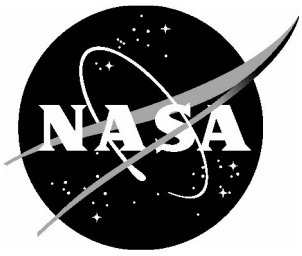
NASA/TM-2011-217166



Evaluation of Electronic Formats of the NASA Task Load Index

Anna C. Trujillo
Langley Research Center, Hampton, Virginia

August 2011

NASA STI Program . . . in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers, but having less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at *http://www.sti.nasa.gov*

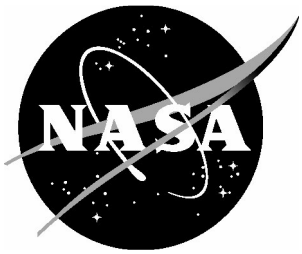
- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at 443-757-5803

- Phone the NASA STI Help Desk at 443-757-5802

- Write to:
  NASA STI Help Desk
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076-1320

NASA/TM-2011-217166



Evaluation of Electronic Formats of the NASA Task Load Index

Anna C. Trujillo
Langley Research Center, Hampton, Virginia

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23681-2199

August 2011

Acknowledgments

The authors gratefully acknowledge the significant contributions of the many members of the research team. Deserving of credit are Alan Pope who helped start this line of research; and Hayes Press of Lockheed Martin and Daniel Bruneau of AT&T who helped refine the display formats investigated. Also deserving of credit is Lucas Hempley of Lockheed Martin for his programming of the experiment.

Available from:

NASA Center for AeroSpace Information
7115 Standard Drive
Hanover, MD 21076-1320
443-757-5802

Table of Contents

List of Figures

List of Tables

List of Equations

Abstract

Paper questionnaires are being replaced by electronic questionnaires. The primary objective of this research was to determine whether electronic formats of paper questionnaires change subjects' ratings and, if so, how the ratings changed. Results indicated that there were no satistically significant differences in self-assessment of workload when using the electronic replica or the paper format of the NASA-TLX scale. Variations of the electronic formats were tested to enforce structure to the TLX scale. Respondents had more consistent ratings with these alternative formats of the NASA-TLX. Non-pilots, in general, had lower workload ratings than pilots. The time to input the rating was the fastest for the electronic facsimile and random title formats. Also subjects preferred the electronic formats and thought these formats were easier to use. Therefore, moving questionnaires from paper to electronic media could change respondents' answers.

Introduction

With the use of computers, paper questionnaires are being replaced by electronic questionnaires. A respondent's ratings, though, can subtly change when using an electronic format – computer tests scores may be lower for lower-performing individuals than for higher-performing individuals (Noyes, Garland, & Robbins, 2004); computer-based presentation impaired understanding and production of information (Wästlund, Reinikka, Norlander, & Archer, 2005); method of cognitive processing may change with information presented on a computer (Noyes & Garland, 2003), which may change a respondent's answer; and a higher workload associated with completing an electronic form may affect those ratings (Noyes & Bruneau, 2007). This is not surprising when previous research has found that even with the traditional paper formats of questionnaires, the format may affect a subject's ratings (Riley & Wilson, 1990; Wilson & Riley, 1989). On the other hand, some research has indicated that reading times between paper and electronic formats was the same as long as the electronic version matches the paper version's typeface, font size, etc. (Noyes & Garland, 2003). This research was conducted to determine if electronic questionnaire formats affect responses and, if so, how electronic questionnaire formats change subjective ratings from the traditional paper format.

Objective

The objective was to determine whether electronic formats of paper questionnaires change subjects' ratings and, if so, how the formats change respondents' ratings. The questionnaire investigated was the NASA Task Load Index (NASA-TLX) which uses six continuous scales to arrive at a rating of workload (Byers, Bittner, & Hill, 1989; Hart & Staveland, 1988). In general, this research determined how electronic formats might affect responses to a scaled-response type of questionnaire. Another main objective was to determine the questionnaire format effects between pilots and non-pilots because the scaled-response type of questionnaire is employed widely outside the pilot community. A secondary objective was to determine the effects of the media and formats of the questionnaires on subject preferences.

Experimental Variables

Subjects

Twenty people participated as subjects. Ten were certificated pilots with at least a current Private Pilot license (Federal Aviation Administration, August 28, 2008). The rest of the subjects were non-pilots. The average age of the pilots was 48 years and the average age of the non-pilots was 40 years. The pilots averaged 22 years of piloting experience and they had an average of 7314 hrs of total piloting time. The slight difference in mean age – pilot population being older – was not considered to be a significant effect on usage or acceptability of electronic forms.

NASA-TLX Formats

Each subject saw five NASA-TLX formats – the standard paper format and four electronic formats, which were counterbalanced across the subject populations (*i.e.*, pilots and non-pilots). The electronic formats were: (1) electronic facsimile, (2) random, (3) random title, and (4) random description. The electronic facsimile provided a direct comparison of paper versus electronic media. The random formats were developed to evaluate potential advantages when using electronic formats. Specifically, the random formats forced respondents to process which workload measure they were responding to rather than just quickly filling in their ratings based more on pattern recognition than actually taking the time to read and comprehend each scale (Trujillo, 2008). The random title and random description formats were tested to determine if the descriptions were needed after the initial training.

Paper NASA-TLX

The paper NASA-TLX was the basic paper version of the NASA-TLX (Fig. 1). Subjects indicated their rating on each scale by placing a mark on the continuous scale.

Electronic Facsimile NASA-TLX

The electronic facsimile NASA-TLX was basically the paper version of the NASA-TLX translated to a computer screen (Fig. 2). When the questionnaire first appears, no ratings markers are present. In order to make a rating, subjects had to touch each scale at the location they wanted to mark. At that point, a vertical bar appeared at the subject's touch location on the scale. This rating method was also employed by all the electronic versions described below.

Random NASA-TLX

The random NASA-TLX showed each of the scales, with its associated title and description, in random order one at a time (Fig. 3). For example, the subject would rate his mental demand and, after completing that rating, the scale for temporal demand would take the previous scale's place.

Random Title NASA-TLX

The random title NASA-TLX again showed each of the scales in random order one at a time (Fig. 4). But, for this format, only the title of each scale was shown such as "Physical Demand." A subject could ask for the description of any title after he had completed all six subscales of the NASA-TLX and before the beginning of the next run but he was unable to change his ratings for the run just completed.

Random Description NASA-TLX

Lastly, the random description NASA-TLX showed each of the scales in random order one at a time (Fig. 5). For this format, though, only the description of each scale was shown. Similar to the random title NASA-TLX, a subject could ask for the title of any description after he had completed all six subscales of the NASA-TLX and before the beginning of the next run.

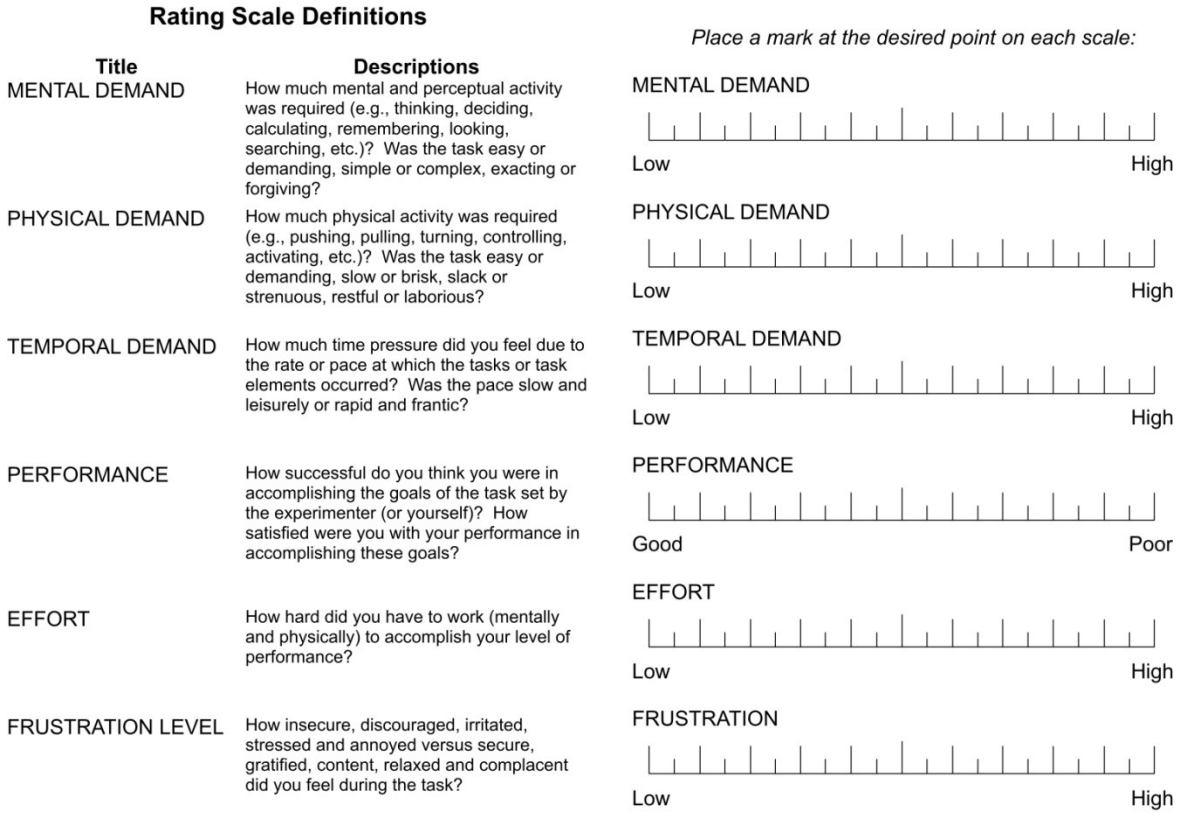


Figure 1. Paper Version of the NASA-TLX

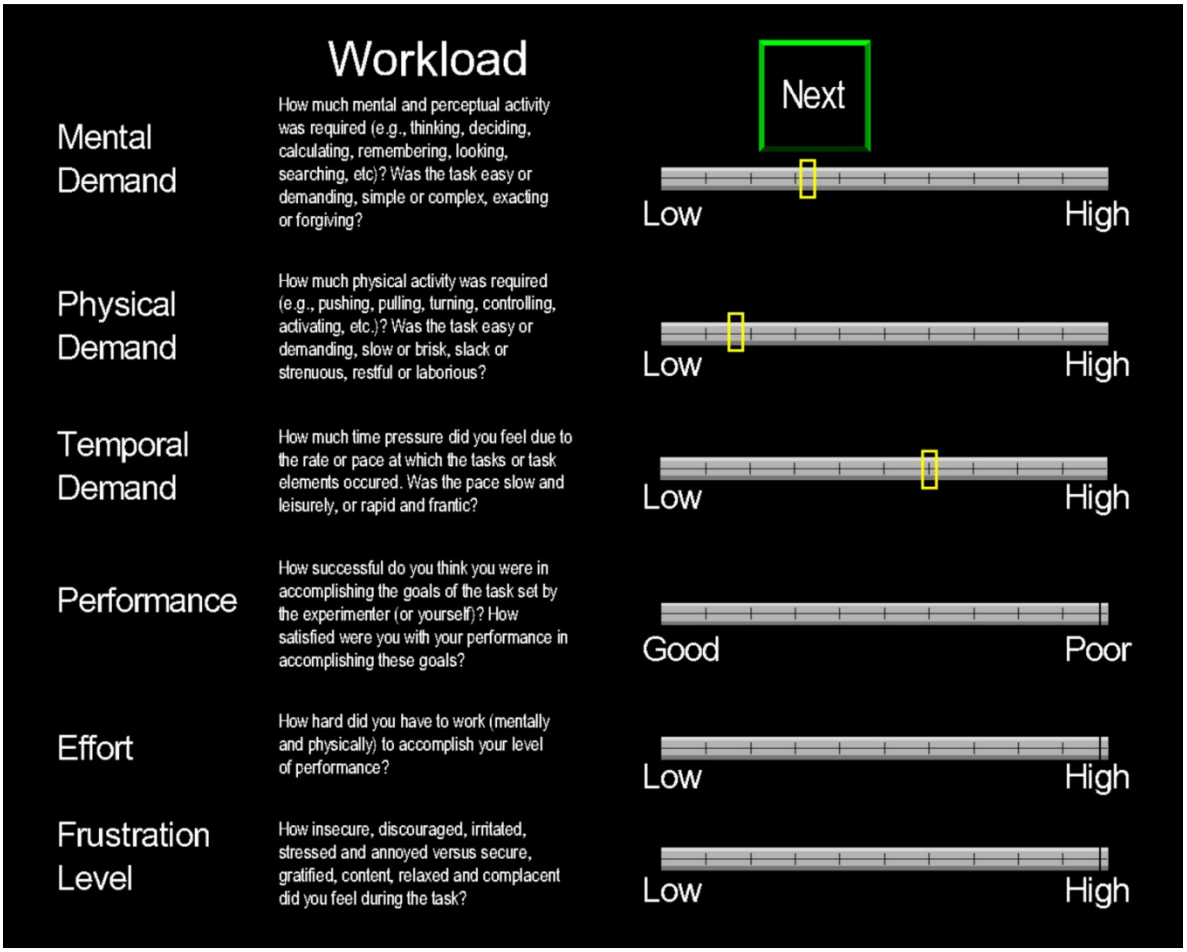
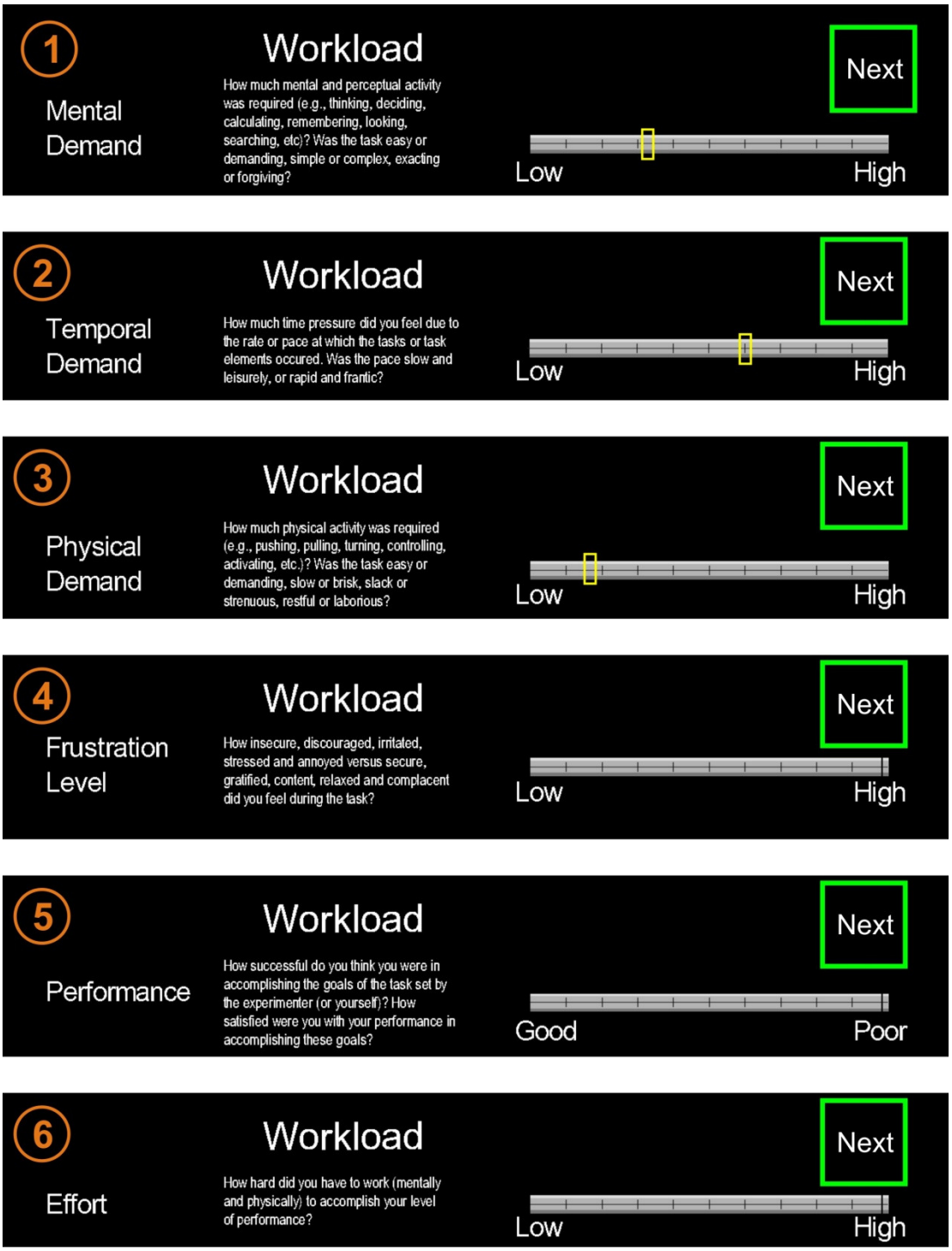


Figure 2. Electronic Facsimile Version of the NASA-TLX



NOTE: Each scale with its associated title and description shown one at a time in random order
Figure 3. Random Version of the NASA-TLX

NOTE: Each scale with its associated title shown one at a time in random order

Figure 4. Random Title Version of the NASA-TLX

NOTE: Each scale with its associated description shown one at a time in random order

Figure 5. Random Description Version of the NASA-TLX

Control Task Difficulty

This experiment had subjects perform a compensatory tracking task that required them to keep a pseudo-randomly moving target centered (Fig. 6). The subjects controlled the target with a right-handed sidestick.

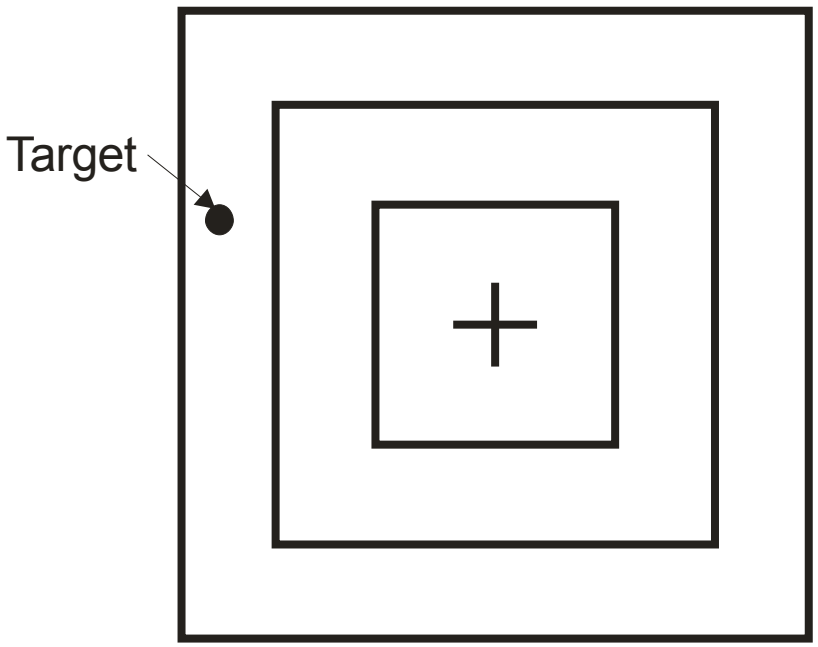


Figure 6. Tracking Task

During each data run, the difficulty level of keeping the target centered was constant. Ten control task difficulty levels were created and each subject saw each difficulty level in random order during each NASA-TLX format condition. The control task difficulty level was manipulated by changing the speed the target moved and the angle difference between the target's previous direction and its current direction. In order to allow for sufficient control authority, the joystick had stick shaping based on control task difficulty level (Eq. 1). The target angle changed every 1 sec randomly between ±45 degrees (Eq. 2). Target speed also changed 1 sec and was dependent on the control task difficulty level (Eq. 3). Therefore, the target distance from the center was dependent on the target speed (Eq. 4). A pretest verified that the control task difficulty levels would modulate workload and therefore tease out possible differences among the NASA-TLX formats tested.

$$stick\_shape = CTD * 2.77 + 2.23$$
$$\text{where } CTD \triangleq \text{Control Task Difficulty} = \{CTD \mid 1 \le CTD \le 10\} \tag{1}$$

$$target\_angle_{t+1} = target\_angle_t + angle_\Delta + \tan^{-1}\frac{y_{stick\_position}}{x_{stick\_position}}$$
$$\text{where} \quad angle_\Delta = \{angle_\Delta \mid -45 \le angle_\Delta \le 45\} \tag{2}$$
$$\left(x_{stick\_position} \vee y_{stick\_position}\right) = stick\_shape * stk\_pox_{x \vee y}$$
$$stk\_pos \triangleq \text{Stick Position} = \{stk\_pos \mid -1 \le stk\_pos \le 1\}$$

$$target\_speed = CTD * 2.17 - 1.17 \tag{3}$$

$$target\_distance_{t+1} = target\_distance_t + target\_speed + \sqrt{x^2_{stick\_position} + y^2_{stick\_position}} \tag{4}$$

Dependent Variables

The primary dependent variable was the subjects' NASA-TLX ratings. The time taken to complete the NASA-TLX ratings and the workload incurred to complete the NASA-TLX ratings were also analyzed.

Secondary dependent variables were associated with common piloting tasks – keeping a target centered, detecting whether a displayed number increased or decreased, and answering a question that required basic multiplication skills. A subject's accuracy and time to answer the questions were recorded and analyzed.

At the end of the experiment, subjects completed a final paper questionnaire (Appendix A). This final paper questionnaire asked subjects to rate on a continuous scale how easy the NASA-TLX formats were for rating the control task difficulty and the associated workload to complete the various NASA-TLX formats. The questionnaire also asked for subject preferences, and likes and dislikes by display type.

Hypotheses

This research tested several hypotheses encompassing effects of the NASA-TLX formats and the target tracking. Specifically for the NASA-TLX formats, it was hypothesized that subjects will more accurately reflect the workload of the tracking task with the random title format than with the formats with the descriptions in them (*i.e.*, paper, electronic facsimile, random, and random description formats). With the random formats, subjects will not be influenced by the other ratings on the questionnaire. Also, with just the titles available, subjects will fully read the available information (*i.e.*, the title) whereas the formats with the descriptions available, subject may not fully read all the available information. It was also hypothesized that subjects will complete the NASA-TLX the fastest with the paper and electronic facsimile formats and will complete the NASA-TLX the slowest with the random formats, especially for the random description format. This is because the paper and electronic facsimile formats require the least number of button pushes; whereas, the random formats require more button pushes. The time to complete the NASA-TLX with the random description format will be the longest because the order of the questions keep changing requiring subjects to read the long text description and this format also includes several button pushes. Lastly, regarding the NASA-TLX format, it was hypothesized the subjects will have an overall preference for the electronic facsimile format because all the information is on one screen and it requires the least amount of button pushes of the electronic formats.

For NASA-TLX ratings, it was hypothesized that the ratings will increase with the control task difficulty level. The difficulty of keeping the target centered as indicated by the CH rating at each difficulty level also will indicate an increase in workload for subjects. Also, pilots will have lower workload ratings than non-pilots because the tracking task is a familiar task to them.

Finally, the hypothesis regarding the subject's ability to keep the pseudo-randomly moving target centered was that the root mean square distance of the pseudo-randomly moving target from the center of the tracking task diagram would be smaller for pilots than for non-pilots. This is simply because pilots routinely perform this type of task for their job.

Procedure

When subjects first arrived, they signed a consent form before being given a detailed verbal briefing on the experiment tasks. Subjects then moved to the simulator where they completed two practice runs, which behaved exactly like the data runs, with the first NASA-TLX format. After the practice runs,

subjects completed 10 data runs where each run had a prescribed control task difficulty level. The order a subject saw a particular control task difficulty level was randomized for each format. During each run, subjects had to keep a moving target centered for 1 minute using a right-handed side stick. They also had to indicate whether a displayed number increased or decreased and answer a question that required basic multiplication skills. At the end of each run, subjects completed the NASA-TLX and the workload of determining their NASA-TLX rating. After the 10 data runs with the first NASA-TLX format, subjects completed at least one practice run with the next NASA-TLX format and then the 10 data runs with that NASA-TLX format. This was repeated until subjects had seen all five NASA-TLX formats. At the end of the simulation runs and questions, subjects completed the final paper questionnaire (Appendix A).

Apparatus

The simulations ran on two PCs running Windows™ XP Professional. These had a redraw refresh rate of 60Hz and a graphics update rate of 30Hz. The target tracking task was displayed on a 30-inch LCD screen centered in front of and slightly above the subject's eye level (Fig. 7). The display indicating the increasing or decreasing number and the information to answer the multiplication question was on a screen to the right of the subject. The questions were answered using a touch screen to the subject's left. The NASA-TLX questionnaire was also presented on this left screen at the end of the run. These two displays were 19-inch LCD screens with an Elo Touchsystems IntelliTouch overlay for touch-screen capability. The side stick used was a Saitek Cyborg evo joystick (Saitek Ltd., 2003). Subjects used their right hand to manipulate the side stick.

Data Analysis

Data were analyzed using SPSS® for Windows v16. The data were analyzed using a 3-way ANOVA with pilot status (pilot vs. non-pilot), NASA-TLX format, and control task difficulty as the independent variables.

Results

Learning or Fatigue Effects on Tracking Error

The subject run number did not significantly affect the root mean square distance of the pseudo-randomly moving target from the center of the target. Therefore, no perceptible learning effects or fatigue effects appeared to affect the ability of the subject to keep the target centered.

Common Piloting Tasks

The NASA-TLX format did not significantly affect the secondary dependent variables associated with target tracking, detecting a value change, and answering a multiplication problem. Neither the answer accuracy nor the time to answer the questions was dependent on the NASA-TLX format.

NASA-TLX Ratings

When comparing the NASA-TLX ratings, collapsed across all formats, pilot status was significant across all six component measures in the NASA-TLX and the combined rating of the NASA-TLX (Table 1). The combined rating was the average of the six independent NASA-TLX scores. For these analyses, the combined rating did not include component weightings from a paired comparison of the 6 independent NASA-TLX scores. This methodology increases the reliability of the measurements because another

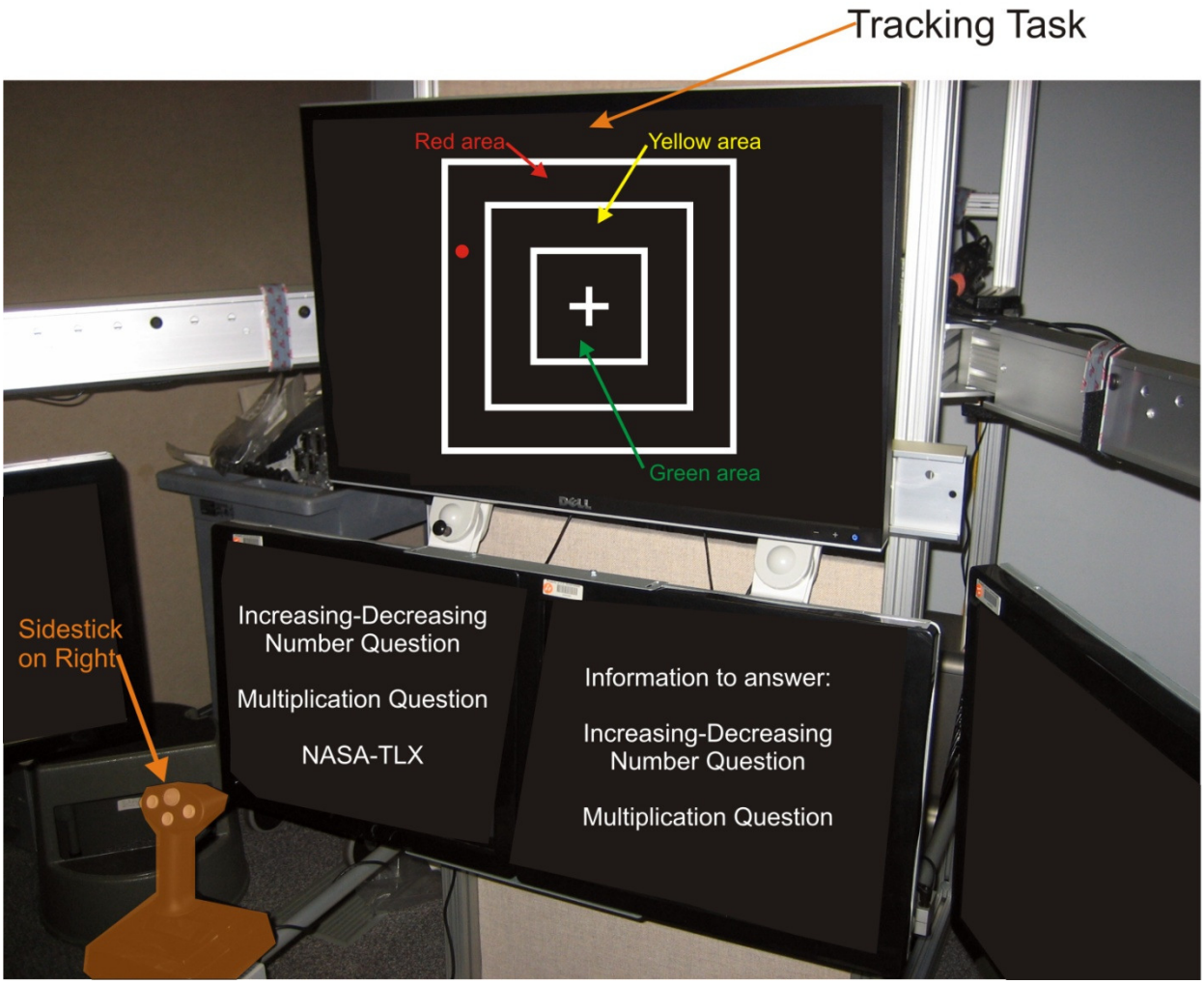


Figure 7. Display Configuration

source of measurement error is not introduced (Bustamante & Spain, 2008). In general, non-pilots rated the workload approximately 8 points lower than pilots (Table 2). This may be due to pilots possibly working harder to keep the target closer to the center of the display than the non-pilots as seen in the pilots' lower target distance from the display center (Fig. 8).

Table 1. Pilot Status Significance for NASA-TLX Ratings

NASA-TLX Measure	$F_{(1, 900)}$	**p-value**
Mental Demand	57.9	≤0.01
Physical Demand	119.1	≤0.01
Temporal Demand	33.0	≤0.01
Performance	9.3	≤0.01
Effort	40.6	≤0.01
Frustration Level	37.5	≤0.01
Combined NASA-TLX	70.7	≤0.01

Table 2. NASA-TLX Rating by Pilot Status

NASA-TLX Measure	Pilot Status	
	Non-Pilot Rating	Pilot Rating
Mental Demand	27.1 ± 1.0	36.2 ± 1.0
Physical Demand	25.7 ± 0.9	38.4 ± 1.0
Temporal Demand	27.8 ± 1.0	34.9 ± 1.0
Performance	32.0 ± 1.1	35.5 ± 1.0
Effort	30.7 ± 1.0	38.5 ± 1.0
Frustration Level	24.9 ± 1.0	32.4 ± 1.0
Combined NASA-TLX	28.0 ± 0.8	36.2 ± 0.9

NOTE: Number formats are NASA-TLX Rating ± 1 standard error of the mean

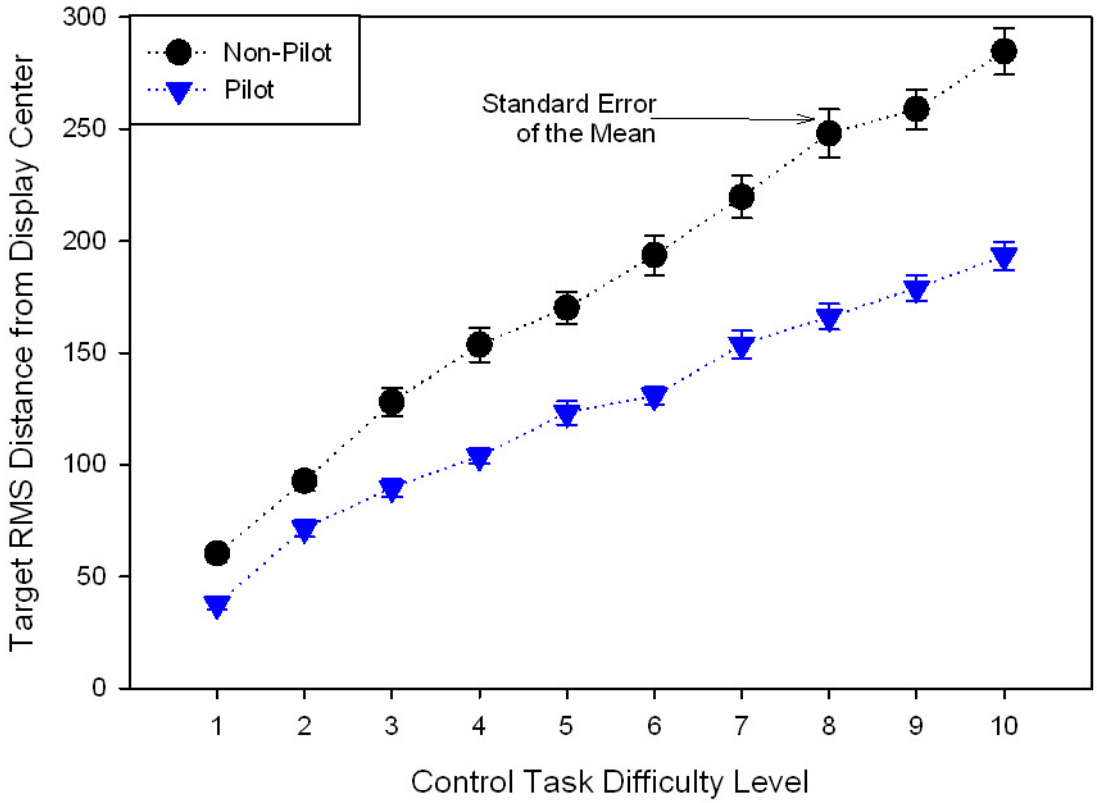


Figure 8. Target RMS Distance from the Display Center by Control Task Difficulty Level and Pilot Status

The control task difficulty level was also found to be significant across all six component measures in the NASA-TLX and the combined rating of the NASA-TLX. This result further verified that the difficulty levels developed for the control task modulated workload. In general, subjective ratings across all six measures in the NASA-TLX and the combined rating of the NASA-TLX increased linearly with an average slope of 5.4 (Table 3 and Figure 9).

Table 3. Linear Regression of NASA-TLX Rating with Control Task Difficulty

NASA-TLX Measure	Control Task Difficulty Linear Coefficient	R^2
Mental Demand	5.3	0.74
Physical Demand	5.4	0.74
Temporal Demand	5.2	0.72
Performance	5.7	0.76
Effort	5.8	0.75
Frustration Level	4.8	0.68
Combined NASA-TLX	5.4	0.80

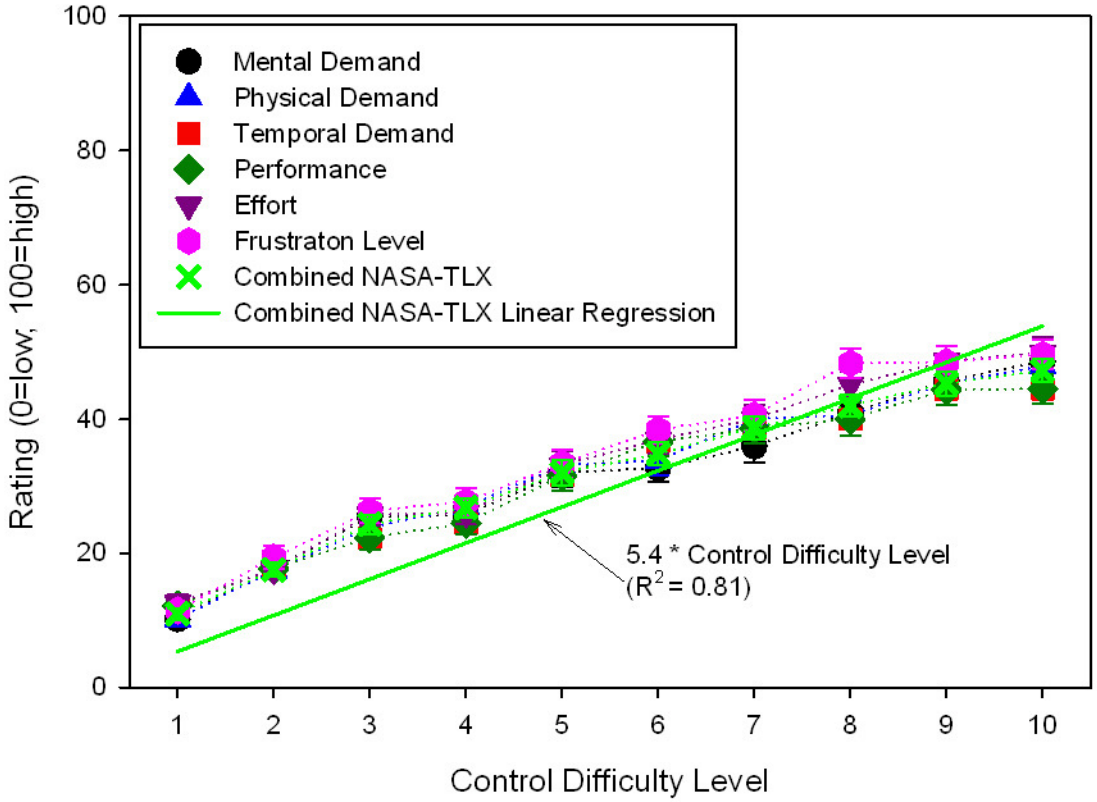


Figure 9. Target RMS Distance from the Display Center by Control Task Difficulty and NASA-TLX Measure

NASA-TLX format was significant for effort ($F_{(4,900)} = 5.90$; $p \leq 0.01$) and frustration level ($F_{(4,900)} = 10.21$; $p \leq 0.01$) (Fig. 10). As seen in Figure 10, the paper and electronic paper formats were different from the random formats. Also, the ratings when using the random formats appear to be more consistent than the ratings for the paper and electronic facsimile formats.

NASA-TLX format by pilot status was also significant for physical demand ($F_{(4,900)} = 6.67$; $p \leq 0.01$) and performance ($F_{(4,900)} = 3.22$; $p \leq 0.02$) (Fig. 11). Again, as seen in Figure 11, the paper and electronic paper formats were different from the random formats. As with effort and frustration level, the physical demand and performance ratings are more consistent with the random formats that with the paper and electronic facsimile formats.

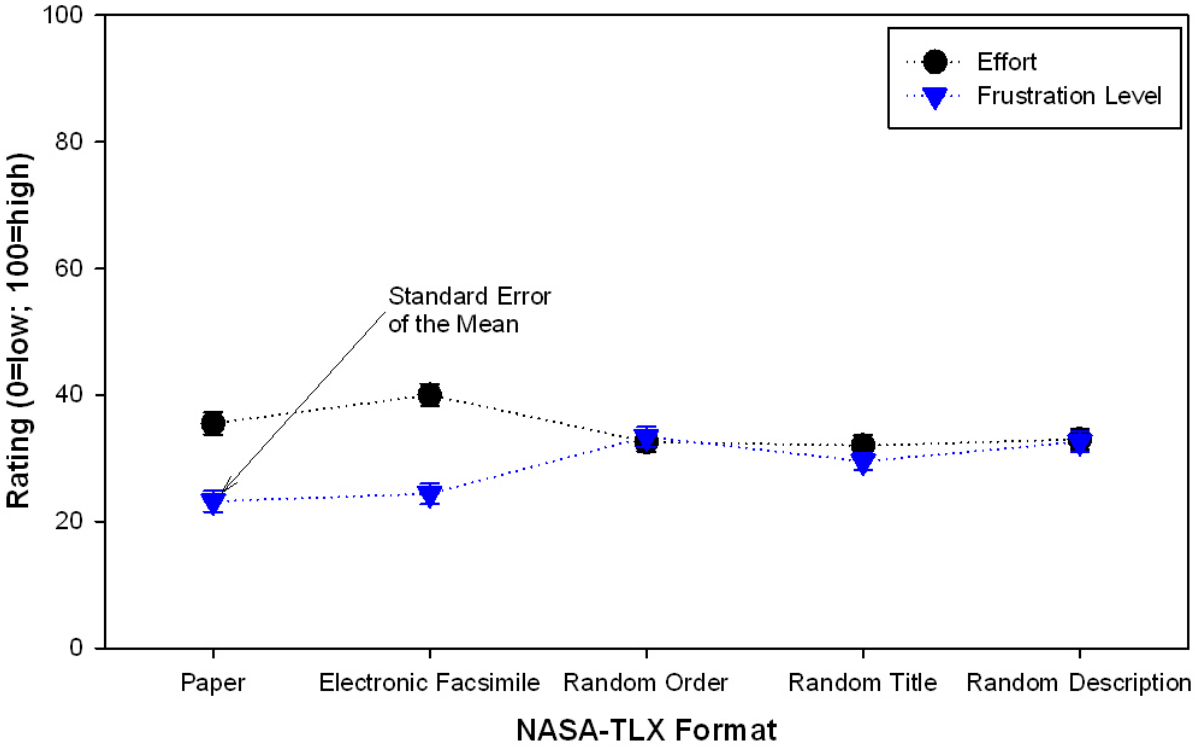


Figure 10. NASA-TLX Rating by Format for Effort and Frustration Level

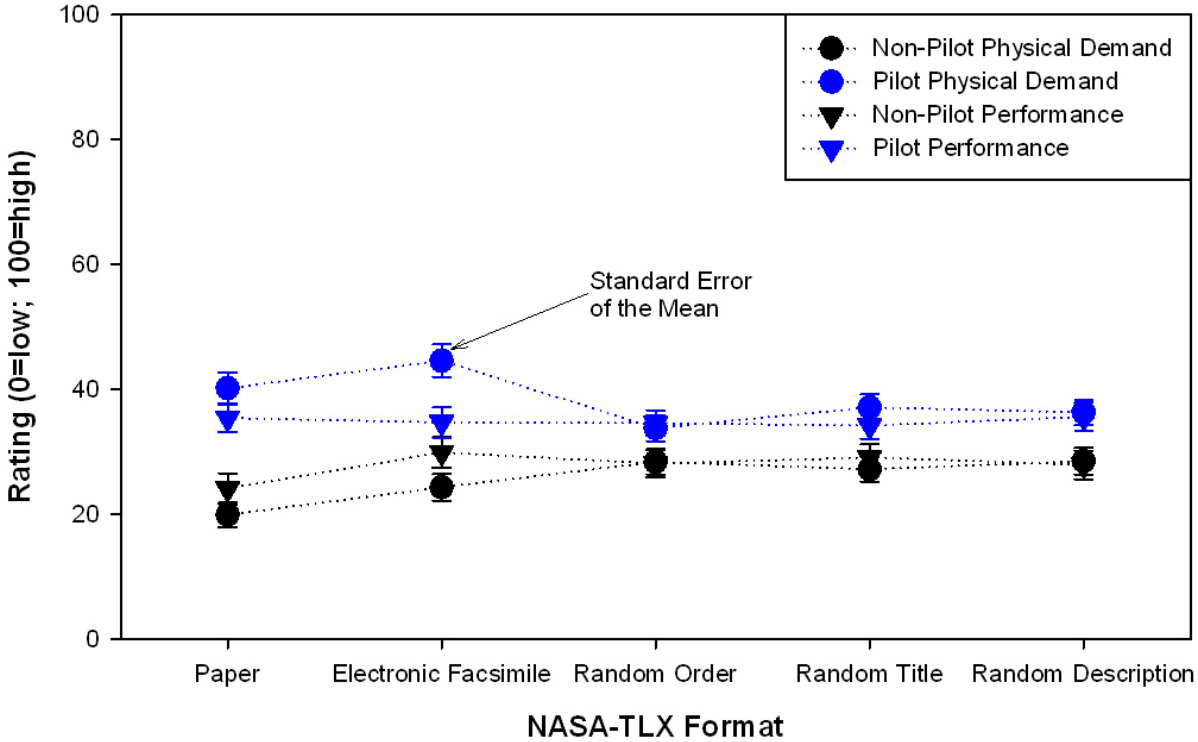


Figure 11. NASA-TLX Rating by Format and Pilot Status for Physical Demand and Performance

Time to Complete NASA-TLX Rating

The time to input the NASA-TLX rating was the fastest for electronic facsimile and random title formats ($F_{(4,900)} = 31.89$; $p \leq 0.01$) (Table 4). This was most likely because the electronic facsimile NASA-TLX has the fewest button pushes and the random title had the least amount of material to read. Furthermore, a Bonferroni test produced the following groupings: random title and electronic facsimile; random; random description and paper.

Table 4. Time to Complete NASA-TLX Rating by Format Type

NASA-TLX Format	Mean	Standard Error of the Mean
Random Title	19.9	0.60
Electronic Facsimile	19.9	0.75
Random	24.3	0.84
Random Description	28.1	0.78
Paper	29.6	0.87

NOTE: Groupings, indicated by color, based on Bonferroni post-test

Subjective Ratings

Subjects were asked about their preferences on using the electronic versions of the NASA-TLX after all the simulation data runs. Subjects compared the electronic versions to the baseline paper version of the NASA-TLX.

Overall, subjects found it easier to make their NASA-TLX rating using the electronic facsimile format when compared to the baseline paper version ($F_{(4, 90)} = 12.85$; $p \leq 0.01$) (Fig. 12). They also indicated that the workload to make their ratings was lower for the electronic facsimile version ($F_{(4, 90)} = 4.49$; $p \leq 0.01$). Lastly, subjects also preferred the electronic facsimile NASA-TLX ($F_{(4, 88)} = 5.14$; $p \leq 0.01$).

Subject Comments

Many subjects commented on their likes and dislikes for each NASA-TLX format (Table 5). Several subjects commented that they liked to see all the scales at once (11 comments) partially because they could then compare their ratings on each scale (5 comments). Some respondents, though, did indicate that with all the scales available at once, the screen was cluttered and there was too much to read (16 comments). Several subjects also commented that they preferred the electronic facsimile format because it entailed the fewest button pushes (7 comments).

Subjects were also asked how often they read the descriptions on the paper NASA-TLX. Pilot status was significant ($F_{(1, 18)} = 6.65$; $p \leq 0.02$) with non-pilots reading the descriptions only 45% of the time while pilots read the descriptions 70% of the time.

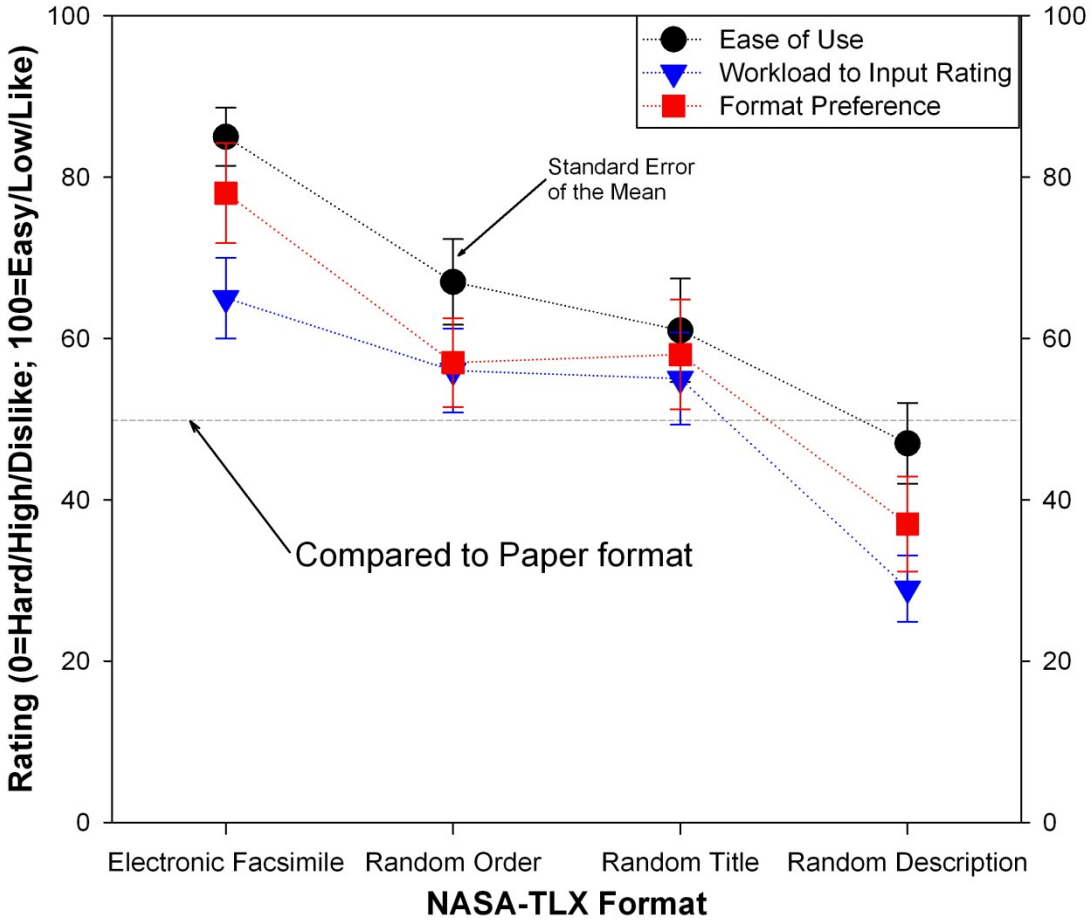


Figure 12. Subjective Ratings

Table 5. Subjects Comments About NASA-TLX Formats

Comment	Count
I want all the scales up (paper and electronic facsimile) so that I can compare my ratings	5
I liked electronic facsimile because it had fewer button pushes	7
I liked paper and electronic facsimile because everything was available	11
There was too much to read on the paper and electronic facsimile	16
I want the descriptions (paper, electronic facsimile, random, random descriptions)	4

Discussion and Concluding Remarks

Many advantages promote electronic questionnaires as a replacement to paper formats. The formats of traditional paper questionnaires have been found to affect a subject's rating (Riley & Wilson, 1990; Wilson & Riley, 1989). Consequently, the transition from paper to electronic format can subtly change results (Noyes & Bruneau, 2007; Noyes, *et al.*, 2004; Noyes & Garland, 2003; Wästlund, *et al.*, 2005). This research looked at how respondents' answers may change by having subjects use five different formats of the NASA-TLX Rating Scale that requires respondents to give scaled ratings.

Results indicated that all NASA-TLX subscales and the combined workload rating linearly increased with the control task difficulty levels used in this experiment. No significant difference between the paper and the electronic replica of the NASA-TLX was found. These data suggests that an electronic analog of the paper TLX form is equivalent. The data also show that non-pilots typically rated the workload lower than pilots did. This may be due to pilots possibly working harder to keep the target closer to the display center than the non-pilots did. Non-pilots were also less likely to read the descriptions contained in the NASA-TLX workload scale. Variations in the electronic formats were tested which took advantage of the media to enforce structure to the rating methodology. The ratings for effort, frustration level, physical demand, and performance were more consistent among the random formats than for the paper and electronic facsimile formats.

Overall, subjects preferred the electronic formats. Several commented that they liked to see all the scales at once so that they could compare their ratings but some respondents did indicate that with all the scales available at once, the screen was cluttered and there was too much to read. Several subjects also commented that they preferred the electronic facsimile format because it entailed the fewest button pushes. This preference is supported by the subjects' ratings for Workload to Input Rating and Ease of Use.

Not surprisingly, the time to input the rating was the fastest for the electronic facsimile and random title formats. The electronic facsimile format entailed the fewest button pushes while the random title format had the least amount of material to read on each screen. Lastly, subjects did not always read the descriptions associated with each scale.

Thus, with a modest time penalty and equal subjective ratings for ease of use, workload to input rating, and format preference, the random format would help alleviate the screen clutter while supplying all the pertinent information about the scale as compared to just having the titles available. This format may also influence subjects to read the description of the scale more often.

Therefore, moving questionnaires from paper to electronic media could change respondents' answers. Specifically, the above results suggest that when using scaled questionnaires, it is best to have all the scales that are related be on the same page with descriptions. This will minimize button pushes and subjects can compare their ratings among the related scales. Plus having the descriptions available may encourage respondents to read and consider them when making their rating. In any case, even with descriptions available, it is very important to go over these descriptions during training because many subjects will not carefully read these during the test.

References

Bustamante, E. A., & Spain, R. D. (2008). *Measurement Invariance of the NASA TLX*. Paper presented at the 52nd Human Factors and Ergonomics Society Annual Meeting, New York, NY.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and Raw Task Load Index (TLX) Correlations: Are Paired Comparisons Necessary? *Advances in Industrial Ergonomics and Safety*, 481-485.

Cooper, G. E., & Harper, R. P. (1969). *The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities*: Technical Report 567, AGARD.

Federal Aviation Administration (August 28, 2008). Electronic Code of Federal Regulations - Title 14: Aeronautics and Space Subpart E-Private Pilots Section 61.103 Retrieved September 2, 2008, from http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=%2Findex.tpl

Harper, R. P., & Cooper, G. E. (1986). Handling Qualities and Pilot Evaluation (Wright Brothers Lecture in Aeronautics). *Journal of Guidance, Control, and Dynamics, 9*(6), 515-529.

Hart, S. G., & Staveland, L. E. (1988). Development of a NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. S. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: Elsevier Science Publishers B. V.

Noyes, J. M., & Bruneau, D. P. J. (2007). A Self-Analysis of the NASA-TLX Workload Measure. *Ergonomics, 50*(4), 514-519.

Noyes, J. M., Garland, K., & Robbins, L. (2004). Paper-Based Versus Computer-Based Assessment: Is Workload Another Test Mode Effect? *British Journal of Educational Technology, 35*(1), 111-113.

Noyes, J. M., & Garland, K. J. (2003). VDT Versus Paper-Based Text: Reply to Mayes, Sims and Koonce. *International Journal of Industrial Ergonomics, 31*, 411-423.

Riley, D. R., & Wilson, D. J. (1990, Aug 20-22). *More on Cooper-Harper Pilot Rating Variability.* Paper presented at the 8th Atmospheric Flight Mechanics Conference, Portland, OR.

Saitek Ltd. (2003). *Saitek evo User Manual: v1.0 SD-evo 260603*: Saitek Ltd.

Trujillo, A. C. (2008). *Paper to Electronic Questionnaires: Effects on Structured Questionnaire Forms*. Paper presented at the HCI International 2009.

Wästlund, E., Reinikka, H., Norlander, T., & 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.

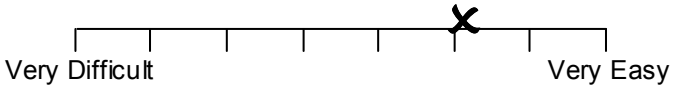
Wilson, D. J., & Riley, D. R. (1989, Aug 14-16). *Cooper-Harper Pilot Rating Variability.* Paper presented at the AIAA Atmospheric Flight Mechanics Conference, Boston, MA.

Appendix A – NASA-TLX Format Final Questionnaire

These questions deal with the NASA-TLX Workload Scale. For all questions, please use the following scales. For all scales, place a mark **anywhere** along the horizontal line of the rating scale like the ones shown below.

The first group of questions asks you about how easy or difficult it is to use the scale compared to the paper version.

Compared to the paper format, using an electronic TLX rating format to make your rating was



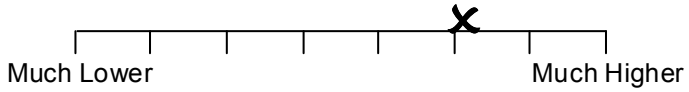
For the scale ends:

Very Difficult = it is hard, frustrating or stressful, and/or it requires a lot of time to use the specified format

Very Easy= it is simple to determine, it is effortless or straightforward, and/or it is readily apparent in a short time to use the specified format

The second group of questions asks you about the workload involved in using the scale compared to the paper version.

Compared to the paper NASA-TLX rating format, the workload of using an electronic NASA-TLX rating format to make your rating was



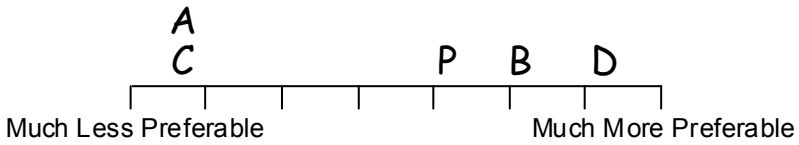
For the scale ends:

Much Lower = it is easier, it takes less effort, and/or it is less stressful to use the specified format

Much Higher = it is harder, it takes more effort, and/or it is more stressful to use the specified format

The third and fourth group of questions deals with how preferred and how well you feel your performance matched your ratings.

On the scale below, order the TLX formats (A, B, C, D, P=paper) by preference.



For the scale ends:

Much Less Preferable = the format is less liked, less desired

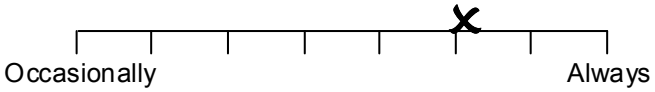
Much More Preferable = the format is more likes, more desired

Rating Does Not Match Performance = It was hard or difficult to use the format to give the rating that you feel best matches your performance

Rating Matches Performance = It was simple or easy to use the format to give the rating that you feel best matches your performance

The last question deals with how you used the formats.

On the paper version, I read each of the descriptions.



For the scale ends:

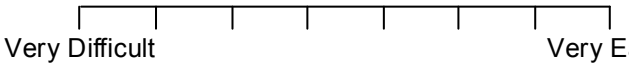
Occasionally = you did it once in a while, probably more often at the beginning

Always = you did it for nearly all of the runs

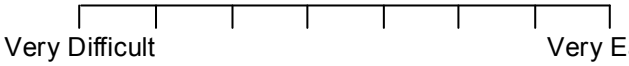
Remember that you may place a mark anywhere along the horizontal line. For the "Much Less Preferable" to "Much More Preferable" and for the "Rating Does Not Mach Performance" to "Rating Matches Performance" questions, please use the letters or numbers mentioned in the question (*e.g.*, A, B, C, and P for the example given above). Furthermore, if you want to rate 2 or more of the displays the same, just stack the appropriate display numbers vertically as shown in the example above. Do not hesitate to place marks in the end regions of the horizontal line if either of these ratings accurately represents your subjective perception. There are no right or wrong answers.

Question 1

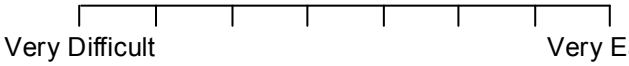
a. Compared to the paper NASA-TLX rating format, using NASA-TLX rating format 1 to make your rating was



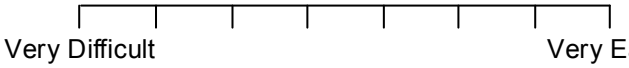
b. Compared to the paper NASA-TLX rating format, using NASA-TLX rating format 2 to make your rating was



c. Compared to the paper NASA-TLX rating format, using NASA-TLX rating format 3 to make your rating was

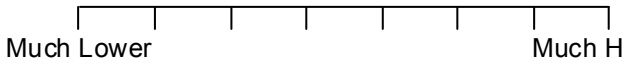


d. Compared to the paper NASA-TLX rating format, using NASA-TLX rating format 4 to make your rating was

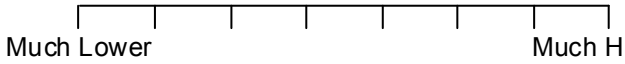


Question 2

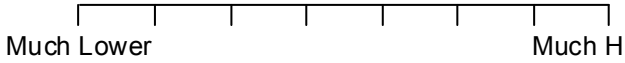
a. Compared to the paper NASA-TLX rating format, the workload of using NASA-TLX rating format 1 to make your rating was



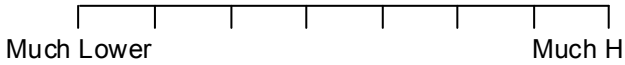
b. Compared to the paper NASA-TLX rating format, the workload of using NASA-TLX rating format 2 to make your rating was



c. Compared to the paper NASA-TLX rating format, the workload of using NASA-TLX rating format 3 to make your rating was

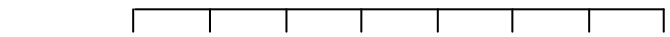


d. Compared to the paper NASA-TLX rating format, the workload of using NASA-TLX rating format 4 to make your rating was



Question 3

a. On the scale below, order the NASA-TLX formats (1, 2, 3, 4, 5=paper) by preference.



Much Less Preferable Much More Preferable

b. For the format you rated the most preferable, why was it the most preferable? _____

_____

_____

_____

_____

_____

c. For the format you rated the least preferable, why was it the least preferable? _____

_____

_____

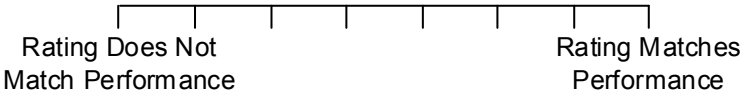
_____

_____

_____

Question 4

a. On the scale below, order the NASA-TLX formats (1, 2, 3, 4, 5=paper) by how well you think your rating matches your performance.



b. For the format you rated the most closely matching your performance, why does it match? ___

_____

_____

_____

_____

_____

c. For the format you rated the least closely matching your performance, why does it <u>not</u> match?

_____

_____

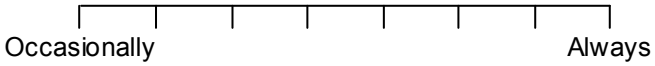
_____

_____

_____

Question 5

On the paper version, I read each of the descriptions.



REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01- 08 - 2011	Technical Memorandum	

4. TITLE AND SUBTITLE

Evaluation of Electronic Formats of the NASA Task Load Index

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Trujillo, Anna C.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

457280.02.07.07.02.01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

NASA Langley Research Center
Hampton, VA 23681-2199

8. PERFORMING ORGANIZATION REPORT NUMBER

L-19919

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSOR/MONITOR'S ACRONYM(S)

NASA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

NASA/TM-2011-217166

12. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified Unlimited
Subject Category 01
Availability: NASA CASI (443) 757-5802

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Paper questionnaires are being replaced by electronic questionnaires. The primary objective of this research was to determine whether electronic formats of paper questionnaires change subjects' ratings and, if so, how the ratings changed. Results indicated that there were no statistically significant differences in self-assessment of workload when using the electronic replica or the paper format of the NASA-TLX scale. Variations of the electronic formats were tested to enforce structure to the TLX scale. Respondents had more consistent ratings with these alternative formats of the NASA-TLX. Non-pilots, in general, had lower workload ratings than pilots. The time to input the rating was the fastest for the electronic facsimile and random title formats. Also subjects preferred the electronic formats and thought these formats were easier to use. Therefore, moving questionnaires from paper to electronic media could change respondents' answers.

15. SUBJECT TERMS

Electronic format; NASA Task Load Index; Ratings; Workloads

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			STI Help Desk (email: help@sti.nasa.gov)
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	33	(443) 757-5802