COLLOCATION AND PATTERN RECOGNITION EFFECTS ON SYSTEM FAILURE REMEDIATION

Anna Trujillo
NASA Langley Research Center
Hampton, VA

Hayes Press
Lockheed Martin
Hampton, VA

Previous research found that operators prefer to have status, alerts, and controls located on the same screen. Unfortunately, that research was done with displays that were not designed specifically for collocation. In this experiment, twelve subjects evaluated two displays specifically designed for collocating system information against a baseline that consisted of dial status displays, a separate alert area, and a controls panel. These displays differed in the amount of collocation, pattern matching, and parameter movement compared to display size. During the data runs, subjects kept a randomly moving target centered on a display using a left-handed joystick and they scanned system displays to find a problem in order to correct it using the provided checklist. Results indicate that large parameter movement aided detection and then pattern recognition is needed for diagnosis but the collocated displays centralized all the information subjects needed, which reduced workload. Therefore, the collocated display with large parameter movement may be an acceptable display after familiarization because of the possible pattern recognition developed with training and its use.

Introduction

Currently, most of the displays in control rooms can be categorized as status screens, alert and procedures screens (or paper), or control screens (where the state of a component is changed by the operator). With the advent and use of graphical displays, various types of input devices and the associated computing power available to compute and display information, it is now possible to combine these different elements of information and control onto a single display. The primary focus of this line of research is whether these pieces of information should be collocated.

Previous research found that operators like to have status, alerts and procedures, and controls located on the same screen or have status and alerts and procedures on one display with controls on another [1-3]. This research was done with displays that were not specifically designed for collocation. This follow-on experiment tested two displays specifically designed for collocation.

One of the collocated displays focused on collocation with little to no pattern recognition available but with large parameter movement. The other collocated display incorporated some pattern recognition but with less parameter movement. The baseline display incorporated features to permit pattern recognition with large parameter movement but without collocation. The amount of parameter movement available referred to how much the parameter could travel with respect to the size of the display area it had available to move around in. The baseline display had a normalized area of movement of 1, the collocated displays had a normalized area of parameter movement of 0.4 and an average of 2.0.

The collocated displays are thought to be of benefit because they would incorporate all pertinent information onto a single display and previous research has indicated subjects prefer this [4], especially for relatively stable displays where some type of collocation may be beneficial because of belonging to the same object [5].

On the other hand, collocation may become a detriment for diagnosis because attention must be paid to each cue [6]; although this may be due to smaller parameter movement typically available on collocated displays. During non-normal situations, more human processing must be done to recognize a change because an operator's expectation no longer matches what he sees [7]. This would be hindered with smaller parameter movement.

The pattern matching aspects of the displays would aid in the detection of non-normal situations [6, 8]. Pattern matching would become especially important in systems that are typically stable or in displays with smaller parameter movement.

Therefore, this experiment tested two collocated displays against a typical baseline configuration for detecting, diagnosing, and correcting for system failures. The displays also incorporated different levels of parameter movement and pattern matching in order to see if other factors besides collocation affected performance in handling non-normal situations.

Objectives

This experiment was conducted to determine whether collocation aided in detecting, diagnosing, and mitigating a system failure. In order to fully meet the objectives, three independent variables were controlled. These independent variables were (1) display format, (2) pilot status, and (3) display order.

For display format, each subject saw the baseline display and one of the collocated displays. The collocated displays were the Dial-on-Control (DoC) and MultiDimensional Object (MDO) display formats.

Both pilots and non-pilots participated in the experiment because the process control task was not specific to aviation. Non-pilots participated for a couple of reasons. First, non-pilots may not be as biased or familiar with the standard display format that is often used in aviation displays. Second, the collocated displays needed to be easy to understand by both pilots and non-pilots because of compatibility issues with current pilots and the possible use of these displays in other industries with control rooms.

Display order was controlled in order to satisfy the objective that the display formats were to be quickly and easily understood. Display order referred to the order subjects saw the baseline display and the collocated display.

Display Format

Each subject saw two display formats: (1) standard status displays and controls (baseline) and (2) one of the collocated displays – DoC or MDO. All the display formats modeled the same 3 systems – fuel system, power plant, and heat exchanger.

The power plant modeled encompassed a reservoir (RES) that supplied a pump (PMP) which fed an engine (ENG) (Fig. 1). The fuel system consisted of a tank (TNK) that fed two pumps (EDP and ADP) whose combined output was shown with overall system parameters (SYS) (Fig. 2(a)). The heat exchanger consisted of a reservoir (RES) that fed two pumps (PRIM PMP and AUX PMP) whose combined output was shown with system parameters (SYS) (Fig. 2(b)). There were two parameters associated with each component with corresponding alert levels. For this study, warning alerts were red in color, cautions were amber, and advisories were cyan. Normal values, which were the remaining range, were shown in green.

The collocated displays were designed so that all three types of information were located on one screen [1-3, 9]. Both were pictorial in format, which suggested less processing would be required [10] especially if patterns could be learned and discerned [11].

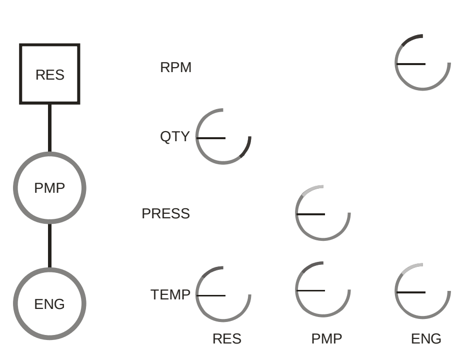

Figure 1. Baseline Engine Display.

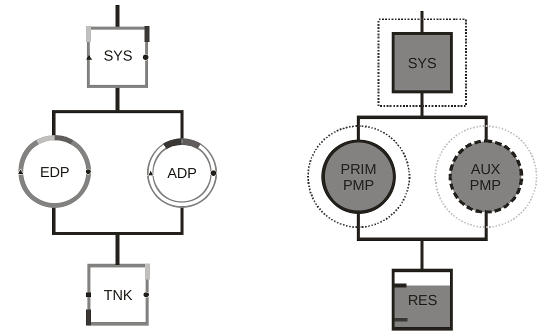

Figure 2. Collocated Displays: (a) Fuel System DoC and (b) Heat Exchanger MDO.

Furthermore, the information in a single location could enhance emergent features [12] and increase the likelihood of noticing a non-normal situation developing [5].

Baseline Display The baseline display format kept the status information separate from the control screen. Status information was presented with standard dial formats (Fig. 1) whose normalized area of parameter movement is 1. When all the parameters were at their expected values, the dial pointers were horizontal. This aspect of the display encouraged check reading because pattern matching could be employed [8]; any parameter deviation had a dial pointer departing from horizontal, which entailed large parameter movement.

The control screen mimicked the functional layout of the generic system (Fig. 1). Components that had no change of state, such as the RES, were shown with white squares. Components that could change state (*i.e.*, turn on and off), such as the PMP, were shown with circles. A single outlined circle indicated a component that was on while a double circle denoted a component that was off. The outline color of the component announced the highest alert range the component's parameters were in. A failed component was shown with a red outline and a red X.

Dial-on-Control (DoC) Display The dial-on-control format was a collocated display with the parameter information integrated into the control display (Fig. 2(a)). It had a normalized area of parameter movement of 0.4. This display shared some of the conventions employed in the baseline display. Components with no change of state were square while components that could change state were circles. Also, a single outlined circle indicated a component that was on where a double circle designated a component that was off.

Each component was split in half vertically. The left half of the component registered either pressure or quantity while the right half of the component indicated temperature. Pressure was shown with a triangle icon, quantity with a rectangle, and temperature with a circle. The icons traveled around the component outline. When all the parameters were at their expected values, the icons would be at the horizontal middle of the component outline. Therefore, this display incorporated collocation and pattern matching but with limited movement.

The appropriately color-coded alert range was indicated at either the top or bottom of the component outline. The rest of the outline, not including the alert ranges, of the component was green.

If a parameter reached an alert range, the icon changed from white to black and the component name was displayed in the same color as the component's parameters highest classification; otherwise the component's name was displayed in white. A failed component was displayed with a red X through the component and the component's name was displayed in red, indicating a warning.

Multi-Dimensional Object (MDO) Display As with the DoC display, the MDO display collocated the parameter information with the control display but the parameter information was more integrated pictorially into the control display (similar to [13]); therefore, this display supported collocation with no pattern matching because subjects were unfamiliar with this display (Fig. 2(b)) but it did have large parameter movement, a normalized area of parameter movement of 2.0. The additional incorporation of the parameter information was thought to enhance visual processing of the display in a glance such as was found with polar-star displays [9]. As with the other two displays, components with no change of state were square while components that could change state were circles. For the components with a change of state (*i.e.*, turn on and off), a solid white outline indicated a component that was on while a thick, long-dash white outline indicated a component that was off. A failed component had a red X through it.

Pressure was indicated by size. If pressure increased, the amount of component fill grew proportionally. If pressure decreased, the colored fill shrank proportionally. The beginning of a pressure alert range was shown with a dotted colored outline indicating the alert level. If the pressure alert range was reached, the dotted colored outline turned to a solid red, amber, or cyan indicating the alert level and the component name turned black in color.

Temperature was indicated by fill color. If the temperature increased, the fill color changed from green to the alert range color from the center out. If the temperature decreased, the fill color changed from green to the alert range color from the outside in. The beginning of the high temperature alert range was indicated by the outside edge of the colored component fill and the beginning of a low temperature alert range was the center of the colored component fill. If a high temperature alert were reached, the fill color was displayed in the same as the alert range color with a dotted green outline at the edge. If a low temperature alert range were reached, the fill color was displayed in the same as the alert range color with a small black circle in the middle. Also, the component name was displayed in black.

Quantity was indicated by fill level. If the quantity increased, the fill level oise and if the quantity decreased, the fill level fell. A small white horizontal line on the side of the component outline indicated normal fill level. A small alert-color-coded line on the side of the component outline showed the beginning of an alert range. When an alert range was reached, the component name turned black and the top of the fill level changed to the color coded alert range.

Pilot Status

An evaluation between pilots and non-pilots was desired because pilots may have had more experience with the baseline display configuration and non-pilots may be less biased towards the collocated displays. Therefore, half of the subjects were current certificated pilots with at least a class III medical certificate [14]. The rest of the subjects were non-pilots who were familiar with computers but did not play flight-simulation computer games.

Experiment Design

Subjects

Twelve people participated as subjects. Six were certificated pilots with a Class III medical certificate. They were also qualified to fly in an aircraft with either electronic displays or an electronic alerting system. The rest of the subjects were non-pilots who were comfortable working with a computer but reported that they did not play flight-simulation computer games. The average age of the pilots was 47 years and the average age of the non-pilots was 42 years. The pilots had an average of 16 years experience and over 780 hrs of flight experience.

Dependent Variables

The dependent variables consisted of the subjects' ability to track the randomly moving object, to detect, diagnose and mitigate the system failure, and their recollection of the problem.

The tracking task average magnitude from center of the target had a normalized range of 0 (centered) to $\sqrt{2 \text{ x } 50^2}$ (in one of the corners) (20" diagonal screen size). The tracking task absolute angle of the target was 0 straight up and was measured in radians.

The elapsed time it took subjects to detect, diagnose, and mitigate the system failure was recorded. Also measured was their accuracy in indicating the system with the failure and their diagnosis accuracy of the component parameter affected.

At the end of each display format, subjects completed the NASA-TLX workload measure questionnaire [15, 16] and a Cooper-Harper (CH) controllability scale rating [17, 18].

Procedure

When subjects first came in, they signed a consent form and then were given a verbal briefing on the experiment tasks. After this briefing, subjects moved to the simulator where they completed two practice runs, which behaved the same as the data runs, with the first display format, either baseline or one of the collocated displays. After the practice runs, subjects completed 12 data runs. During each run, subjects had to keep a randomly moving target centered using a left-handed joystick. They also had to monitor for a single failure that would occur in one of the systems. Once they identified the system with the failure, subjects then corrected the failure by following a checklist. At the end of each run, subjects had to answer questions about where the failure occurred (system, component, and parameter), and complete the NASA-TLX and CH controllability rating scale. At the end of the 12 data runs with the first display, subjects completed two practice runs with the second display and then the 12 data runs with that display. At the end of the simulation runs and questions, subjects completed a final questionnaire.

Data Analysis

Data was analyzed using SPSS® for Windows v13.0 [19]. The data was analyzed using a 1-way ANOVA. In all cases, significance was set at $p \leq 0.05$.

Results

Tracking Task

Pilot status and display type were significant for the average magnitude of the randomly moving target from the center of the display ($F_{(1, 17)}$=11.81, $p \leq 0.01$; $F_{(2, 17)}$=11.28, $p \leq 0.01$ respectively) and for the average absolute angle of the target ($F_{(1, 17)}$=6.93, $p \leq 0.01$; $F_{(2, 17)}$=4.48, $p \leq 0.02$ respectively) (Table 1). In both cases, non-pilots had the smaller deviation. For the average magnitude error, the MDO display had the greatest error but for the average absolute angle error, the DoC display had the greatest error. For the CH rating, display format was significant ($F_{(2, 35)}$=8.769, $p \leq 0.01$) with the DoC display having the lowest rating which subjects said allowed for more controllability.

Table 1. Tracking Task (Averages).

Independent Variable	Magnitude (max≅707)	Absolute Angle (rad)	CH Rating
Pilot Status			
Non-Pilot	55	0.32	
Pilot	65	0.42	
Display			
Baseline	55	0.37	2.56
DoC	58	0.46	2.61
MDO	71	0.28	3.06

Detection and Diagnosis

Pilot status, display format, and pilot status by display format were significant for accurately detecting the system with the problem ($F_{(1, 17)}$=10.46, $p \leq 0.01$; $F_{(2, 17)}$=6.82, $p \leq 0.01$; $F_{(2, 17)}$=12.22, $p \leq 0.01$ respectively) (Table 2) while display was significant for the time it took subjects to notice the problem ($F_{(2, 17)}$=7.95, $p \leq 0.01$) (Table 3). Interestingly, non-pilots were more accurate in determining the system with the problem and the MDO display had the least accuracy for determining the system with the problem, especially for pilots (Fig. 3). The time it took to accurately notice the problem was greatest for the DoC display.

Overall diagnostic accuracy had pilot status, display format, and pilot status by display format as significant factors ($F_{(1, 17)}$=14.94, $p \leq 0.01$; $F_{(2, 17)}$=7.94, $p \leq 0.01$; $F_{(2, 17)}$=12.90, $p \leq 0.01$ respectively). Again, in general, non-pilots were more accurate in diagnostic accuracy and the MDO display had the worst diagnostic accuracy (Table2). This particularly showed up for pilots (Fig. 3). The time to make this diagnosis had display format as a significant factor ($F_{(2, 17)}$=3.58,

Table 2. Detection and Diagnosis Accuracy.

Independent Variable	Detect Problem Accuracy (0=incorrect, 1=correct)	Overall Accuracy (0=incorrect to 2= all correct)
Pilot Status		
Non-Pilot	0.96	1.74
Pilot	0.84	1.43
Display		
Baseline	0.96	1.66
DoC	0.94	1.81
MDO	0.78	1.33

Table 3. Time to Detection and Diagnose (sec).

Independent Variable	Notice Problem	Overall Problem
Display		
Baseline	78	12
DoC	95	11
MDO	76	17

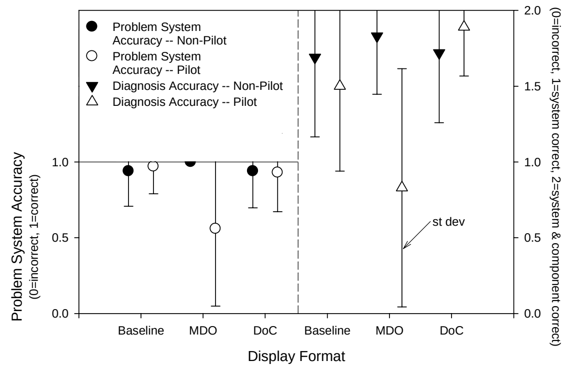

Figure 3. Problem System Accuracy and Diagnosis Accuracy by Display Format and Pilot Status.

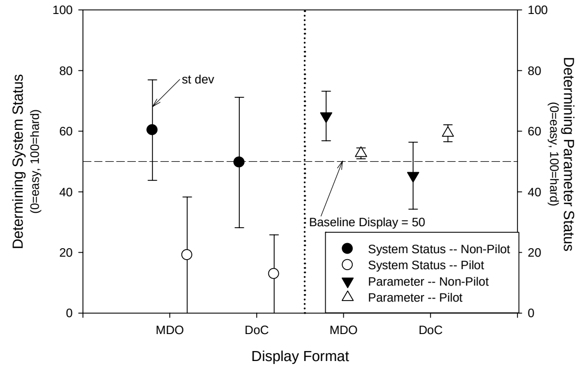

Figure 4. Difficulty in Determining System and Parameter Status by Display Format and Pilot Status.

p≤0.04) where the MDO display time exhibited the highest detect and diagnoses times (Table 3).

Workload

The NASA-TLX workload rating was not significant but subjects were also asked about the workload for determining the status of the system and its components in the final questionnaire. For this workload, display format was significant in determining the affected component ($F_{(2, 5)}$=18.76, p≤0.01) where the two collocated displays were rated as having lower workload (Table 4).

Table 4. Final Questionnaire Workload for Determining Component with Failure.

Independent Variable	Questionnaire: Determine Component (0=low, 100=high)
Display	
Baseline	50
DoC	38
MDO	35

Subjective Input

In the final questionnaire, subjects gave opinions on how easy or hard it was to use the collocated displays. For determining system status, pilot status and pilot status by display format were significant ($F_{(1, 5)}$=5.93, p≤0.03; $F_{(2, 5)}$=4.25, p≤0.03 respectively) (Table 5, Fig. 4). Pilots reported that the MDO display was the hardest to use for determining system status. Also asked was how easy or hard it was to determine the system parameter affected. For this, pilot status by display format was significant ($F_{(2, 5)}$=4.34, p≤0.03). Here, pilots rated the DoC display as harder but non-pilots rated the MDO display as harder (Fig. 4).

Table 5. Final Questionnaire Results for Determining System Status (0=easy, 100=hard).

Independent Variable	Rating
Pilot Status	
Non-Pilot	52
Pilot	43

Interestingly, subjects reported no difference among the displays for intuitiveness, usefulness, and overall preference. There was also no significant difference when looking at the pilot status by display format interaction but some interesting patterns did show up (Fig. 5). For intuitiveness, pilots did not think the DoC display was intuitive but non-pilots did think it was intuitive. For usefulness, non-pilots did not find the MDO display very useful. Lastly, for overall preference, pilots preferred the MDO display while non-pilots preferred the DoC display.

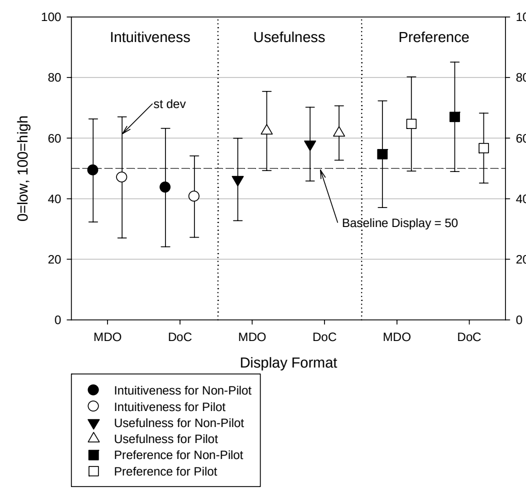

Figure 5. Intuitiveness, Usefulness, and Preference Subjective Data (not significant for display by pilot interaction)

Discussion and Conclusions

Twelve subjects participated in this experiment evaluating two displays designed for collocating system information against the baseline that consisted of dial status displays, a separate alert area, and a controls panel. During the data runs, subjects had to keep a randomly moving target centered on a display using a left-handed joystick and locate a system problem in order to correct it using the provided checklist.

In general non-pilots performed slightly better than pilots in the tracking task. This may have been due to a couple of reasons. Non-pilots may have had a smaller error on the tracking task because they were concerned about losing control and their ability to regain it. Many non-pilots reported that they focused on the tracking task. Also, pilots reported being more comfortable with not being in the exact center for the tracking task because they realize they must also attend to the monitoring of other information and a slight deviation from path within certain bounds, such as the FAA private pilot practical test standards [20], is acceptable.

While diagnosing problems, subjects did best when the displays had pattern recognition (baseline and DoC). This is not surprising because of the detection and diagnosis task, especially for pilots who work with dial-type displays on a regular basis. Interestingly, the time to notice the failure, the detection aspect, was lowest with the baseline and MDO displays across subjects. This suggests that the large parameter movement is the most helpful for detecting parameter deviations. Then, the workload for the two collocated displays was lower than for the baseline display.

In general, the two collocated displays did not perform any worse than the traditional baseline display setup. In fact, the results showed that for detecting a problem, large parameter movement is best. Then for diagnosing the problem, pattern recognition dominates. But the workload is the least for the collocated display, which centralized all the information the subjects needed to detect, diagnose, and correct for a system failure. Therefore, the MDO display may be an acceptable display after familiarization because of its large parameter movements, collocation, and possible pattern recognition developed with training and its use.

References

[1] A. P. Bartolone and A. C. Trujillo, "Glass-Cockpit Pilot Subjective Ratings of Predictive Information, Collocation, and Mission Status Graphics: An Analysis and Summary of the Future Focus of Flight Deck Research Survey," NASA Langley Research Center, Hampton, NASA-TM 2002-211419, January 2002.

[2] A. C. Trujillo, "Experience and Grouping Affects When Handling Non-Normal Situations," presented at Human Factors and Ergonomics Society 45th Annual Meeting, 2001.

[3] A. C. Trujillo, "Response Times in Correcting Non-Normal System Events When Collocating Status, Alerts and Procedures, and Controls," presented at 2nd International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres, Manchester, England, 2001.

[4] A. F. Stokes and C. D. Wickens, "Aviation Displays," in Human Factors in Aviation, E. L. Wiener and D. C. Nagel, Eds. San Diego, CA: Academic Press, Inc., pp. 387-431, 1988.

[5] G. Davis, "Characteristics of Attention and Visual Short-Term Memory: Implications for Visual Interface Design," Philosophical Transactions: Mathematical, Physical and Engineering Sciences, vol. DOI: 10.1098/rsta.2004.1462, pp. 18, 2004.

[6] P. M. Jones, C. D. Wickens, and S. J. Deutsch, "The Display of Multivariate Information: An Experimental Study of an Information Integration Task," Human Performance, vol. 3, pp. 1-17, 1990.

[7] J. Hawkins and S. Blakeslee, On Intelligence, 1st ed. New York, NY: Times Books, 2004.

[8] M. S. Sanders and E. J. McCormick, Human Factors in Engineering and Design. USA: McGraw-Hill Publishing Co., 1987.

[9] D. L. Mahaffey, R. L. Horst, and R. C. Munson, "Behavioral Comparison of the Efficacy of Bar Graphs and Polar Graphics for Displays of System Status," presented at 1986 IEEE International Conference on Systems, Man and Cybernetics, 1986.

[10] L. F. Weinstein and C. D. Wickens, "Use of Nontraditional Flight Displays for the Reduction of Central Visual Overload in the Cockpit," The International Journal of Aviation Psychology, vol. 2, pp. 121-142, 1992.

[11] A. F. Stokes, C. D. Wickens, and K. Kite, Dispay Technology: Human Factors Concepts. Warrendale, PA: Society of Automotive Engineers, Inc., 1990.

[12] M. A. Buttigieg and P. M. Sanderson, "Emergent Features in Visual Display Design for Two Types of Failure Detection Tasks," Human Factors, vol. 33, pp. 631-651, 1991.

[13] R. Albert, N. Syroid, Y. Zhang, J. Agutter, F. Drews, D. Strayer, G. Hutchinson, and D. Westenskow, "Psychophysical Scaling of a Cardiovascular Information Display," presented at IEEE Visualization 2003, Seattle, Washington, 2003.

[14] Federal Aviation Administration, "FAA Code of Federal Regulations, Title 14, Vol. 1, Chapt. 1, Part 61, Subpart D - Third-Class Airman Medical Certificate," 2006.

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

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

[17] G. E. Cooper and R. P. Harper, "The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities," AGARD April 1969.

[18] R. P. Harper and G. E. Cooper, "Handling Qualities and Pilot Evaluation," in Wright Brothers Lectureship in Aeronautics, 1984.

[19] SPSS Inc., "SPSS for Windows, v13.0," 13.0 ed. Chicago, IL: SPSS, Inc., 2004.

[20] Federal Aviation Administration, "Private Pilot Practical Test Standards for Airplane (SEL, MEL, SES, MES)," Flight Standards Service, pp. 115, 2002.