Dr., Detroit, MI 48202 (e-mail: av9852@wayne.edu).

REFERENCES

Battiste, V., & Bortolussi, M. (1988). Transport pilot workload—A comparison between two subjective techniques. In *Proceedings of the Human Factors and Ergonomics Society 32nd Annual Meeting* (pp. 150-154). Santa Monica, CA: Human Factors & Ergonomics Society.

Baulk, S. D., Kandelaars, K. J., Lamond, N., Roach, G. D., Dawson, D., & Fletcher, A. (2007). Does variation in workload affect fatigue in a regular 12-hour shift system? *Sleep & Biological Rhythms*, **5**, 74-77.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics & safety* (Vol. 1, pp. 481-485). London: Taylor & Francis.

Cha, D., & Park, P. (1997). User required information modality and structure of in-vehicle navigation system focused on the urban commuter. *Computers & Industrial Engineering*, **33**, 517-520.

Eggemeier, F. T. (1988). Properties of workload assessment techniques. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 41-62). Amsterdam: Elsevier.

Greenwood-Ericksen, A., Oron-Gilad, T., Szalma, J. L., Stafford, S., & Hancock, P. A. (2006). Workload and performance: A field evaluation in a police shooting range. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 1953-1957). Santa Monica, CA: Human Factors & Ergonomics Society.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 904-908). Santa Monica, CA: Human Factors & Ergonomics Society.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: Elsevier.

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

Kaber, D. B., Onal, E., & Endsley, M. R. (2000). Design of automation for telerobots and the effect on performance, operator situation awareness, and subjective workload. *Human Factors & Ergonomics in Manufacturing*, **10**, 409-430.

Kramer, A. F. (1991). Physiological metrics of mental workload: A review of recent progress. In D. L. Damos (Ed.), *Multiple-task performance* (pp. 329-360). London: Taylor & Francis.

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

Miyake, S. (2001). Multivariate workload evaluation combining physiological and subjective measures. *International Journal of Psychophysiology*, **40**, 233-238.

Moroney, W. F., Biers, D. W., Eggemeier, F. T., & Mitchell, J. A. (1992). A comparison of two scoring procedures with the NASA Task Load Index in a simulated flight task. *Aerospace and Electronics Conference, 1992. NAECON 1992, Proceedings of the IEEE 1992 National* (Vol. 2, pp. 734-740). Dayton, OH: NAECON.

NASA (1986). *Task Load Index (TLX): Computerized version* (Version 1.0). Moffett Field, CA: Human Research Performance Group, NASA Ames Research Center.

Noyes, J. M., & Bruneau, D. P. J. (2007). A self-analysis of the NASA-TLX workload measure. *Ergonomics*, **50**, 514-519.

Park, P., & Cha, D. (1998, October). *Comparison of subjective mental workload assessment techniques for the evaluation of in-vehicle navigation system usability.* Paper presented at Session T56 of the 5th World Congress on Intelligent Transport Systems 1998, Seoul.

Reilly, S., Grasha, A. F., Matthews, G., & Schafer, J. (2003). Automatic-controlled information processing and error detection in a simulated pharmacy-verification task. *Perceptual & Motor Skills*, **97**, 151-174.

Rubio, S., Díaz, E., Martín, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology: An International Review*, **53**, 61-86.

Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, **35**, 991-1009.

Turner, C. W., Safar, J. A., & Ramaswamy, K. (2006). The effects of use on acceptance and trust in voice authentication technology. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 718-722). Santa Monica, CA: Human Factors & Ergonomics Society.

APPENDIX
NASA-TLX Rating Scale Descriptions

Title	Endpoints	Description
Mental Demand	Low/High	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed, and annoyed or secure, gratified, content, relaxed, and complacent did you feel during the task?

(Manuscript received March 3, 2008;
revision accepted for publication July 14, 2008.)