DOT/FAA/TC-06/13

Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

Optimal Design of Event Lists (ODELs) Phase 2: Does List Length Impact Performance?

Vicki Ahlstrom, Human Factors Team – Atlantic City, ATO-P
Bonnie Kudrick, L-3 Communications, Titan Corporation

December 2006

Technical Report

U.S. Department of Transportation
Federal Aviation Administration

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT/FAA/TC-06/13		

4. Title and Subtitle Optimal Design of Event Lists (ODELs) Phase 2: Does List Length Impact Performance?	**5. Report Date** December 2006 **6. Performing Organization Code** AJP-6110
7. Author(s) Vicki Ahlstrom, Human Factors Team – Atlantic City, ATO-P Bonnie Kudrick, L-3 Communications, Titan Corporation	**8. Performing Organization Report No.** DOT/FAA/TC-06/13
9. Performing Organization Name and Address Federal Aviation Administration Human Factors Team – Atlantic City, ATO-P William J. Hughes Technical Center Atlantic City International Airport, NJ 08405	**10. Work Unit No. (TRAIS)** **11. Contract or Grant No.**
12. Sponsoring Agency Name and Address Federal Aviation Administration Human Factors Research and Engineering Group 800 Independence Avenue, S.W. Washington, DC 20591	**13. Type of Report and Period Covered** Technical Report **14. Sponsoring Agency Code** ATO-P

15. Supplementary Notes

16. Abstract

This report documents the second study on the optimal design of event lists (ODELs). This study investigated the impact of event list length on user performance in searching for information. Participants searched for information in each of four different lists, each with a different length: 5 items, 15 items, 25 items, and 35 items. Researchers measured task completion time, accuracy, subjective ratings, and rankings. The shortest list (5 items) led to significantly slower response rates. The slower response rates appear to be due to the time it took users to scroll through the table. Participants made significantly more errors on the longer lists than on the shorter lists. Subjective data mirrored the performance data. Users preferred the mid-size lists (15 items and 25 items) and rated them higher than the longest list (35 items) and shortest list (5 items). Based on the results of this ODELs study, we would recommend a list size of 15 or 25 items over a list size of 5 or 35 items if speed and accuracy are important to the task.

17. Key Words	18. Distribution Statement
Event Lists List Length List Size Technical Operations Visual Search	This document is available to the public through the National Technical Information Service, Springfield, Virginia, 22161. A copy is retained for reference by the William J. Hughes Technical Center Library.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	21	

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Table of Contents

List of Illustrations

Executive Summary

Technical Operations (TO) is the part of the Federal Aviation Administration responsible for maintaining the thousands of systems and equipment that allow aircraft to move from one airport to another safely and efficiently. As part of their maintenance tasks, TO Specialists routinely engage in visual searches for information on system events. This information often appears in an electronic table, referred to in the TO environment as an event list.

As the National Airspace System continues to be modernized, there is an increased reliance on automation for monitoring system and equipment status. The shift from the visual inspection of wires and fuses to automated electronic monitoring increases the quantity and complexity of event lists. As the number of event lists increase, TO Specialists are faced with the question of where to put all of the lists of information. If event lists are made smaller, they will take up less space on the screen and, thus, more systems could be monitored from a single screen. Making the list size smaller, however, could involve tradeoffs, and designers do not want to alter the displays in such a way that they negatively impact user performance. Therefore, it is important to know whether changing the length of the list will impact user performance. This study asks whether list length impacts user performance when searching for information.

To address this question, we constructed a search task using four representative list sizes: 5 items, 15 items, 25 items, and 35 items. The lists mimicked event list displays in the TO environment, although the data within the lists were fictional. For each trial, an information item appeared as a Find statement. Participants searched for the item within the event list. The researchers collected data on response time and accuracy. After the participants completed all four trials, they rated each of the four conditions and ranked them in order of preference.

Results indicated that list length did have an impact on search performance. Participants had significantly longer mean response times for the shortest list. Further analysis indicated that this result could be explained by the amount of scrolling that was necessary, with longer scrolling distances leading to longer response times. List length had an impact on accuracy of search performance as well. Participants made significantly more mistakes on the longest list length (35 items) than they did on the mid-size lists (15 items and 25 items). They also rated mid-size lists higher than short lists (5-item lists) and long lists (35-item lists) for ease of task completion. The mid-size list lengths received higher overall rankings. These results suggest that the optimal list size is between 15 or 25 items.

For the conditions that we researched in this study, a list size of 15 items led to optimal performance followed by a list size of 25 items. The increased scrolling necessary to find items in the 5-item list led to the slowest performance and, although performance was fast in the 35-item list size, it was more error prone than the other conditions. Based on the results of this study, we recommend that designers use an event list that shows 15 or 25 items rather than 5 or 35 items to optimize search time and accuracy.

One of the limitations of this study is that we did not include an automated search as a variable. In this study, we forced the participants to visually search for items in the event lists. We made this decision based on our observations of TO Specialists in operational settings. We do not know, however, whether automation could counterbalance the negative performance aspects found for 5- item and 35-item lists in this study.

1. INTRODUCTION

Event lists are one of the key methods used for a monitor and control environment to organize large quantities of information. As the amount of information available to a single individual increases, so does the importance of having the information organized in such a way that the individual user can quickly and easily locate items of interest. This is especially important in an environment such as Technical Operations (TO), where finding the right information for the task at hand is important for the safety and efficiency of the National Airspace System (NAS).

Some TO Specialists monitor more than 20 different screens of information. This has led to suggestions for consolidation of monitoring so that, instead of having a dedicated display for each system, a TO Specialist could monitor multiple systems from a single display. This would necessitate putting information from multiple systems on a single screen. In order to do this, event lists may have to be limited in size. If event lists are small, they take up less screen space and, thus, more systems could be monitored from a single screen. Making the list size smaller, however, could involve tradeoffs. The length of the event list could have an impact on the ease with which TO Specialists can make selections. Designers do not want to alter the displays in such a way that they negatively impact performance. As finding items in a list is a frequent task, even a change that slowed performance slightly could have a major impact on overall system performance. Thus, decision makers in the TO organization have asked human factors researchers from the Human Factors Team – Atlantic City to investigate whether list length impacts performance. This document is the second in a series to study the optimal design of event lists (ODELs). This phase of the ODELs study asks whether the length of event lists, also known as tables of information, impacts user performance when searching for information.

A number of studies have investigated the size issue in different contexts with mixed results. Zelinsky and Sheinberg (1995) compared search times for O- and Q-shaped objects in 5-item lists and 17-item lists. They found that participants made significantly fewer saccades in a search task with the 17-item list than in a search task with the 5-item list. They also found that participants had significantly longer mean saccade latencies for the 17-item display than for the 5-item display. Their study implies that smaller-sized displays would produce better search performance. Conversely, Hendrickson (1989) looked at search in the context of window size for a menu-based system and found that fixation durations were significantly longer for 4-item windows than for 8- or 16-item windows.

A complicating factor is that to display the same amount of information with a smaller list, the users must scroll through the list to view information that is not readily visible. According to Zhai, Smith, and Selker (1997), the time it takes to acquire the scroll bar varies in relation to the distance that the cursor has to travel and the size of the target. Chipman, Bederson, and Golbeck (2004) found that acquiring the scroll bar can take up to 2 seconds and has a negative impact on performance. Additionally, to use a scroll bar to move through a document, the user must take his or her eyes off the target to focus on the scroll bar. Longer lists require less scrolling because more information is visible.

With this in mind, we expect that it would take a longer time to find items on shorter lists than on the longer lists because the user would be required to make both an eye movement and a motor movement to scroll the “window.” For the longer list display, the user would only be required to make an eye movement to locate an item, and then select the item with the mouse.

Researchers have found that eye movement is significantly faster than motor movement (Sibert & Jacob, 2000; Ware & Mikaelian, 1986).

1.1 Purpose

This is the second study in a series of studies with the overall purpose of identifying the ODELs. This study examines the impact of list length on user performance.

1.2 Task Design

In formulating this task, we visited several TO locations and looked at the length of the lists currently used by TO personnel. In addition to observational data, we conducted structured interviews with TO personnel at each site. These interviews provided us with information on specific tasks related to how the TO personnel use event lists. We used these data to structure the lists and tasks for the experiments.

Based on the observational data and feedback from the subject matter experts at the TO locations, we found that the TO Specialists use the event lists for two general tasks: (a) searching for specific items of information and (b) monitoring events or systems for change. The current study focuses on searching for specific items of information.

We also used our familiarization visits of TO facilities to collect information on list characteristics of current systems. We found that the majority of the lists that TO Specialists use were organized in chronological order with the most recent event appearing at the bottom of the list and older events, scrolling out of view, at the top of the list.

In addition to looking at the list characteristics of current systems, we looked at the ambient environment. TO facilities are often collocated with air traffic control facilities. These facilities are often dimly lit. We measured ambient light levels at an operational facility so that we could replicate the lighting in the laboratory for the experiment.

2. METHOD

This section describes the methods used in this study. Subsections describe the participants, apparatus, and procedures.

2.1 Participants

Twenty-six people took part in the study. Participants ranged in age from 20 to 63. Fourteen of the 26 participants had taken part in Phase 1 of the ODELs study (Ahlstrom & Kudrick, 2006). Participants reported having normal or corrected to normal vision. None of the participants wore bifocals. We eliminated the data from two participants because of technical problems encountered while collecting oculometer data. We obtained valid data from 7 females and 17 males.

We used a convenience sample, recruiting our participants from those available at the Federal Aviation Administration's (FAA's) William J. Hughes Technical Center (WJHTC). We chose this sample of individuals instead of TO Specialists for several reasons; one reason was that of economy. We used local personnel to eliminate the travel cost necessary to bring in participants. We were also concerned that TO personnel may have developed biases after many years of working with a particular system. Our study population was similar to the TO personnel in many characteristics such as age and level of computer experience, but without the operational experience in finding information in event lists.

2.2 Apparatus

We conducted this study at the Research, Development, and Human Factors Laboratory at the FAA's WJHTC. We wrote a custom data collection program in Visual Basic to display the probe questions and event lists and collect reaction time data. The leftmost monitor displayed probe questions one at a time in the form of Find statements, and the rightmost monitor displayed the event list. We designed the Find statements to mimic the information searches done by the TO personnel when using event lists. The characters were approximately 2 mm high.

We used two computers for this study, one for the oculometer and one for the experiment. Prior to each session, we synchronized the two computers. The participants wore an Applied Science Laboratories Model 5000 oculometer comprised of eye and head tracking components. The oculometer captured eye and head movements while recording x, y, and z point-of-gaze coordinates at the rate of 60 Hz (Ahlstrom & Friedman-Berg, 2005). We used Screen Pro recording software to capture events on the screen throughout the experiment. To mimic the lighting in TO environments, we dimmed the overall ambient lighting level (approximately +.38 foot-lamberts).

The table of information contained 100 rows of event information. Rows were alternately white or gray (see Figure 1). There were four different window sizes, varying the number of rows that the participant could see at once. The four sizes were 5 items, 15 items, 25 items, and 35 items. The tables of information were arranged in reverse chronological order with the most recent event at the bottom of the list. The window always started at the bottom of the list, showing the participant the most recent events. The scroll bar slider was always located at the bottom of the scroll bar.

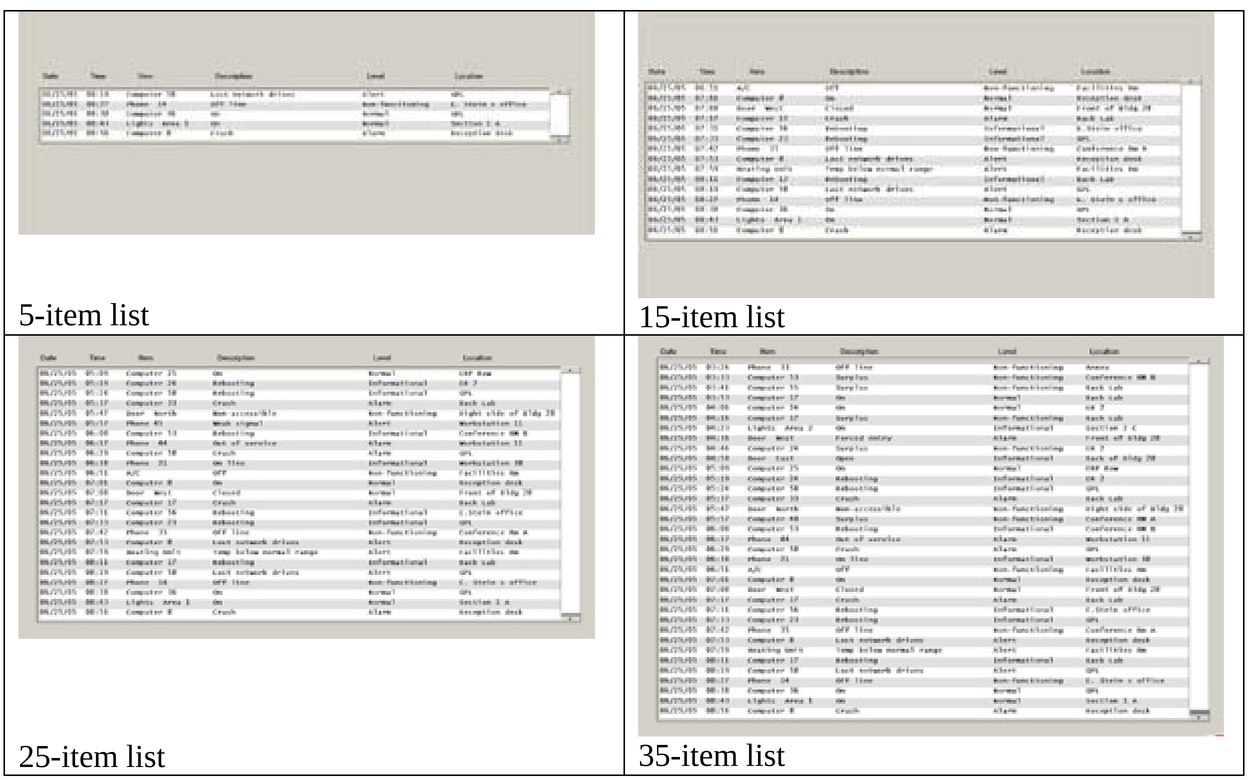

Figure 1. Examples of the four window size conditions studied.

2.3 Procedure

A researcher provided each participant with a printed statement of informed consent. The participants read and signed the informed consent form before continuing with the study. A researcher then read instructions to each participant. We asked participants to provide background information by filling out a paper questionnaire. Information on the questionnaire included gender, job title, number of years in current position, and any vision problems.

Prior to each session, a researcher placed the oculometer on the participant's head and began calibration procedures. The researcher calibrated the system using a nine-dot calibration grid (Willems, Heiney, & Sollenberger, 2005). The calibration screen consisted of a series of nine numbered black dots on a white screen that were displayed across the two screens. Occasionally, the researcher would have to repeat this process until certain that the calibration was accurate. The same researcher performed the calibration for all of the participants. We will present the detailed results of the oculometer data in a different report.

Previous experience with the oculometer showed that it is preferable to limit the time that the participant is in the study with the oculometer to intervals of less than 1 hour. Although the oculometer data is not the focus of the current report, it was a limiting factor in the data collection trial duration. We ran a pilot study to determine the time that it would take to test the four conditions. Based on the results of the pilot study, we decided to use 15 different Find statements per condition. Although each participant had the same Find statements for each of the four conditions, we randomized the order of presentation.

The method was nearly identical to an earlier study (Ahlstrom & Kudrick, 2006). In order to maintain congruency with tasks described by TO Specialists at the field sites visited, we created the experiment to mimic a monitor and control interface for a building. Because we were not using TO Specialists, we were concerned that lack of familiarity with acronyms and terminology used for TO systems would impact the task. Thus, instead of using actual NAS events, we created events related to the functioning of a building such as the monitoring of phones, doors, lighting, and air conditioning.

Each participant completed one practice session to become familiar with the experimental procedures. The practice session consisted of all four conditions. Total time was approximately 1 hour. During and immediately after the practice session, we encouraged the participants to ask questions in order to clarify their understanding of their responsibilities as participants.

In each trial, we presented each participant with a Find statement on the leftmost monitor of the two-screen display. The event list appeared on the rightmost monitor. The participant read each Find statement then searched the event list for the information item. Once we presented a Find statement, the task of the participant was to quickly and accurately locate the Find statement in the event list. If desired, the participant could look back at the Find statement at any time by shifting the focus of their gaze to the leftmost monitor. When the participant found the item, they used the mouse to click on the item. The search was self-terminating. A feedback message appeared at the bottom of the screen indicating whether the answer was correct. Once the participant answered correctly, a new Find statement appeared on the left screen. If the participant selected an incorrect answer, feedback appeared at the bottom of the screen that the answer was incorrect. This process continued until the participant chose the correct answer. Each participant took part in all four conditions. We randomly assigned the order of the conditions. At the conclusion of the experimental session, the participants rated and ranked each list. We also gave them the opportunity to provide comments on any aspect of the study.

3. RESULTS

This section describes the results of the response time measures, the accuracy measures, and the subjective data including both the rating and ranking data. A separate report will present the eye movement data results.

3.1 Task Completion Time

Figure 2 shows the mean response time and 95% confidence intervals for each of the four conditions (Cousineau, 2005; Loftus & Masson, 1994). A pairwise comparison demonstrated a significantly longer response time for the 5-item condition than the other three conditions (all p values < .001). We believe that the increased response time in the 5-item condition was due to the amount of scrolling required. We put the target locations into bins of 20 for the 5-item condition. Items in the first bin (locations 1-20) would require less scrolling than items in the second bin (locations 21-30), and so on.

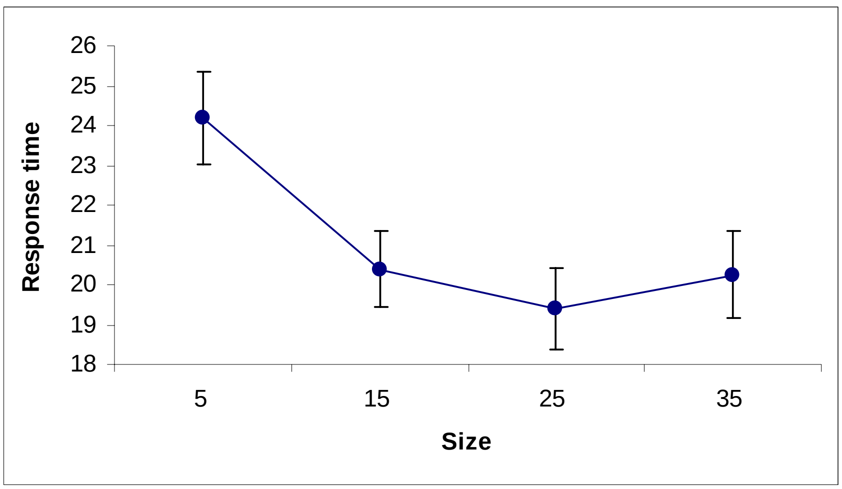

Figure 2. Mean response time and 95% confidence intervals for each condition.

When we plotted the response time against the scrolling distance, we found an r-square of .97 for the 5-item condition (see Figure 3). As indicated in Figure 3, distance 1 = target locations 1-20; distance 2 = target locations 21-40; distance 3 = target locations 41-60; distance 4 = target locations 61-80; and distance 5 = target locations 81-100.

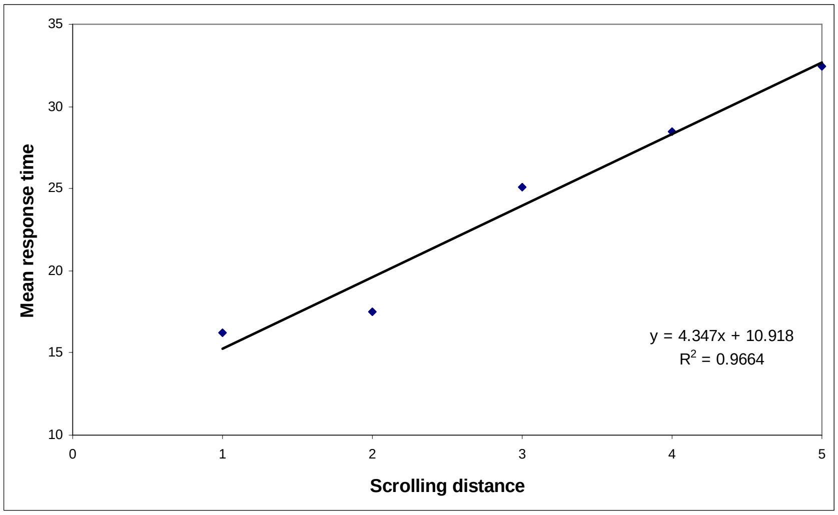

Figure 3. Mean response time as a function of scrolling distance for the 5-item list condition.

We performed a similar analysis for the 35-item condition. For the 35-item condition, we put the items into one of three bins. Bin 1 was for targets in location 1-35, which required no scrolling. Bin 2 was for targets in locations 36-70, which required up to one page of scrolling. Bin 3 was for targets in locations 71-100, which required more than one page of scrolling. We plotted response times by location for items in the three bins (see Figure 4). The *r*-square for response time by location was .98, suggesting that scrolling accounted for 98% of the variance. As indicated in Figure 4, distance 1 = target locations 1-35, which required no scrolling; distance 2 = target locations 36-70, which required up to one page of scrolling; and distance 3 = target locations 71-100, which required more than one page of scrolling.

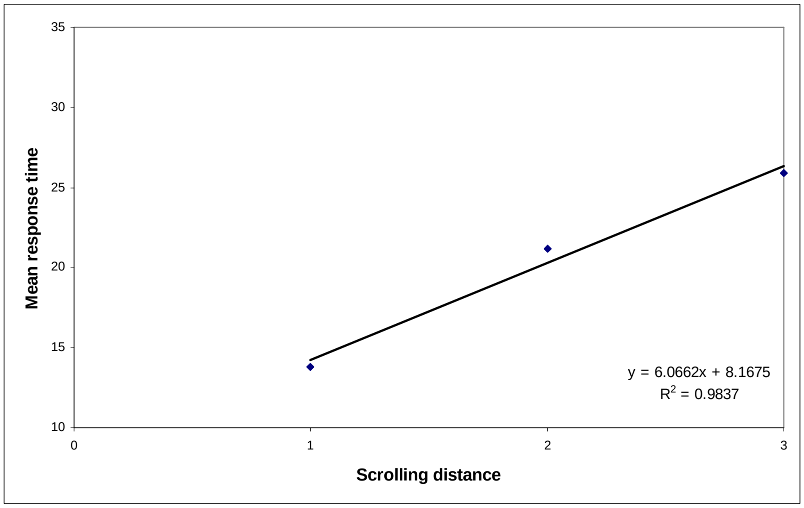

Figure 4. Mean response time as a function of target location for the 35-item list condition.

The high *r*-square indicates that the target distance explained a significant amount of the variance. It is possible, however, even without scrolling, that the target distance would explain a significant amount of the variance. This would be the case if a participant were conducting a serial search, for example, from the bottom to the top of the list. If the increase in response time were due to target location and not scrolling, items at the bottom of the list would have faster response times than items at the top of the list when all of the items were presented on a single page (requiring no scrolling). In order to address this issue, we plotted target location against response time for each of the four conditions for targets that were visible without scrolling. Thus, for the 5-item condition, we plotted the first 5 items against response time. For the 15-item condition, we plotted the first 15 items against response time, and so on.

Figure 5 shows the response times by location for the first 35 items in the 35-item list condition. We fitted the data with a simple regression line and analyzed it using a general linear model. We found that the slopes of the regression lines were not statistically significantly different than zero, which indicated that in the absence of scrolling there was no statistical significance of position on response time.

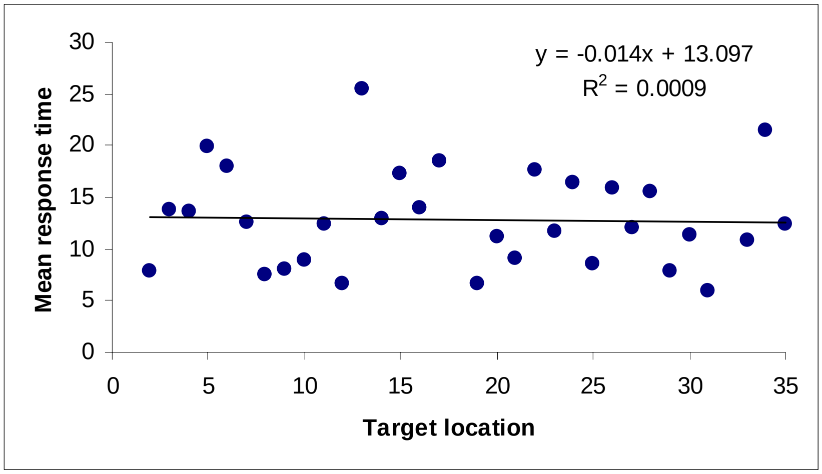

Figure 5. Mean response time as a function of item location for target locations in the 35-item condition.

3.2 Accuracy

We determined the accuracy by measuring the number of errors made for each condition, as accuracy and errors are inversely related. As Figure 6 shows, there was a decrease in the number of errors in the 25-item condition, followed by an increase in errors in the 35-item condition. There was no significant difference between the number of errors made in the 5- and 15- item conditions, the 5-and 25-item conditions, or the 15- and 25-item conditions. A chi-square test showed that the differences between the 5- and 35-item conditions and the 15- and 35-item conditions were statistically significant at a $p < .05$ level ($\chi^2 = 4.40$, $df = 1$, $p < .036$). The number of errors between the 25-item condition and the 35-item condition was also statistically significant ($\chi^2 = 6.14$, $df = 1$, $p < .013$). Participants made more errors in the 35-item condition than in any of the other conditions.

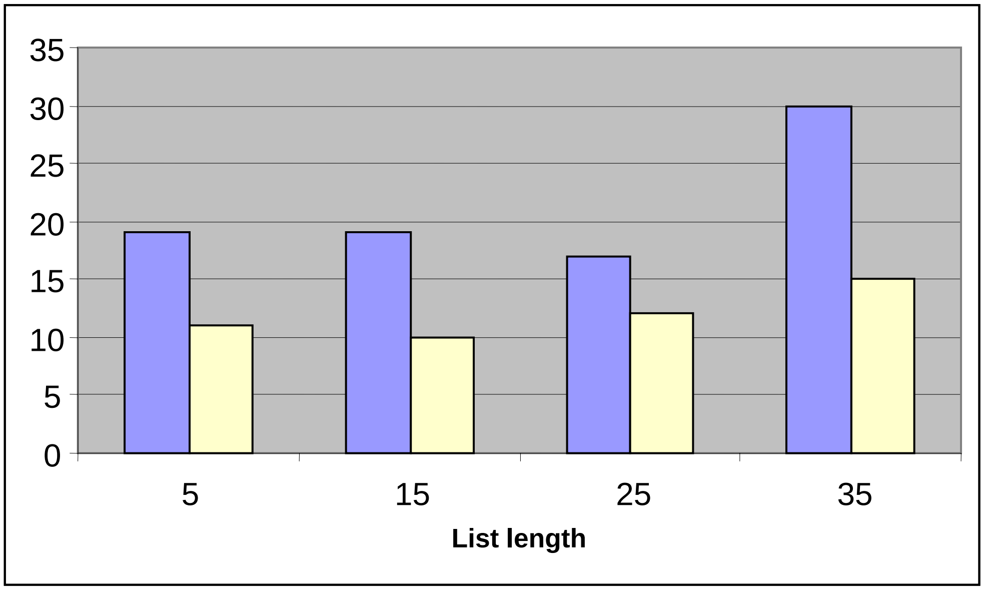

Figure 6. The number of incorrect responses (dark bar) and the number of participants making incorrect responses (light bar) in each of the four conditions.

In addition, Figure 6 shows a decrease in the number of people making errors in the 15-item condition, increasing slightly in the 5- and 25-item conditions. There was a statistically significant difference between the number of people making errors in the 35-item condition versus the number of people expected to make errors based on the frequencies in the 15-item condition ($\chi^2 = 4.44$, $df = 1$, $p < .035$). Chi-square tests between the other conditions did not show statistical significance. More people made errors in the 35-item condition than in the 15-item condition.

3.3 Subjective Rating Data

Participants rated the task difficulty for the four conditions. Combining the easy categories (Somewhat Easy and Very Easy) and the difficult categories (Somewhat Difficult and Very Difficult), we fit a polynomial line to the data (see Figure 7). The polynomial line shows an increase in the easy rating (E) peaking at the 25-item mark and then decreasing toward the 35-item mark; the inverse is true for the difficult rating (D). The number of participants rating the condition as difficult starts high and then quickly decreases, reaching a minimum at the 25-item mark before increasing again at the 35-item mark. We performed a pairwise comparison allowing for multiple comparisons, demonstrating significance between the 5-item condition and the 15- and 25-item conditions ($p < .001$).

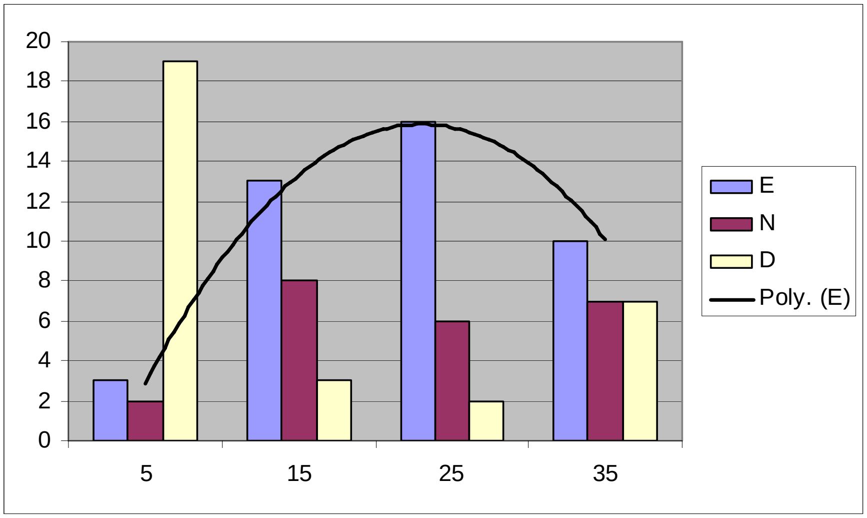

Figure 7. Subjective ratings of easy (E), neutral (N), and difficult (D) for the four conditions with a polynomial line fitted to the easy (E) data.

3.4 Ranking Data

Study participants ranked the four conditions based on their preferences with 1 = most preferred and 4 = least preferred. As Figure 8 illustrates, the 15-item condition was significantly preferred over the 5-item condition, and the 25-item condition was preferred over both the 5- and 35-item conditions. We conducted pairwise comparisons to test for differences in the participants' preferences for the four conditions. The participants preferred the 15-item condition over the 5-item condition (p = .003) and preferred the 25-item condition to both the 5-item (p = .024) and 35-item (p = .039) conditions. Note that no participants ranked the 15- or 25-item conditions fourth.

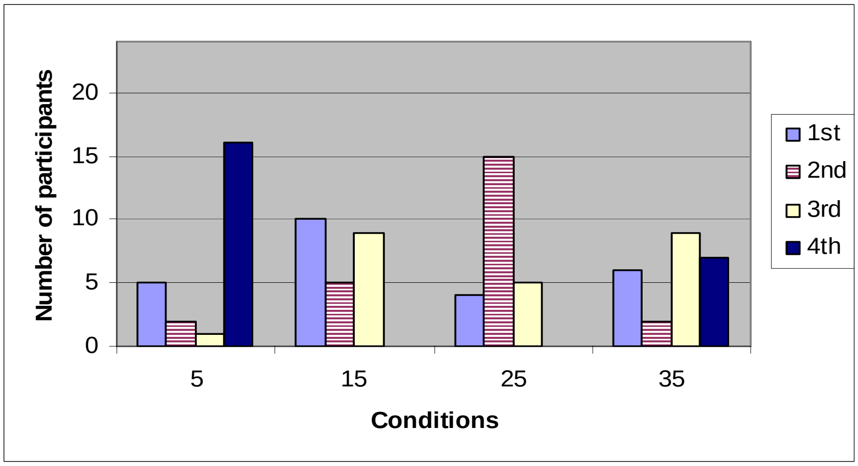

Figure 8. The number of participants who ranked each condition first, second, third, and fourth.

4. CONCLUSION

The results of this study show that size does indeed matter for lists. Response times were significantly slower for lists of five items compared to lists of longer lengths. As response times increased linearly with the location of the item in the list, the results suggest that the longer response times are due to scrolling. These findings imply that as the size of the list decreases beyond a certain level (15 items), the amount of scrolling required to find the item leads to significantly increased response time. We found that, in absence of scrolling, response time did not appear to be significantly affected by position.

Although there were no significant differences in response time between the 15-, 25-, and 35-item lists, there were differences in error rates. There were significantly more errors in the longest list (35 items) over the lists that were shorter in length. In order to separate the tendency for errors in a certain condition from the frequency of errors across individuals, we analyzed not only the number of errors but also the number of participants making the errors. Significantly, more people made errors in the 35-item condition than in the 15-item condition. These findings suggest that as the list length increases beyond a certain length (25 items), not only do the number of errors increase, but the number of people making errors increases as well. This result implies more of a global effect rather than an individual effect.

The subjective rating data are in congruence with the response time and accuracy data. Participants rated the 5-item and 35-item lists as more difficult than the 15-item and 25-item lists. Participants also ranked the different conditions. More participants ranked the 15-item condition first (most preferred) and the 25-item condition second than any other condition. None of the participants ranked the 15- or 25-item conditions fourth (least preferred).

Based on the results of this study, it would appear that a list size of approximately 15 items would lead to the fastest performance times and least errors of the four sizes studied, followed by a list size of 25 items. Lists displaying only 5 items out of 100 available should probably be avoided if time is a critical factor. Designers should probably avoid lists that are 35 items in length if accuracy is most important.

In operational applications, there are often tradeoffs and compromises. It is likely that there will be some systems and situations requiring the use of list sizes with more than 25 items. If a system requires a list size of more than 25 items, we recommend the designer implement some way to counterbalance the potential increase in errors. Although we have not tested the efficacy of any potential means for counterbalancing the effects found in this study, one potential for decreasing search time is by allowing for an automated list search such as a "Find" function.

Additionally, we recommend that users avoid scrolling lists of 5 items. If the system requires a list of less than 15 items, then the designer should be aware of the potential negative impact on performance and should provide an effective means for counterbalancing the increased need for scrolling. Although we did not address the issue in this study, prolonged dragging motions such as those required for frequent and extended scrolling not only impact performance, but can also have ergonomic consequences, such as increased carpal tunnel pressure (Ahlstrom & Kudrick, 2005; Ahlstrom & Longo, 2003).

We found that, in absence of scrolling, response time did not appear to be significantly affected by position. We could take this result as circumstantial evidence that the participants did not conduct a serial search of items either from top to bottom or from bottom to top. Analysis of the search behavior, however, was beyond the scope of the present paper. We intend to explore search behavior in more detail in a companion paper through analyzing the eye movement data collected in this study using the oculometer.

Additionally, we did not investigate whether the direction of the scrolling influenced our findings. Charness, Bosman, and Elliott (1995) found an orientation effect for scroll bar dragging speed. They found that people dragged scroll bars faster from top to bottom in the vertical direction than from left to right in the horizontal direction. They interpreted this finding as evidence that gravity helps to speed scrolling. In our study, we placed the most recent events at the bottom of the screen. In order for the participants to see additional information, they must scroll from the bottom to the top (rather than from the top to the bottom as mentioned in the Charness et al. study). This may slow scrolling behavior and further increase any performance gap due to scrolling between the different conditions.

References

Ahlstrom, U., & Friedman-Berg, F. (2005). *Subjective workload ratings and eye movement activity measures* (DOT/FAA/CT-05/32). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Ahlstrom, V., & Kudrick, B. (2005). *Human factors criteria for the design and acquisition of non-keyboard interaction devices* (DOT/FAA/CT-05/09). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Ahlstrom, V., & Kudrick, B. (2006). *Optimal design of event lists (ODELs) Phase 1: Does list format facilitate visual search for information* (DOT/FAA/TC-06/12). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Ahlstrom, V., & Longo, K. (Eds.). (2003). *The human factors design standard: Human interface guidelines* (HF-STD-001*).* Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Charness, N., Bosman, E. A., & Elliott, R. G. (1995, August). Senior-friendly input devices: Is the pen mightier than the mouse? Abstract presented at the *103rd Annual Convention of the American Psychological Association,* New York.

Chipman, L. E., Bederson, B. B., & Golbeck, J. (2004). Slide bar: Analysis of a linear input device. *Behavior and Information Technology*, *23*(1), 1-9.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson’s method. *Tutorial in Quantitative Methods for Psychology, 1*(1), 42-45.

Hendrickson, J. J. (1989). Performance, preference, and visual scan patterns on a menu-based system: Implications for interface design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Wings for the mind* (pp. 217-222). New York: ACM Press.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in-within subject designs. *Psychonomic Bulletin & Review*, *1*, 476-490.

Sibert, L. E., & Jacob, R. J. K. (2000). Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 281-288). The Hague, The Netherlands. New York: ACM Press.

Ware, C., & Mikaelian, H. H. (1986). An evaluation of an eye tracker as a device for computer input. In *Proceedings of the SIGCH/GI Conference on Human Factors in Computing Systems and Graphics Interface* (pp. 183-188). Toronto, Canada. New York: ACM Press.

Willems, B., Heiney, M., & Sollenberger, R. (2005). *Study of an ATC baseline for the evaluation of team configurations: Effects of allocating multisector control functions to a radar associate or airspace coordinator position* (DOT/FAA/CT-05/07). Atlantic City International Airport, NJ: Federal Aviation Administration, William J. Hughes Technical Center.

Zelinsky, G., & Sheinberg, D. (1995). Why some search tasks take longer than others: Using eye movements to redefine reaction times. In J. Findlay, R. Kentridge, & R. Walker (Eds.), *Eye movement research: Mechanisms, processes, and applications* (pp. 325-336). Amsterdam: Elsevier.

Zhai, S., Smith, B. A., & Selker, T. (1997, July). Improving browsing performance: A study of four input devices for scrolling and pointing tasks. *Proceedings of INTERACT97: The Sixth IFIP Conference on Human-Computer Interaction,* Sydney, Australia, 286-292.

Acronyms

FAA	Federal Aviation Administration
NAS	National Airspace System
ODELs	Optimal Design of Event Lists
TO	Technical Operations
WJHTC	William J. Hughes Technical Center