Effects of Expected-Value Information and Display Format on Recognition of Aircraft Subsystem Abnormalities

Michael T. Palmer and Kathy H. Abbott

Effects of Expected-Value Information and Display Format on Recognition of Aircraft Subsystem Abnormalities

Michael T. Palmer and Kathy H. Abbott
Langley Research Center • Hampton, Virginia

National Aeronautics and
Space Administration
Code JTT
Washington, D.C.
20546-0001

Contents

Symbols and Abbreviations

The abbreviations in parentheses are used in selected figures and appendixes.

F		ratio of mean square of effect to mean square of associated error term
p		probability value
EPR	(Epr)	engine pressure ratio
N_1	(N1)	fan rotational ratio, r/min
EGT	(Egt)	exhaust gas temperature, °C
N_2	(N2)	compressor rotational speed, r/min
FF		fuel flow, lb/hr
VIB	(Vib)	vibration
PRS	(Pres)	pressure, psi
QL	(QuanL)	fuel quantity left, lb (measured in increments of 1000)
QC	(QuanC)	fuel quantity center, lb (measured in increments of 1000)
QR	(QuanR)	fuel quantity right, lb (measured in increments of 1000)

Summary

Advances in computer technology have increased the capability of system designers to generate and present information to aid flight crews of commercial transport aircraft. Aiding flight crews in managing in-flight subsystem failures is one area that has a high potential benefit, but interface and display technology currently may be inadequate to achieve this benefit. The purpose of this study, therefore, was to identify improved methods for presenting system parameter information for the detection of an abnormal condition and for the identification of the system status. Specifically, two workstation experiments using static displays were conducted concurrently. The primary objective of the first experiment, the Expected Value Experiment, was to determine whether including expected-value-range information on traditional parameter display formats had any significant effect on the ability of a subject to detect abnormal conditions and identify the status of the affected systems. The primary objective of the second experiment, the Parameter Format Experiment, was to determine if using a nontraditional parameter display format, which presented relative deviation from the expected value, was significantly better than using a traditional parameter display format with expected-value ranges included for the same task. The inclusion of expected-value-range information onto traditional parameter display formats had essentially no effect on the subjects' performance for the given task in this study. However, the subjective evaluation results indicate that the subjects show moderate support for including this information. The nontraditional column deviation parameter display format (hereafter called the column deviation format) resulted in significantly fewer errors when compared with traditional parameter display formats with the expected-value-range information included. In addition, error rates for the column deviation format remained relatively stable as the scenario complexity increased, whereas error rates for the traditional parameter display formats with expected-value ranges increased dramatically. The subjective evaluation results also indicate that the subjects thought that their own performance was better with this column deviation format and that they generally preferred it over the others. It is therefore recommended that the column deviation format be considered for display applications that require rapid recognition of out-of-tolerance conditions, especially if a large number of system parameters must be monitored.

Introduction

Background

Advances in computer technology have increased the capability of system designers to generate and present information to aid flight crews of commercial transport aircraft. Aiding flight crews in managing in-flight subsystem failures is one area that has a high potential benefit, and research to develop systems for aiding failure management is under way for aircraft (ref. 1) and for other complex human–machine interface applications, such as nuclear power plant control rooms and computerized hospital operating rooms (ref. 2). Appropriate presentation of the information generated by these systems, however, is very important. Much of this new information may have a higher degree of uncertainty than that which pilots are currently using because it includes estimates of the system state in addition to the raw sensor data. The uncertainty of this information may make current interface and display technology inadequate. This study attempted to resolve one particular display issue concerning the presentation of system parameter information for the recognition and the identification of abnormal system behavior.

During the development of the Engine Indication and Crew Alerting System (EICAS) displays at The Boeing Company during the early 1980's, Boeing engineers found that the conventional round electromechanical gauges are actually very good. However, Graham (ref. 3) noted that the commercial aircraft industry seems to have forgotten "exactly what their virtues are and how their features are used, individually and collectively, by the aircrew." Graham further states that the translation of these displays into electronic form is a "deceptively difficult task," and he correctly cautions that although we attempt to design electronic displays so that they are better than the electromechanical gauges that they replace, we must first make sure that the new displays are at least as good as the old ones.

A point that Graham did not address in that paper, however, is that the conventional gauges are only remarkably good displays on average; that is, they do a good overall job for the wide variety of tasks for which they are used. Specific tasks may benefit greatly from displaying system parameter information in a different form. For example, research has recently been performed by Abbott (ref. 4) which examined the benefits of presenting engine parameter information in a new format for recognition of

abnormal engine performance. This new column deviation format displayed the difference between the current and the expected value for each parameter. Abbott demonstrated that the new format significantly improved recognition of abnormal engine conditions, especially when the parameter values were still within their operating limits. Such was the case with the Air Florida Boeing B-737 accident near Washington, District of Columbia, in 1982 (ref. 5), in which ice impaired the performance of the engine pressure ratio (EPR) sensors; this situation misled the crew into believing that the engines were developing full thrust when they were not. The EPR gauges incorrectly indicated normal operation; however, fuel flow and exhaust gas temperature (EGT) indications were significantly lower than those expected for the existing conditions.

One important purpose of aiding the flight crew during fault recognition and system status identification is to increase their situation awareness. Failure to understand all the implications of certain system failures on the capability of other aircraft systems has been cited as a contributing factor in several accident and incident cases. These cases include the American Airlines, Incorporated, McDonnell Douglas Corporation DC-10 accident at O'Hare airport in Chicago, Illinois, in 1979 (ref. 6), in which an engine separation during takeoff resulted in an unrecognized asymmetric leading-edge slat condition, and the China Airlines Boeing B-747 incident enroute to Los Angeles, California, in 1985 (ref. 7), in which the autopilot's compensation for loss of thrust on an engine masked the asymmetric thrust condition until the autopilot was unable to handle it. Another example of the potentially devastating effects of misunderstanding system failures and their effects was the British Midland Airways Boeing B-737-400 accident near Kegworth, Leicestershire, England, in 1989 (ref. 8); in this accident, the flight crew mistakenly shut down the wrong engine after misinterpreting the symptoms resulting from a fan blade separation on the left engine. In this case, the flight data recorder proved that the instruments were presenting the information necessary to correctly identify the affected engine. The displays did not, however, lead the crew to take the correct action. Because the affected engine was still producing thrust and the secondary feedback from the air-conditioning system suggested that the correct action had indeed been taken, the crew's error was not noticed until it was too late to restart the good engine before impact with the ground. Thus, enhancing the crew members' awareness of the true state of their aircraft should help them not only with the current situation but also with any future re-

lated or unrelated problems that may arise during that flight.

The purpose of this study, therefore, was to identify improved methods for presenting system parameter information for the detection of an abnormal condition and for the identification of the system status. Specifically, two workstation experiments using static displays were conducted concurrently to identify particular characteristics that make one format preferable to another for displaying system parameter information. The primary objective of the first experiment, the Expected Value Experiment, was to determine whether including expected-value-range information on traditional parameter display formats had any significant effect on the ability of a subject to detect abnormal conditions and identify the status of the affected systems. The primary objective of the second experiment, the Parameter Format Experiment, was to determine if using a nontraditional parameter display format, which presented relative deviation from expected value, had any significant effect on the ability of a subject to detect abnormal conditions and identify the status of the affected systems compared with traditional parameter display formats with expected-value ranges included. Although these two experiments were conducted concurrently, they will be treated separately for the purposes of describing the experimental method and results.

Experimental Variables

In the Expected Value Experiment, the experimental variables examined were the presence or the absence of an expected-value range on a traditional parameter display format, the type of traditional format used, the level of application of a "display-by-exception" philosophy, and the level of scenario complexity for each system condition viewed. In the Parameter Format Experiment, the experimental variables examined were the type of parameter display format, the level of application of a display-by-exception philosophy, and the level of scenario complexity for each system condition viewed. Each of these experimental variables is explained below. Because the exact form of the expected-value-range information in the first experiment depends on the traditional parameter display format that it is added to, these two variables are explained and discussed together.

Expected-Value Range and Parameter Display Format

As stated above, the primary objective of the Expected Value Experiment was to determine whether

including expected-value-range information on traditional parameter display formats had any significant effect on the ability of a subject to detect abnormal conditions and identify the status of the affected systems. Here, expected-value-range information refers to the range of values in which a parameter would normally be expected to be, given the current operating conditions and system state. This range is normally much smaller than that for all normal conditions and system states. Traditional parameter display formats are those visual representations of a parameter's value which have evolved from electromechanical dials, gauges, and pointers used in some current (and all previous) generations of commercial aircraft. Sketches of two of these traditional parameter display formats are shown in figures 1 and 2. The circular gauges (fig. 1) are similar to those used on the Boeing B-757-200 aircraft (ref. 9) for primary engine parameters, and the vertical-scale sliding-pointer indicators (fig. 2) are similar to those used on the Boeing B-757-200 aircraft for some secondary parameters. To maintain consistency with the terminology used by the subjects, the vertical-scale sliding-pointer indicators were referred to throughout this report as "bow tie bargraphs" (although they are not strictly bargraphs).

Traditional parameter display formats do not normally include any indication of what the ideal value of the associated parameter should be for the current conditions. Although fixed caution and warning limits are often shown, these are not changed dynamically as the environmental conditions or the state of the underlying system changes. For example, the ranges of normal values for a turbofan engine's oil pressure and temperature increase considerably when the engine thrust is increased from idle to a takeoff power setting. The traditional caution and warning limits for these parameters, however, do not account for such changes, and they give no indication to the flight crew that these parameters are expected to be within a higher range, given the current thrust setting.

To investigate the potential advantages of including expected-value information, the two traditional parameter display formats introduced above, circular gauges and bow tie bargraphs, were each enhanced by adding the expected-value range. This range was presented as a white arc or line that extended 10 percent above and below the expected (modeled) value of the parameter, given the current condition. Examples are shown in figures 3 and 4, respectively, in which expected-value-range information has been included on the circular gauge format (as an arc segment) and on the bow tie bargraph format (as a line segment).

Both of these traditional parameter display formats were included in the experiment, since the specific type of format might affect how the subjects use the expected-value information.

Although the parameter formats used were similar to those on current commercial aircraft, there was one major difference in their operation. For this first experiment, we wanted to provide the same visual color cues whether the expected-value-range information was included or not. Therefore, since exceeding the expected-value range caused the color of the current parameter value and its surrounding box to turn either amber or red, this color change was provided regardless of whether the expected-value range itself was shown or not. Although providing these color cues clearly supplies a visual indication that does not currently exist for traditional parameter display formats, the color change was included so that any differences in subject performance detected with respect to the expected-value-range information would not be confounded by differences in color cues.

Also as stated above, the primary objective of the second experiment, the Parameter Format Experiment, was to determine if using a nontraditional parameter display format that presented relative deviation from expected value had any significant effect on the ability of a subject to detect abnormal conditions and to identify the status of the affected systems compared with traditional parameter display formats with expected-value ranges included. The nontraditional parameter display format used in the second experiment was similar to the Engine Monitoring and Control System (E-MACS) concept developed recently at Langley Research Center (ref. 4) as part of a task-oriented display design effort. This display is referred to as a column deviation format, and it is shown with labels in figure 5 for a single display element.

The operational concept of the column deviation format is considerably different from that of the more traditional formats. The horizontal centerline of this display represents the expected value of the parameter as provided by a subsystem model, and it is shown in green. A solid column of color fill extends either up or down from this centerline, thus representing either higher or lower values than those expected. The horizontal lines immediately above and below the centerline are amber, and they represent a 10-percent deviation from the expected (modeled) value. When the difference between the actual and expected values of the parameter reaches this level, the column of color fill reaches one of these lines and also turns amber. The horizontal lines above and below the amber lines are red, and they represent a

15-percent deviation from the expected value. Again, when the difference reaches this level, the column of color fill reaches one of these lines and turns red. One addition that Abbott (ref. 4) made to the basic column deviation format is that fixed caution and warning limits are represented; this is accomplished by ensuring that the height of the column of color fill is extended when the actual value is near one of those limits. Therefore, even if no difference exists between the actual and expected values, if the actual value has reached a high caution limit, the column of color fill will reach the upper 10-percent deviation line and turn amber.

Because the expected value of each parameter is an inherent part of the column deviation format, the two traditional parameter formats used for comparison had the expected-value ranges (as described above) added to them. This addition kept the information content of the displays similar, even though the method of presenting the information was different. Specifically, an extra processing step was required to determine the relative deviation of a parameter from its expected value when using the traditional parameter display formats because the column deviation format presented this information directly.

Levels of Display by Exception

Based on Allen's previous work (ref. 9), it was concluded that varying the amount of information contained in each display by using different levels of a display-by-exception philosophy would probably be a factor in how well the subjects could perform the stated task. The underlying concept of display by exception is that only those items requiring immediate attention should be shown. Therefore, four different levels of display by exception were used; each represented a different level of application of the display-by-exception philosophy. For level 1, all parameters for all systems were always displayed, thus representing a total absence of display by exception. This approach is similar in concept to many current-generation cockpits with fixed electromechanical instruments, and it is demonstrated by the sample display shown in figure 6 (using circular gauges). For level 2, primary engine parameters were always displayed, but parameters for other systems were displayed only when those systems were affected by a problem. All parameters for each affected system were shown together as a group. This approach is similar in concept to the Boeing EICAS display philosophy, and it represents an application of the display-by-exception philosophy at the systems level (except for the engines). A sample display demonstrating this level is shown in figure 7, and it de-

picts the same system condition used in figure 6. For level 3, only parameters of the same type were displayed for affected systems when one or more of those parameters were out of tolerance. For example, all hydraulic pressure readings were displayed if even one was abnormal, but no hydraulic quantity and temperature readings were shown unless one or more of those, too, were abnormal. This approach represents an application of the display-by-exception philosophy at the parameter-type level, and it is demonstrated by the sample display in figure 8; this display, again, depicts the same system condition shown in figures 6 and 7. For level 4 (the last display exception level), only out-of-tolerance parameters were displayed for the affected systems. This approach represents an application of the display-by-exception philosophy at the level of each individual parameter reading. Figure 9 presents a sample display demonstrating the use of this level with the same system condition used for figures 6 through 8. A summary of the four different levels of the display-by-exception philosophy is presented in table I.

Scenario Complexity

It was also expected, based on Allen's work (ref. 10), that the complexity of the system state shown to the subjects would have a significant effect on their task performance, so seven different statically displayed scenarios with varying levels of complexity were used. The level of complexity for each scenario was manipulated by controlling the number of systems (one, two, or three) affected by a problem and the number of abnormal parameter readings (single or multiple) within each affected system. Because Allen found that increases in the number of affected systems resulted in more errors than increases in the number of abnormal parameters within each system, it was expected that the following scenario descriptions would result in an approximately monotonic increase in complexity level: (1) no abnormal parameters; (2) a single abnormal parameter in a single system; (3) multiple abnormal parameters in a single system; (4) a single abnormal parameter in each of two systems; (5) multiple abnormal parameters in each of two systems; (6) a single abnormal parameter in each of three systems; and (7) multiple abnormal parameters in each of three systems. Each abnormal parameter in the scenarios had a value that was either out of the normal operating range for that parameter (i.e., in a caution or warning range) or was different from the expected (modeled) value by more than 10 percent. Both conditions resulted in the same color change. A summary of the seven

different levels of scenario complexity is presented in table II.

To make the recognition of parameter conditions as realistic as possible, the scenarios used to implement three of the complexity levels were derived from the National Transportation Safety Board accident reports (refs. 11 and 12). In addition, normal engine values for all seven levels of scenario complexity were generated using a Boeing B-737-100 Pratt & Whitney JT8-D turbofan nonlinear engine model that produced thrust values sufficient for a Boeing B-737-100 to cruise at Mach 0.70 at 33 000 ft. Slight variations from the throttle angles required for trimmed flight conditions were used to create the engine parameter values for the different scenarios. For abnormal engine conditions, the operating ranges and limits for the engine parameters were first obtained from the *Boeing Model 737-100—Operations Manual* (ref. 13) for the Boeing B-737-100 that was equipped with Pratt & Whitney JT8-D turbofan engines. Values significantly lower and higher than normal were then selected as means for low and high values, and a corresponding range of variation for each mean was also selected.

The operating ranges and limits for the subsystem parameters were also obtained from the Boeing manual (ref. 13). Low, normal, and high means and corresponding ranges of variation then were selected. To obtain the exact parameter values used for this experiment for each scenario (except for the modeled normal engine conditions), low, normal, and high means were randomly varied within their ranges using the "minimal standard" pseudorandom number generator described by Park and Miller (ref. 14).

One difference between the screen displays tested in this study and the displays used in current cathode-ray-tube-equipped (CRT-equipped) aircraft is that all system parameters were displayed using the same format. For example, if engine parameters were displayed using circular gauges, then all oil, hydraulic, electric, and fuel system parameters were also displayed using circular gauges. This situation is not generally the case in practice, where space constraints force creative compromises. However, to maintain consistency and avoid introducing yet another experimental factor, different parameter formats were not mixed on the same display for this study. Another difference is that current CRT-equipped aircraft also use a caution and warning system, such as Boeing's EICAS, which is integrated with the parameter display and serves to direct the crew's attention to the appropriate parameters by displaying textual messages. This study only investigated the pilot's ability to use certain types of information on the parameter display itself.

Expected Value Experiment

The following sections describe the experimental method, the experimental hypotheses, the test procedures, the data analysis, and the results and discussion for the Expected Value Experiment. Because this experiment was conducted jointly with the Parameter Format Experiment, the relationship between these objective results and the subjective data are covered jointly with the results of that experiment in the section of this report entitled "Subjective Evaluation Results."

Experimental Method

The Expected Value Experiment examined the effects on subject task performance of adding expected-value-range information to traditional circular gauge and bow tie bargraph parameter display formats. Twelve test subjects were selected to participate in this study. All subjects held an Airline Transport Rating on their pilot licenses, and all were "type rated" in the Boeing B-737 aircraft. Nine test subjects were currently employed as flight crew members of a major commercial air carrier, and three test subjects were employed as engineering test pilots. In general terms, each subject's task was to view on a graphics workstation a static display that represented a particular combination of the independent variables, and then to clear the screen and verbally report which aircraft systems were affected by a problem and which parameters within those affected systems were abnormal.

The independent variables for the Expected Value Experiment included the presence of expected-value-range information (included or not included), the type of traditional parameter display format (circular gauges or bow tie bargraphs), the level of application of the display-by-exception philosophy (levels 1 through 4, as defined in table I), and the level of scenario complexity (levels 1 through 7, as defined in table II). The experiment was constructed as a repeated-measures, full factorial design, in which each of the 12 test subjects was tested once on each combination of the independent variables (i.e., there were no replications within experimental cells). This testing resulted in a total of 112 treatments per subject for the $2 \times 2 \times 4 \times 7$ design. All sequences of presentation of the independent variables were either counterbalanced or selected randomly to reduce the effects of learning and ordering bias on the results.

The dependent measures for this experiment were the visual, verbal, and total response times, the accuracy of the subjects' responses while identifying the affected systems and abnormal parameters, and the preference data from the subjective evaluation. Both visual and verbal response times were obtained to detect differences in strategy between pilot subjects because it was expected that some pilots would perform most of the recognition and identification tasks while viewing the display and others would perform these steps after clearing the display from the screen by relying on short-term memory. The total response time, which was an unweighted sum of the visual and verbal response times, was also computed to obtain an overall measure of the task speed. The accuracy of the subjects' verbal responses was obtained by counting the number of errors in system identification (i.e., missed/extra systems), the errors in parameter identification (i.e., missed/extra parameters), the errors in parameter status (i.e., caution/warning or high/low status), and the total number of errors. Each of these specific error types was measured to detect any differences in the types of mistakes that were made with the different combinations of treatment conditions. The total number of errors, which was an unweighted sum of the other error types, was also computed to obtain an overall measure of task accuracy. The subject preferences as expressed in the subjective evaluation were also examined, although they were not evaluated statistically. The data collected gave general indications of the subject's preferences for or against each level of the experimental variables, and they also provided feedback about the perceived ease of use. Standard workload evaluation techniques, such as the modified Cooper-Harper rating scale, were not used because of the limited nature of the task in this experiment.

Experimental Hypotheses

The experimental hypotheses, in terms of expected results from the analysis of the dependent measures data, are presented below for the Expected Value Experiment. Justifications for these expectations are provided where appropriate.

Expected-Value Range

It was expected that the inclusion of expected-value-range information on both of the traditional parameter display formats would lead to a decrease in visual response time and a decrease in the number of parameter identification and status errors, primarily because the new information should have made it easier to detect and remember the status of the abnormal parameters. It was also expected that the pilots would favor having this information included on the displays. The justification for these hypotheses was that expected-value ranges presented the parameter information in a manner more consistent with its intended use. Because the experimental task was to detect abnormal conditions and report the status of the affected parameters, the situation-specific reference provided by the expected-value range should have made that task easier.

Parameter Display Formats

Of the two traditional parameter display formats (circular gauges and bow tie bargraphs), the circular gauges most closely approximate the electro-mechanical round dials that most current pilots have used throughout their careers. It was believed to be unlikely that the brief training period in this experiment would have overcome the familiarity associated with years of experience. Therefore, it was expected that both visual and verbal response times would be slightly lower for the circular gauge parameter format, that fewer parameter identification and status errors in the response would be made, and that the pilots would subjectively favor it over the bow tie bargraph format.

Levels of Display by Exception

It was expected that the levels of display by exception which gave the least amount of information would lead to a decrease in both visual and verbal response times and a decrease in the number of system identification errors made. The justification for this hypothesis is that the levels with the least information displayed fewer parameters that were not indicating abnormal readings. Therefore, scanning the display for the affected systems and the parameters that were abnormal should have been significantly easier because the display was less cluttered. It was also expected that the response accuracy would decrease with level 4, since displaying only the abnormal parameters may remove the visual cues that help to distinguish whether those parameters belong to the left or the right system.

Scenario Complexity

As described earlier, the scenarios were designed to present a monotonically increasing amount of complexity; scenario 1 was the most simple and scenario 7 was the most complex. It was expected that greater scenario complexity would result in significantly higher visual and verbal response times, as well as significantly more errors in the response.

The issue of fault and status complexity is extremely important in the evaluation of display format concepts, especially since Allen (ref. 10) found that the type of error committed more frequently with multiple affected systems is that one or more of these systems would be forgotten or ignored. This omission indicates a potentially critical lack of situation awareness. In fact, this lack of situation awareness may cause the crew to respond in an inappropriate manner to the current and any future abnormal conditions. The range of affected systems and fault indications in the scenarios designed for this study was introduced to examine how the number of systems and fault indications contributes to complexity for the different display concepts examined. However, it has also been shown that diagnosis of critical inflight events is more difficult when the necessary system knowledge is not incorporated properly into the diagnostic process (ref. 15). This difficulty suggests that the "perceived" complexity may be a function of the information presentation, the number of affected systems and parameters, and the familiarity of the pilot with those systems and their interactions.

Although scenario complexity was included as an independent variable in this experiment for the above reasons, it is secondary to the main purpose of this experiment, and therefore the results and discussion relating to complexity are presented in appendix A.

Test Procedures

When the test subjects first arrived, they each filled out a pilot background questionnaire to record information such as years of experience, number of flight hours in different aircraft and simulator types, and participation in other research studies. All written test materials are included in appendix B. Each subject then was asked to read a pilot briefing on the current study that explained its goals, described the display formats and levels of display by exception to be tested (referred to as "presentation styles" in the test materials), and explained the actual testing procedure. The briefing also instructed each test subject to place priority on accuracy, rather than speed, during the conduct of the test.

Before the actual experiment trials began, the subjects participated in a brief training exercise on the experiment hardware to familiarize them with the different treatment conditions and the operation of the test program. The training exercise lasted approximately 10 minutes, and it presented a single scenario (not used for the actual experiment) which used each of the different expected-value-range conditions, parameter formats, and levels of display by exception. During this time, the subjects were again instructed not to sacrifice accuracy for speed, and they were trained on the desired verbal protocol for the response accuracy measure.

At this point, the test trials began. For each combination of levels of the experimental variables, the test program first indicated that it was ready to proceed. The subjects then pressed the space bar on the keyboard, which caused the display to appear on the terminal screen. When the subjects finished looking at the display, they again pressed the space bar, which cleared the screen. The test program automatically calculated the interval between these keyboard inputs and recorded it as the visual response time. The subjects then verbally identified the affected aircraft systems and the condition of each parameter for those systems, and then they pressed the space bar one last time. The test program calculated the interval between these keyboard inputs and recorded it as the verbal response time. The verbal responses of the subjects were manually recorded on a form by the experimenter, and each session was also tape recorded to preserve the comments of each subject and provide verification of the written test results.

After completing all test trials, which took approximately 3 hours, the subjects completed subjective evaluations to record their impressions about their own performance, the ease of use of each display, and the display preferences and the reasons for these preferences. The subjective evaluations also recorded suggestions from the subjects concerning how to improve the displays.

Data Analysis

Visual, verbal, and total response time data were analyzed using a mixed-model analysis of variance (ANOVA), with a significance level of 0.05. The differences between the means of the effects found to be significant in the ANOVA were assessed using the Waller-Duncan procedure (ref. 16). Both the mixed-model ANOVA and the Waller-Duncan tests were computed using the General Linear Models (PROC GLM) procedure in the SAS/STAT® statistical analysis computer program (ref. 17).

System identification, parameter identification, parameter status, and total error data were also analyzed using a mixed-model ANOVA followed by the Waller-Duncan tests, where appropriate. Although frequency data such as the error counts in this experiment are usually examined using a stratum-adjusted Pearson chi-square statistic instead of the ANOVA, the more conservative mixed-model ANOVA was

used. This substitution was made because the error data were so sparse that expected cell frequencies in the chi-square analysis were significantly smaller than needed (ref. 18). The effect of analyzing the error count data using ANOVA, even though the data were nonnormally distributed, was that the significance levels were probably slightly higher than those reported by the F-test in the ANOVA tables (ref. 19).

For both the response time and error data analyses, note that the significant main effects that are also included in the significant interactions with other effects normally will not be examined separately in the discussion below. Note also that unless specifically mentioned, inspection of the few statistically significant higher order interactions (three-way interactions and above) did not reveal any discernible patterns, and therefore they are not reported.

Results and Discussion

The results of the data analyses for the Expected Value Experiment are presented below. Results that bear directly on the primary purpose of the experiment are examined first; results that indicate additional findings are examined second. When the results of the ANOVA tests are presented, they consist of the F-test statistic (F, which is the ratio of the mean square of the effect to the mean square of the associated error term), followed by the probability value p.

Primary Results

The primary objective of the first experiment, as stated in the Introduction, was to determine if including expected-value-range information on traditional parameter display formats had any significant effect on the ability of a subject to detect abnormal conditions and identify the status of the affected systems. No statistically significant main effects or interactions involving expected-value-range information were discovered with respect to the visual, verbal, or total response time or the number of system identification, parameter identification, parameter status, or total errors per trial.

This lack of results for the effects of including expected-value-range information on these traditional display formats was unexpected. One possible explanation for this lack of effects is that more than one visual cue was provided which enabled the subjects to perform the experimental task. Specifically, parameters with values that were outside the expected-value range were redundantly coded by a color change as described above in the section entitle "Experimental Variables." Since the subjects'

task was to identify the abnormal parameters and the status of their values, it appears that these subjects were primarily using the color change to detect out-of-range conditions. Another possible explanation for the lack of effects is that there were no within-cell replications in the experimental design. This artifact of the design could have allowed variations in performance within a single subject to have a detrimental effect on the sensitivity of the statistical tests that were involved.

Additional Findings

Parameter format. Many of the additional statistically significant results suggest that the bow tie bargraph format was more difficult for the subjects to use for this task than the circular gauge format. For example, the significant interactions between the parameter display format and the level of scenario complexity for visual ($F_{6,66} = 2.89$, $p < 0.05$), verbal ($F_{6,66} = 3.77$, $p < 0.01$), and total ($F_{6,66} = 4.16$, $p < 0.01$) response times, as shown in figures 10, 11, and 12, respectively, all indicated that the response times for the bow tie bargraphs were significantly higher than those for the circular gauges and that these differences were largest for the more complex scenarios.

These results were consistent with expectations, since the circular gauge format was much closer in appearance to the electromechanical round dials that today's pilots have used for most of their career. The bow tie bargraph format, although used in several current-generation aircraft, is still relatively new. Interestingly, the fact that the verbal and visual response times were longer for the bow tie bargraph format indicated that the subjects may have had more difficulty mentally decoding bargraph information as well as visually scanning it. The interactions with scenario complexity also suggested that this familiarity with the circular gauge format was most important for the complex scenarios, since less excess mental processing capacity was presumably available.

Level of display by exception. The additional experimental results that were related to the differences among the four levels of the display by exception favored the levels presenting the least amount of information with respect to errors in the response. These results, however, were inconclusive with respect to response time. The significant ($F_{18,198} = 10.22$, $p < 0.0001$) interaction between the level of display by exception and the level of scenario complexity for system identification errors, as shown in figure 13, indicated that the subjects made fewer mistakes when using the levels that presented the

least amount of information; however, this occurred only for two of the more complex scenarios (i.e., the ones with more affected systems and more abnormal parameters, such as scenarios 5 and 6). As discussed in appendix A, the small number of system identification errors for scenario 7 indicated that it was not a good example of the level of scenario complexity which it was designed to represent and that scenarios 5 and 6 were generally perceived by the subjects to be more difficult.

The response time results were less compelling. As expected, the significant ($F_{18,198} = 4.50$, $p < 0.0001$) interaction between the level of display by exception and the level of scenario complexity (fig. 14) showed that the visual response times were widely separated for the simpler scenarios and that the quickest times resulted from those levels of display by exception with the fewest gauges to scan. However, the mean times for the different levels tended to converge as the complexity of the scenario increased. This tendency to converge was consistent with expectations, since the differences between the amount of information presented in the various levels diminished as the number of abnormal parameters increased. The significant interactions between the level of display by exception and the level of scenario complexity for verbal ($F_{18,198} = 2.49$, $p < 0.01$) and total ($F_{18,198} = 3.27$, $p < 0.0001$) response times, as shown in figures 15 and 16, respectively, showed interesting results for scenario 6, which was determined to be the most complex scenario (as discussed in appendix A). In this scenario, the levels of display by exception which displayed the least amount of information resulted in the longest verbal response times, but they had the fewest errors (fig. 13). It was unclear why the mean verbal response times increased as the total amount of information displayed decreased, especially since reducing the amount of displayed information appeared to have reduced the apparent complexity of the scenario, as evidenced by the fewer system identification errors per trial. Although the instructions given to the subjects emphasized accuracy over speed, there was no obvious bias that should have led them to follow more or less this advice for the different levels of display by exception.

Parameter Format Experiment

The following sections describe the experimental method, the experimental hypotheses, the test procedures, the data analysis, and the results and discussion for the Parameter Format Experiment. Again, because this experiment was conducted jointly with the Expected Value Experiment, the relationship between the objective results and the subjective data are covered jointly with the results of that experiment in the section entitled "Subjective Evaluation Results."

Experimental Method

The Parameter Format Experiment examined the differences in subject task performance when using the circular gauge and the bow tie bargraph parameter formats (with expected-value-range information added) and the column deviation format. The same 12 test subjects were used for this experiment as for the Expected Value Experiment, and the experimental task was identical.

The independent variables for the Parameter Format Experiment included the type of parameter display format (circular gauges and bow tie bargraphs with expected-value-range information included and column deviations), the level of application of the display-by-exception philosophy (levels 1 through 4, as defined in table I), and the level of scenario complexity (levels 1 through 7, as defined in table II). This experiment was also constructed as a repeated-measures, full-factorial design, in which each of the 12 test subjects was tested once on each combination of the independent variables (i.e., there were no replications within experimental cells). This experiment resulted in a total of 84 treatments per subject for the 3 × 4 × 7 design. As before, all presentation sequences of the independent variables were either counterbalanced or selected randomly to reduce the effects of learning and ordering bias on the results.

The dependent measures for the Parameter Format Experiment were the same as those for the Expected Value Experiment: the visual, verbal, and total response times; the number of errors in system identification (i.e., missed/extra systems), the errors in parameter identification (i.e., missed/extra parameters), the errors in parameter status (i.e., caution/warning or high/low status), and the total number of errors; and the preference data from the subjective evaluation.

Experimental Hypotheses

Parameter Display Formats

The earlier observations concerning the pilots' preexisting familiarity with circular gauges also applied for this experiment; however, Abbott's results (ref. 4) showed that error detection rates were much higher for a column-deviation-based display than for

a combination circular gauge and bow tie bargraph display similar to those used in current "glass" cockpits. Therefore, it was expected that this ease of recognition would result in both lower response times and fewer errors in the response compared with circular gauges and bow tie bargraphs (with expected-value-range information). It also was expected that using circular gauges, because of their familiarity, would produce faster response times and fewer errors in the verbal response than in the bow tie bargraphs. It was unclear, however, because of the novelty of the column deviation format, whether the pilots would subjectively favor this display over the others.

Other Factors

All the expectations described above for the effects of presentation style and scenario complexity on the subjects' task performance in the Expected Value Experiment also apply to the Parameter Format Experiment.

Test Procedures

Because the Parameter Format Experiment was conducted concurrently with the Expected Value Experiment, the test procedures used were identical.

Data Analysis

Visual, verbal, and total response time data for the Parameter Format Experiment were analyzed using a mixed-model analysis of variance (ANOVA), with a significance level of 0.05. The differences between the means of the effects found to be significant in the ANOVA were assessed using the Waller-Duncan procedure (ref. 16). As with the Expected Value Experiment, both the mixed-model ANOVA and the Waller-Duncan tests were computed using the General Linear Models (PROC GLM) procedure in the SAS/STAT® statistical analysis computer program (ref. 17).

System identification, parameter identification, parameter status, and total error data were also analyzed using a mixed-model ANOVA followed by the Waller-Duncan tests, where appropriate. For the reasons given in the section entitled "Data Analysis" for the Expected Value Experiment, the error count data in the Parameter Format Experiment were also investigated using the more conservative mixed-model ANOVA instead of the usual chi-square analysis.

For the response time and error data analyses, note that the significant main effects that are also included in the significant interactions with other effects normally will not be examined separately in the

discussion below. Note also that unless specifically mentioned, inspection of the few statistically significant higher order interactions (three-way interactions and above) did not reveal any discernible patterns and therefore are not reported.

Results and Discussion

Primary Results

The primary objective of the second experiment, as stated in the Introduction, was to determine if using a nontraditional parameter display format (the column deviation format), which presented relative deviation from expected value, had any significant effect on subject performance compared with traditional parameter display formats with expected-value ranges included. The results relating to errors in the responses clearly indicated that the subjects found it easier to identify the presence of abnormal parameters in a particular system when the information was presented using the column deviation format compared with the circular gauge and bow tie bargraph formats. As shown in the significant ($F_{12,132} = 3.01$, $p < 0.001$) interaction between parameter format and scenario complexity for the mean number of system identification errors in figure 17, far fewer errors were made identifying affected systems using the column deviation format, especially for the most complex scenarios (i.e., scenarios 5, 6, and 7). This reduction of errors suggested that presenting parameter information using the column deviation format reduced the apparent task complexity. This result was consistent with expectations, since the column deviation format was expressly designed for detecting abnormalities during parameter monitoring tasks. This finding also agrees with the results found by Abbott in reference 4.

In addition, the response time results suggested that the subjects could perform the task more quickly using the circular gauge and column deviation formats than they could with the bow tie bargraph format, especially for the most complex scenarios. For example, the significant interactions between parameter format and scenario complexity for the mean visual ($F_{12,132} = 2.25$, $p < 0.05$), verbal ($F_{12,132} = 2.88$, $p < 0.01$), and total ($F_{12,132} = 2.99$, $p < 0.001$) response times, shown in figures 18, 19, and 20, respectively, indicated that it usually took the subjects as long or longer to respond both visually and verbally to the scenarios when using the bow tie bargraph format than when using either the circular gauge or column deviation format. The largest differences among the parameter formats occurred in the more complex scenarios, as expected, but the

similarity in response times for the more simple scenarios prevented the main effect of parameter format (across all levels of complexity) from reaching statistical significance.

Many other issues beyond those examined in this particular study, however, still remain concerning the use of column deviation formats for parameter presentation. For example, the ability to rapidly scan column deviation display formats may be affected if the parameter information is distributed among the different elements of a system schematic. Moreover, it was shown by Kieras (ref. 20) that such a distribution of parameter data within a system schematic (so that each parameter is in close proximity to its related component) is the preferred method to present this information for diagnosis. The pictorial system status format of Summers (ref. 21) also distributed the parameter information within the schematic, and this format was shown to be better than a textual format for abnormal condition detection and system identification. It is therefore recommended that additional studies be performed to determine if different methods of presenting system parameter information, depending on the set of tasks to be performed, will enhance overall performance.

Additional Findings

The observations made with respect to differences in performance among the four levels of display by exception in the Expected Value Experiment were also supported by the additional results of this experiment: the error results generally favored the levels presenting the least information, but the response time results were inconclusive. As in the first experiment, the subjects usually made fewer mistakes with the levels of display by exception which presented the least amount of information. The significant ($F_{18,198} = 6.92$, $p < 0.0001$) interaction between the level of exception and the level of scenario complexity for the mean number of system identification errors, as depicted in figure 21, showed that the greatest contribution to the decrease in system identification errors as information was removed from the display occurred during one of the most complex scenarios (i.e., scenario 6). However, the significant ($F_{18,198} = 2.27$, $p < 0.01$) interaction between the level of display by exception and the level of scenario complexity for the mean number of parameter status errors, as shown in figure 22, does not offer evidence that reducing the amount of displayed information reduces the perceived complexity of the scenario.

As in the first experiment, the response time results were statistically significant, but they were inconclusive. For example, the significant ($F_{18,198} = 2.18$, $p < 0.01$) interaction between the level of display by exception and the level of scenario complexity for mean visual response time, seen in figure 23, followed the pattern discovered in the first experiment. The visual response times for the four levels of display by exception tended to converge as the scenarios became more complex because the differences in amount of information presented among the four levels decreased as the number of abnormal parameters increased.

Subjective Evaluation Results

A subjective evaluation in the form of a structured questionnaire (the "Subjective Evaluation Parameter Presentation Study," as seen in appendix B) was administered to the subjects after all experimental trials for the Expected Value Experiment and the Parameter Format Experiment were completed. In questions 1 through 12, the subjects were asked to evaluate their own performance for each parameter format, expected-value-range status, and level of display by exception (referred to as "presentation style" in the questionnaire), with regard to how easily and quickly they could detect out-of-tolerance conditions and remember these conditions for the verbal response. Summaries of the responses to questions 1 through 12 on the subjective evaluation are presented in figures 24 through 35, respectively. In questions 13 and 14, the subjects were asked to rank the three combinations of format, expected-value range, and level of display by exception that they liked the best and the three combinations that they liked the least. Summaries of these responses are presented, grouped by test condition, in figures 36 through 40. In questions 15 and 16, the subjects were asked to provide the reasons for their rank selections from questions 13 and 14. Finally, in questions 17 through 19, the subjects were asked for general comments on how the display concepts could be improved. Summaries of all the responses to questions 15 through 19 are not presented, but selected comments are incorporated in the following discussion.

For questions 1 through 12 of the subjective evaluation, average responses to each question for all 12 test subjects were computed. Because these questions all dealt with deviations from a null response, the average responses were calculated by assigning specific values to each of the blocks on the rating scale used to answer the questions. The extreme left-hand block, which corresponded to answers such as "very slowly" or "very difficult," and the extreme right-hand block, which corresponded to answers such as

"very quickly" or "very easy," were assigned values of −2 and 2, respectively. The blocks just left of center and just right of center, which corresponded to answers such as "somewhat slowly" or "somewhat difficult," and "somewhat quickly" or "somewhat easy," respectively, were assigned values of −1 and 1. The center block corresponded to the answer "neither," and it was assigned a value of 0. An average response for all 12 test subjects that exceeded 0.75 was considered moderate support for assuming a nonnull response, while an average response that exceeded 1.50 was considered strong support.

The results from questions 13 and 14 of the subjective evaluation were combined to provide a single average response for each condition. The mean response was calculated using a value ranging from −3 for the least-liked response from question 14, to 3 for the best-liked response from question 13. No null response existed for these questions. The average responses were considered to be different from one another only if these differences exceeded 0.75.

Parameter Display Formats

The responses to the first four questions of the subjective evaluation, summarized in figures 24 through 27, indicated that many of the pilots believed their performance was better using the column deviation format compared with the circular gauge or bow tie bargraph formats. Specifically, the responses to questions 1, 3, and 4 all strongly indicated that the subjects believed that when using the column deviation format, they could more quickly detect out-of-tolerance parameters, more quickly determine whether those out-of-tolerance parameters were too high or too low, and more easily remember (for the verbal response) whether those out-of-tolerance parameters were too high or too low. The responses to question 1 also moderately indicated that the subjects believed that when using the circular gauge parameter format they could more quickly detect out-of-tolerance parameters.

In addition, as seen in figures 38, 39, and 40, respectively, the combined average responses to questions 13 and 14 for circular gauges and bow tie bargraphs (with expected-value-range information) and column deviations moderately indicated that the subjects preferred column deviations over both circular gauges and bow tie bargraphs. Interestingly, when the responses were subdivided by level of display by exception, it was seen that only the column deviation format received positive feedback all four styles. The combined average responses for circular gauges and bow tie bargraphs (with expected-value-

range information) suffered most from negative feedback for level of the display by exception in which all parameters were displayed all the time. All three of the parameter formats, however, were strongly preferred when displayed using level 3 of the display by exception, in which all parameters of the same type were displayed for affected systems when one of those parameters was out of tolerance.

Comments from the subjects also indicated that, in general, they preferred the column deviation format. Remarks such as "easy to scan, analyze, and remember," "more logical," "by far the easiest to read and interpret," "seemed to make abnormal indications much more prominent," and "jumps out and grabs your attention" were common. Several pilots also commented, however, that they preferred circular gauges and bow tie bargraphs because of "clarity and familiarity" and because they are "common on most aircraft that I operate" and "what I am most trained to use."

These results usually agreed with the objective results of the Parameter Format Experiment because the use of the column deviation format resulted in significantly fewer errors in system identification; fewer errors meant that the pilots were having less difficulty remembering which systems were affected by faults. Also, response times with the column deviation format were essentially the same as with the circular gauge format, and both these formats resulted in faster responses than with the bow tie bargraph format, even though the pilots were more familiar with the circular gauge and bow tie bargraph formats.

Expected-Value Range

The responses to questions 5 through 8 of the subjective evaluation, summarized in figures 28 through 31, respectively, indicated that many of the pilots believed that the expected-value ranges helped them more quickly detect problems. Specifically, the responses to questions 5 and 7 moderately indicated that with the expected-value ranges included, the subjects believed they could more quickly detect out-of-tolerance parameters and more rapidly determine whether these parameters were too high or too low. As seen in the responses to questions 6 and 8, however, the pilots did not in general believe that expected-value ranges helped them remember the scenario conditions any better for the verbal response.

The combined average responses from questions 13 and 14 of the subjective evaluation, for circular gauges and bow tie bargraphs both without

(figs. 36 and 37) and with (figs. 38 and 39) expected-value ranges included, offered strong support that the subjects preferred having the expected-value-range information on the display. This preference is seen particularly when the combined average responses for circular gauges without expected-value ranges (-1.364) and with expected-value ranges (0.733) are compared because this is the parameter format with which the pilots were most familiar.

Comments from the pilots also indicated that they generally preferred having the expected-value-range information included on the traditional parameter display formats. One pilot "liked [the] expected-value range (but didn't realize at first that I was using it)." Another pilot observed that "expected-value ranges would be a welcome addition...hard range marks are of limited value in a dynamic environment." A third pilot, however, remarked that adding expected-value ranges resulted in "too much clutter" and made it harder to "extract what's important."

The preferences of the pilots for the expected-value-range information were not supported by the objective data. However, some of the subjective comments indicated that the pilots were extrapolating the usefulness of this information into situations and conditions beyond those used in these experiments. Other comments suggested that additional training and operational experience may be required before the information is used to its full potential.

Levels of Display by Exception

The responses to questions 9 through 12 of the subjective evaluation, summarized in figures 32 through 35, indicated that many of the pilots believed their performance was better using the two levels of display by exception that presented the least amount of information. Specifically, the responses to all of these questions moderately indicated that the subjects believed that, using levels 3 and 4, they could more quickly detect and more easily remember (for the verbal response) out-of-tolerance parameters and more quickly determine and more easily remember whether those out-of-tolerance parameters were too high or too low.

The responses from questions 13 and 14 of the subjective evaluation, however, did not always indicate support for those beliefs. Although the level presenting the most information (level 1) received negative feedback for four of the five test conditions (as seen in figs. 36 through 39), level 4, which presented the least information, also received some negative feedback (as seen in figs. 36 through 38). Level 3 received the most positive feedback in four of the five

conditions. For the column deviation format (fig. 40), the subjects indicated a higher preference for level 1 than for level 4; however, level 3 again received the most positive feedback.

Comments from the subjects indicated that they were divided concerning whether the levels of display by exception that showed less information were really better than the others. For example, several of the pilots thought that level 1, which displayed all the parameters all of the time, was "too busy" or "too cluttered." Other pilots also thought that "the parameters should not be displayed unless there is a problem." However, many pilots also wanted other parameters shown for comparison. Remarks such as "I like being able to compare with the other similar systems" and "I like having all parameters up if the display is not too busy" were common. Even these pilots, though, tended to like level 3, which displayed all similar gauges for a system even if only one was out of tolerance. One pilot stated that "...style [level] 3 was [the] best declutter mode."

These findings agree with the objective results of the Expected Value Experiment and the Parameter Format Experiments because levels 3 and 4 of the display by exception generally had fewer errors and quicker response times across different parameter formats and levels of scenario complexity.

Conclusions and Recommendations

The inclusion of expected-value information onto traditional parameter display formats had essentially no effect on the performance of the subjects for the given task in this study. However, the subjective evaluation results indicated that the subjects showed moderate support for including this information. Because specific comments also indicated that the subjects, all experienced pilots, were mentally extrapolating the usefulness of this information into situations other than those encountered in this study, it is recommended that display designers consider including this type of information for tasks that include detecting system abnormalities that develop over a period of time.

Use of the nontraditional parameter display format implemented in this study, a column deviation format based on Abbott's (ref. 4) Engine Monitoring and Control System (E-MACS) concept, resulted in significantly fewer errors in system identification when compared with traditional parameter display formats with expected-value information included. In addition, use of the column deviation format appears to have resulted in less perceived complexity for this task because system identification error rates

remained relatively stable as the scenario complexity increased, whereas error rates for the traditional parameter display formats with expected-value ranges increased dramatically. The subjective evaluation results also indicate that the subjects thought that their own performance was somewhat better with this parameter format and that they generally preferred it over the others. It is therefore recommended that the column deviation format be considered for display applications that require rapid recognition of out-of-tolerance conditions, especially if a large number of system parameters must be monitored.

The effects of applying four different levels of the display-by-exception philosophy were mixed. Verbal response times were reduced as more extraneous information was removed from the display, but verbal response times for the most complex scenario were increased. Interestingly, the number of system identification errors was significantly reduced, without causing the expected increase in parameter status errors. Reducing the amount of information on the display also appeared to have reduced the per-ceived complexity of the scenario, since the number of system identification errors for the more complex scenarios decreased significantly as the amount of information decreased. In addition, the subjective results indicated that the subjects believed that the levels of display by exception with the least information actually improved the speed with which they could identify out-of-tolerance parameters, although they still had concerns about being able to see all the parameters when they needed to see them. These findings suggest that display designers should carefully evaluate the impact of including nonessential information on displays that are used primarily for rapid recognition of out-of-tolerance system conditions. When such displays are used for other system tasks, however, the subjective results suggest that all the information that supports those tasks should be presented.

Table I. Summary of Levels 1 Through 4 of Display by Exception

Level of display by exception	Description
1	All parameters for all systems are always displayed
2	Primary engine parameters are always displayed, but parameters for other systems are shown only when those systems are affected by problem
3	Only parameters of same type are displayed for affected systems when one or more is out of tolerance
4	Only out-of-tolerance parameters are displayed

Table II. Summary of Levels 1 Through 7 of Scenario Complexity

Level of scenario complexity	Description
1	No abnormal parameters are displayed
2	Single abnormal parameter is displayed in single system
3	Multiple abnormal parameters are displayed in single system
4	Single abnormal parameter is displayed in each of two systems
5	Multiple abnormal parameters are displayed in each of two systems
6	Single abnormal parameter is displayed in each of three systems
7	Multiple abnormal parameters are displayed in each of three systems

Appendix A

Results and Discussion of Scenario Complexity

The effects of scenario complexity on response time and accuracy were essentially as predicted for both the Expected Value Experiment and the Parameter Format Experiment; that is, increasing the complexity led to longer response times and more errors. In addition, increasing the complexity of the scenarios also helped to highlight differences between display formats which may otherwise have remained obscured by uniformly good performance on the task. Both experiments, however, indicated that although the experimental measures that were employed definitely detected changes in performance caused by scenario complexity, the set of dimensions originally used to define complexity for this study was apparently incomplete. This incompleteness was due to the level of scenario complexity not being perceived by the subjects as monotonically increasing with scenario number, as intended. For instance, the interactions between parameter format and level of scenario complexity for visual, verbal, and total response times in the Expected Value Experiment, seen in figures 10, 11, and 12, respectively, all showed that regardless of the parameter format, the subjects took longer to respond to situations that were more complex. The interactions between level of display by exception and level of scenario complexity for mean verbal and total response times in the Expected Value Experiment, however, as seen in figures 15 and 16, respectively, also suggested that scenario 6 was significantly different from either scenario 5 (which was designed to be less complex) and scenario 7 (which was designed to be more complex).

The most striking evidence of the differences in perceived complexity between scenario 6 and scenarios 5 and 7 was seen in the error results. The significant differences in mean number of parameter identification $(F_{6,66} = 13.94, \quad p < 0.0001)$, parameter status $(F_{6,66} = 4.53, \ p < 0.001)$, and total $(F_{6,66} = 18.14, \ p < 0.0001)$ errors for each of the seven scenarios in the Expected Value Experiment are shown in figures A1, A2, and A3, respectively. For the parameter identification and the total number of errors (figs. A1 and A3), a Waller-Duncan test showed that the means for the seven levels of scenario complexity were divided into three clusters. The first

cluster contained only scenario 6; the second contained scenarios 5 and 7; and the third contained the rest. For the parameter status errors (fig. A2), the Waller-Duncan test showed that the means were divided into only two overlapping clusters, with scenarios 6, 5, 7, and 4 in the first cluster and scenarios 4, 3, 2, and 1 in the second cluster. Taken together, these results indicate that scenario 6 was seen to be much more difficult than the others; this difficulty extended to the point that the mean total number of errors per trial for scenario 6, as seen in figure A3, was greater than 1.

The error results for the Parameter Format Experiment were similar. The significant differences in mean number of parameter identification $(F_{6,66} = 11.79, \ p < 0.0001)$, parameter status $(F_{6,66} = 2.81, \ p < 0.05)$, and total $(F_{6,66} = 17.66, \ p < 0.0001)$ errors for each of the seven scenarios are shown in figures A4, A5, and A6, respectively. For the parameter identification errors (fig. A4), a Waller-Duncan test showed that the means for the seven levels of scenario complexity were divided into four clusters, three of which overlapped. The first cluster contained only scenario 6; the second contained scenarios 5 and 7; the third contained scenarios 7 and 4; and the fourth contained scenarios 4, 2, 3, and 1. For the parameter status errors (fig. A5), the means were divided into only two overlapping clusters, with scenarios 7, 6, 5, and 4 in the first and 5, 4, 3, 2, and 1 in the second. Finally, the means for the total number of errors (fig. A6) were divided into three disjoint clusters. The first cluster contained scenario 6; the second contained scenarios 5 and 7; and the third contained the rest. As in the Expected Value Experiment, scenario 6 was apparently the most difficult.

Further inspection of the descriptions of the scenarios used for this study indicated that scenario 7 was probably seen by the subjects as less complex than scenario 6 for at least two reasons. First, the description of scenario 7 lists the left and right engines as two separate systems, when in fact the subjects appeared to have considered them together, especially since they exhibited identical behavior in this case. Second, all abnormal parameters in the three affected systems (left engine, right engine, and left electrical system) for scenario 7 had the same status; that is, the parameters were all low and in the warning region. This regularity apparently had the effect of creating a situation that was much less complex to grasp, even though scenario 7 had a total of

nine abnormal parameters, while scenario 6 had only three.

Based on the scenario complexity results of the Expected Value Experiment and the Parameter Format Experiment, a new set of dimensions to define the complexity of a statically viewed scenario should include the following: (1) the number of affected systems of different types; (2) the number of abnormal parameters in each affected system; (3) the number of differences in status (caution or warning range) for the abnormal parameters, both within and between affected systems; and (4) the number of differences in qualitative value (too high or too low) for the abnormal parameters, both within and between affected systems. Additional dimensions, such as the rate of change for each parameter, will clearly be necessary if dynamic scenarios are used.

Appendix B

Written Test Materials

The written materials for this experiment included the Pilot Background Questionnaire, the Pilot Briefing: Parameter Presentation Study, the Trial Response Form (which was filled out by the researcher during the experiment), and the Subjective Evaluation: Parameter Presentation Study. All materials are included here exactly as seen by the subjects. Please note that the experimental variable "level of display by exception" was referred to as "presentation style" in these materials and during all interactions with the subjects.

PILOT BACKGROUND QUESTIONNAIRE

Name: _____

Address: _____

_____

_____

Home Phone: _____   Work Phone: _____

Other Phone: (if applicable): _____

Employer: _____

_____

Rank (Position): _____

Years Flying Commercial: _____   Flying Military: _____

Approximate Total Time (Hours): _____

Flight Experience - Type of aircraft and approximate hours (No GA):

Type   Approx. Hours in Type

_____

_____

_____

_____

_____

Simulator Experience - Type of simulator and approximate hours:

_____

_____

_____

_____

PILOT BACKGROUND QUESTIONNAIRE

Would you be interested in participating in (check each that applies):

      _____ tests which require flying the NASA B-737? (must be 737 qualified)

      _____ tests which require flying one of the NASA simulation facilities?

      _____ evaluations of new displays and flight deck systems? (no flying)

      _____ interviews and studies relating to aircraft safety, automation, etc.?

Have you ever participated in a NASA research project before? _____

    If YES, then either briefly describe the test/interview or give the researcher's name who conducted the test:

_____

_____

_____

How much lead time will you require for scheduling appointments?

    _____

Please list any colleagues who may be interested in participating in NASA studies (include address and phone number, if known):

_____

_____

_____

Please return this form to:

Pilot Questionnaire
Vehicle Operations Research Branch
Mail Stop 156A
NASA Langley Research Center
Hampton, VA 23665

THANK YOU FOR YOUR COOPERATION!

PILOT BRIEFING
PARAMETER PRESENTATION STUDY

Introduction

        The purpose of this study is to identify improved methods for
presenting system parameter information to permit faster and better
recognition of system faults and status. Five different display formats
will be tested. Each of these five formats will have four different
styles of "display-by-exception" applied to them, so that the total
amount of information will be different depending on what has gone wrong.

        The study will be conducted in five parts, one for each of the five
display formats. Within each of these parts, each of the four display-
by-exception styles will be used with a series of aircraft system
states. Your tasks will be to determine which systems are affected in
each state that you view and what the status is of the parameters for
those affected systems.

        Please remember that this study is not attempting to develop new
instrument display formats for immediate commercial aircraft use.
Rather, it is an attempt to develop guidelines for future development of
display formats for system parameter information. This is an important
distinction.

Display Formats

        The five parameter display formats you will see, (1) circular
gauges, (2) bargraphs, (3) circular gauges with expected value range,
(4) bargraphs with expected value range, and (5) column deviations, are
shown in Figures 1 through 5 [figs. B1 through B8]. Basically, they are
all variations on the following form:

```
+-----------------------------------------+
|                                         |
|                                         |
|                                         |
|                 Engines                 |
|                                         |
|                                         |
|      -----------------------------      |
|                     |                   |
|                     |                   |
|         Oil         |    Electrical     |
|                     |                   |
|      ---------------|-------------      |
|                     |                   |
|       Hydraulic     |      Fuel         |
|                     |                   |
+-----------------------------------------+
```

Note that there are really only three different types of display elements: circular gauges, bargraphs, and column deviations. To create formats 3 and 4, a new piece of information, an expected value range, has been added to the circular gauges and bargraphs. This expected value range sweeps out a range of ±10% around the expected value of the parameter at any given time. As long as the pointer lies within this range, the value of the parameter is considered essentially normal for the current conditions. If the pointer lies below the arc, then the value is significantly lower than expected, and if it is above the arc, then it is significantly higher than expected. Either case will generate a caution or warning condition, depending on the magnitude of the difference. Cautions or warnings will also be generated if the parameter enters a fixed caution or warning range for that parameter, even if the expected value also falls within those ranges for the current conditions.

The column deviation format shown in Figure 5 [fig. B5] will require some additional explanation. The height of any bar in this display format does not depict the current value of the parameter it represents; rather, it depicts the <u>difference</u> between the current value and the expected value. A bar that extends below the centerline shows that the current value is lower than expected, while one that extends above the line shows that the current value is higher than expected. The additional two lines both above and below the centerline mark the regions of caution and warning alerts. A caution alert is generated whenever a difference of 10% or more exists between the current and expected values for a parameter, and a warning alert is generated whenever a difference of 15% or more exists. In addition, if a parameter value exceeds either the caution or warning limits for that parameter (whether it matches the expected value or not), the bar height reflects the proximity to the appropriate limit.

The expected values for all the formats are generated by a numerical model of the engine, and since this model cannot perfectly predict the engine output, it is normal for slight differences to exist. On the circular gauges and bargraphs, this means that the pointer will frequently be a little off-center with respect to the expected value range, while on the column deviation display it means that short bars either above or below the centerline may be present.

Also, on the circular gauge and bargraph display formats both with and without the expected value range, labelling will normally appear for all gauges that are displayed. For the column deviation display format, however, no gauge labelling will appear except for those individual parameters that are either significantly different from their nominal value or in a fixed caution or warning region. However, each display area for column deviations is labelled with the name of the appropriate aircraft subsystem. All gauge labels were included in Figure 5 [fig. B5] to let you see where the various parameters are displayed.

To summarize, there are three different display element types: circular gauges, bargraphs, and column deviations. Two additional display formats were created by adding an expected value range to the

circular gauges and bargraphs, which makes a total of five display formats: (1) circular gauges; (2) bargraphs; (3) circular gauges with expected value range; (4) bargraphs with expected value range; and (5) column deviations. The test will be conducted in five sections, one for each of the above formats.

<u>Display-by-Exception Styles</u>

The display formats as shown on the previous pages all use the same style: (1) all parameters for all systems are always displayed. The remaining styles are as follows: (2) primary engine parameters are always displayed, but parameters for other systems are only displayed when those systems are affected by a fault; (3) all parameters of the same type are displayed for affected systems when one of those parameters is out-of-tolerance (e.g. all hydraulic pressure gauges are displayed even if only one of them is low); and (4) only out-of-tolerance parameters are displayed for affected systems. An example of each of these remaining display-by-exception styles, using the circular gauge format with expected value range, is shown in Figures 6 through 8 [figs. B6 through B8].

<u>Parameter Values</u>

The values used for all parameters are representative of a Boeing 737-200 while in cruise flight at about FL330, at about Mach 0.7, on a near standard day. To aid your memory, the following list gives the normal range of values for each parameter under these conditions:

```
Engines:  EPR:  1.65  to   1.80
          N₁ :  79.0  to   83.0  %RPM
          EGT:   320  to    340  °C
          N₂ :  81.0  to   84.0  %RPM
          FF :  1.89  to   2.12  x 1000 pph

Oil:      Pres:   42  to   52  psi
          Temp:   70  to   90  °C
          Quan:  3.0  to  4.0  gal

Hydraulic:    Pres:  2950  to  3050  psi
              Temp:    52  to    62  °C
              Quan:   3.1  to   3.3  qts

Electrical:   Gen Oil Temp:  70  to   90   °C
              Volts:        113  to  117   V
              Frequency:    398  to  402   Hz

Fuel:     Quan:  5.50 to  6.50  x 1000 lbs
          Temp:   −20 to  +20   °C
```

Whenever abnormal parameter values are displayed, they will differ significantly from the expected normal value and will fall outside of the ranges given above.

<u>Test Sequence</u>

To display the first aircraft system state on the computer screen, press the space bar on the keyboard. As soon as your have determined which systems are affected and the status of the parameters for those systems, press the space bar again. This erases the display and records how long you were looking at it. You will then state, verbally: (1) which aircraft systems were affected and (2) which parameters in those systems were abnormal and what the status of each abnormal parameter was (low warning, low caution, normal, high caution, or high warning). After you are done, press the space bar one more time. This records how long it took for you to provide the necessary information. (Some cases will take longer than others because of the number of affected systems and abnormal parameters.) After your answers are recorded by the researcher, press the space bar again to display the next aircraft system state. You will be notified when new formats and exception styles are about to be used. The entire study consists of 140 displays, which should take about two hours to complete.

While participating in this study, please place your emphasis on accuracy rather than speed. It's better to take a little longer and get everything right than to rush and make mistakes. And remember, this study is not testing <u>you</u>, it's testing the <u>displays</u>.

<u>Training Sequence</u>

Before beginning the actual test runs, you will participate in a brief training exercise to familiarize yourself with the different display formats and presentation styles. Don't worry at all about trying to go fast – just get comfortable with what you're going to see and do during the test. And if you have questions at all – PLEASE ASK!

TRIAL RESPONSE FORM

Subject Number: _____   Trial: _____

Affected Systems: Engines   Oil   Hydraulics   Electrical   Fuel

Affected Parameters:

Engine1: Epr _____ N_1 _____   Egt _____ N_2 _____   FF _____   Vib _____

Engine2: Epr _____ N_1 _____   Egt _____ N_2 _____   FF _____   Vib _____

Oil1: Pres _____   Temp _____   Quan _____

Oil2: Pres _____   Temp _____   Quan _____

HydraulicA: Pres _____   Temp _____   Quan _____

HydraulicB: Pres _____   Temp _____   Quan _____

HydraulicS: Pres _____   Temp _____   Quan _____

Electrical1: GenOilT _____   Volt _____   Freq _____

Electrical2: GenOilT _____   Volt _____   Freq _____

Fuel: QuanL _____   QuanC _____   QuanR _____   Temp _____

Response: _____

Trial: _____

Affected Systems: Engines   Oil   Hydraulics   Electrical   Fuel

Affected Parameters:

Engine1: Epr _____ N_1 _____   Egt _____ N_2 _____   FF _____   Vib _____

Engine2: Epr _____ N_1 _____   Egt _____ N_2 _____   FF _____   Vib _____

Oil1: Pres _____   Temp _____   Quan _____

Oil2: Pres _____   Temp _____   Quan _____

HydraulicA: Pres _____   Temp _____   Quan _____

HydraulicB: Pres _____   Temp _____   Quan _____

HydraulicS: Pres _____   Temp _____   Quan _____

Electrical1: GenOilT _____   Volt _____   Freq _____

Electrical2: GenOilT _____   Volt _____   Freq _____

Fuel: QuanL _____   QuanC _____   QuanR _____   Temp _____

Response: _____

SUBJECTIVE EVALUATION
PARAMETER PRESENTATION STUDY

For each of the following twelve questions (1-12), please make a mark
inside the block that best describes your answer. Do not mark on the
block dividers.

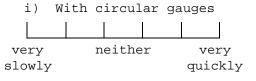Definitions: very slowly: extremely large amount of time required
                neither: neither particularly quickly nor slowly
                very quickly: little or no time required

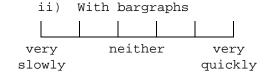   very difficult: intense mental effort required
                neither: neither particularly easy nor difficult
                very easy: little or no mental effort required
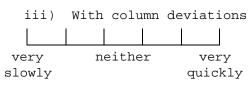
The blocks in between the extremes and the middle of each scale indicate
"somewhat." For example, the block between "very slowly" and "neither"
would correspond to "somewhat slowly," while the block between "neither"
and "very easy" would correspond to "somewhat easy."

NOTE: The following four questions (1-4) refer only to the three
      different types of display elements (circular gauges, bargraphs,
      and column deviations), regardless of whether or not expected
      value ranges were included or which display-by-exception
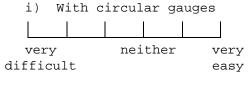   presentation style was used.

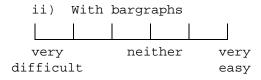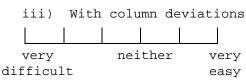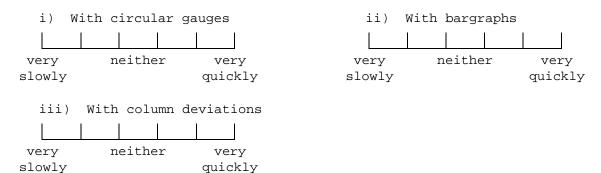1.   How quickly could you detect out-of-tolerance parameters?

        i) With circular gauges   ii) With bargraphs

        very   neither   very   very   neither   very
        slowly   quickly   slowly   quickly

        iii) With column deviations

        very   neither   very
        slowly   quickly

2.   How easy was it to remember which parameters were out-of-tolerance?

        i) With circular gauges   ii) With bargraphs

          very   neither   very   very   neither   very
        difficult   easy   difficult   easy

        iii) With column deviations

          very   neither   very
        difficult   easy

56

3.   How quickly could you determine whether the out-of-tolerance
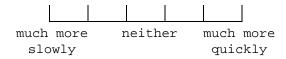   parameters were too high or too low?

         i) With circular gauges   ii) With bargraphs

     |___|___|___|___|___| |___|___|___|___|___|
      very   neither   very   very   neither   very
     slowly   quickly   slowly   quickly

         iii) With column deviations

     |___|___|___|___|___|
      very   neither   very
     slowly   quickly

4.   How easy was it to remember whether the out-of-tolerance parameters
     were too high or too low?

         i) With circular gauges   ii) With bargraphs

     |___|___|___|___|___| |___|___|___|___|___|
       very   neither   very   very   neither   very
     difficult   easy   difficult   easy

         iii) With column deviations

     |___|___|___|___|___|
       very   neither   very
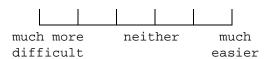   difficult   easy

NOTE: The following four questions (5-8) refer only to the addition of
      expected value ranges on the displays, regardless of the display
      element type or which display-by-exception presentation style was
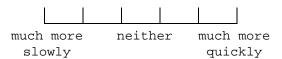   used.

5.   How much more quickly or slowly could you detect out-of-tolerance
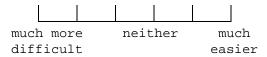   parameters when expected value ranges were added?

                         |___|___|___|___|___|
                         much more   neither   much more
                            slowly   quickly

6.   How much easier or more difficult was it to remember which
     parameters were out-of-tolerance when expected value ranges were
     added?

                         |___|___|___|___|___|
                         much more   neither   much
                          difficult   easier

7. How much more quickly or slowly could you determine whether the out-of-tolerance parameters were too high or too low when expected value ranges were added?

```
      |___|___|___|___|___|
```
      much more   neither   much more
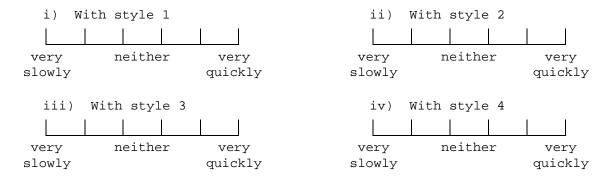   slowly   quickly

8. How much easier or more difficult was it to remember whether the out-of-tolerance parameters were too high or too low when expected value ranges were added?

```
      |___|___|___|___|___|
```
      much more   neither   much
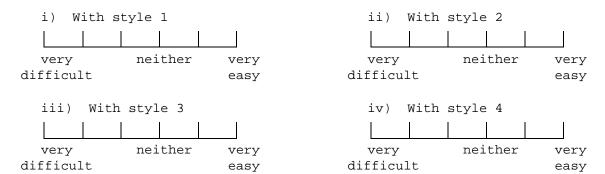   difficult   easier

NOTE: The following four questions (9-12) refer only to the four different display-by-exception presentation styles, regardless of the type of display element or whether or not expected value ranges were included. To refresh your memory, a brief description of each of the presentation styles is printed below.
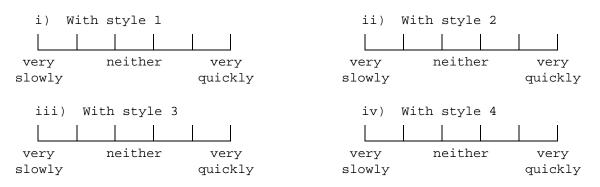
STYLES: 1) all parameters for all systems are always displayed;
          2) primary engine parameters are always displayed, but parameters for other systems are only displayed when those systems are affected by a fault;
          3) all parameters of the same type are displayed for affected systems when one of those parameters is out-of-tolerance (e.g. all hydraulic pressure gauges are displayed even if only one of them is low);
          4) only out-of-tolerance parameters are displayed for affected systems.

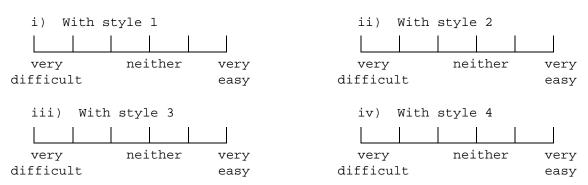9. How quickly could you detect out-of-tolerance parameters?

   i) With style 1   ii) With style 2
```
   |___|___|___|___|___|                  |___|___|___|___|___|
```
   very   neither   very   very   neither   very
   slowly   quickly   slowly   quickly

   iii) With style 3   iv) With style 4
```
   |___|___|___|___|___|                  |___|___|___|___|___|
```
   very   neither   very   very   neither   very
   slowly   quickly   slowly   quickly

10. How easy was it to remember which parameters were out-of-tolerance?

    i) With style 1

```
|___|___|___|___|___|
```
  very   neither   very
difficult   easy

    ii) With style 2

```
|___|___|___|___|___|
```
  very   neither   very
difficult   easy

    iii) With style 3

```
|___|___|___|___|___|
```
  very   neither   very
difficult   easy

    iv) With style 4

```
|___|___|___|___|___|
```
  very   neither   very
difficult   easy

11. How quickly could you determine whether the out-of-tolerance parameters were too high or too low?

    i) With style 1

```
|___|___|___|___|___|
```
  very   neither   very
slowly   quickly

    ii) With style 2

```
|___|___|___|___|___|
```
  very   neither   very
slowly   quickly

    iii) With style 3

```
|___|___|___|___|___|
```
  very   neither   very
slowly   quickly

    iv) With style 4

```
|___|___|___|___|___|
```
  very   neither   very
slowly   quickly

12. How easy was it to remember whether the out-of-tolerance parameters were too high or too low?

    i) With style 1

```
|___|___|___|___|___|
```
  very   neither   very
difficult   easy

    ii) With style 2

```
|___|___|___|___|___|
```
  very   neither   very
difficult   easy

    iii) With style 3

```
|___|___|___|___|___|
```
  very   neither   very
difficult   easy

    iv) With style 4

```
|___|___|___|___|___|
```
  very   neither   very
difficult   easy

For the following ranking questions (13-14), please write the appropriate number in the blank beside each selection you make.

STYLES: 1) all parameters for all systems are always displayed;
              2) primary engine parameters are always displayed, but parameters for other systems are only displayed when those systems are affected by a fault;
              3) all parameters of the same type are displayed for affected systems when one of those parameters is out-of-tolerance (e.g. all hydraulic pressure gauges are displayed even if only one of them is low);
              4) only out-of-tolerance parameters are displayed for affected systems.

13.   Please select the three combinations of display format, expected value range, and presentation style that you like the best (1=best, 2=next best, 3=third best). Pick only three.

_____	circular gauges	no expected value range	style 1
_____	circular gauges	no expected value range	style 2
_____	circular gauges	no expected value range	style 3
_____	circular gauges	no expected value range	style 4
_____	bargraphs	no expected value range	style 1
_____	bargraphs	no expected value range	style 2
_____	bargraphs	no expected value range	style 3
_____	bargraphs	no expected value range	style 4
_____	circular gauges	expected value range	style 1
_____	circular gauges	expected value range	style 2
_____	circular gauges	expected value range	style 3
_____	circular gauges	expected value range	style 4
_____	bargraphs	expected value range	style 1
_____	bargraphs	expected value range	style 2
_____	bargraphs	expected value range	style 3
_____	bargraphs	expected value range	style 4
_____	column deviations		style 1
_____	column deviations		style 2
_____	column deviations		style 3
_____	column deviations		style 4

STYLES: 1) all parameters for all systems are always displayed;
          2) primary engine parameters are always displayed, but
             parameters for other systems are only displayed when those
             systems are affected by a fault;
          3) all parameters of the same type are displayed for affected
             systems when one of those parameters is out-of-tolerance
             (e.g. all hydraulic pressure gauges are displayed even if
             only one of them is low);
          4) only out-of-tolerance parameters are displayed for affected
             systems.

14.   Please select the three combinations of display format, expected
      value range, and presentation style that you liked the least
      (1=worst, 2=next worst, 3=third worst). Pick only three.

_____	circular gauges	no expected value range	style 1
_____	circular gauges	no expected value range	style 2
_____	circular gauges	no expected value range	style 3
_____	circular gauges	no expected value range	style 4
_____	bargraphs	no expected value range	style 1
_____	bargraphs	no expected value range	style 2
_____	bargraphs	no expected value range	style 3
_____	bargraphs	no expected value range	style 4
_____	circular gauges	expected value range	style 1
_____	circular gauges	expected value range	style 2
_____	circular gauges	expected value range	style 3
_____	circular gauges	expected value range	style 4
_____	bargraphs	expected value range	style 1
_____	bargraphs	expected value range	style 2
_____	bargraphs	expected value range	style 3
_____	bargraphs	expected value range	style 4
_____	column deviations		style 1
_____	column deviations		style 2
_____	column deviations		style 3
_____	column deviations		style 4

For each of the following four questions (15-18), please include at
least one sentence for each requested response.

15.　　Please give the reasons for choosing your three most preferred
　　　　combinations from question 13.

_____

_____

_____

_____

16.　　Please give the reasons for choosing your three least preferred
　　　　combinations from question 14.

_____

_____

_____

_____

17.　　How could the format/presentation style combination that you most
　　　　preferred be improved further?

_____

_____

_____

_____

18.　　What situations might exist in which a format/presentation style
　　　　combination other than your most preferred one may be useful or
　　　　necessary? Which one(s)?

_____

_____

_____

_____

19. Please record any other comments, suggestions, or criticisms you may have about any of the display element types, the inclusion of expected value ranges, the display-by-exception presentation styles, or the way the experiment was conducted.

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

References

1. Abbott, Kathy H.; Schutte, Paul C.; Palmer, Michael T.; and Ricks, Wendell R.: Faultfinder: A Diagnostic Expert System With Graceful Degradation for Onboard Aircraft Applications. *Proceedings of the 14th Symposium on Aircraft Integrated Monitoring Systems*, Helmut Hardegen, compiler, DFVLR-MITT-88-04, Deutsche Forschungs- und Versuchsanstalt fuer Luft- und Raumfahrt (Brunswick, Germany), 1988, pp. 353–370.

2. Woods, David D.: *Cognitive Activities and Aiding Strategies in Dynamic Fault Management*. CSEL 92-TR-05, Ohio State Univ., Oct. 1992.

3. Graham, David K.: Electronic Display of Powerplant Parameters. *3rd Aerospace Behavioral Engineering Technology Conference*, Soc. of Automotive Engineers, Inc., Oct. 1984, pp. 6.128–6.224.

4. Abbott, Terence S.: Task-Oriented Display Design— Concept and Example. M.S. Thesis, College of William and Mary, 1989.

5. *Aircraft Accident Report—Air Florida, Inc., Boeing 737-222, N62AF Collision With 14th Street Bridge, Near Washington National Airport, Washington, D.C., January 13, 1983*. NTSB-AAR-82-8, National Transportation Safety Board, Aug. 1982. (Available from NTIS as PB82 910 408.)

6. *Aircraft Accident Report—American Airlines, Inc., DC-10-10, N110AA, Chicago–O'Hare International Airport, Chicago, Illinois, May 25, 1979*. NTSB-AAR-79-17, National Transportation Board, Dec. 1979.

7. *Aircraft Accident Report—China Airlines Boeing 747-SP, N4522V 300 Nautical Miles Northwest of San Francisco, California, February 19, 1985*. NTSB-AAR-86-03, National Transportation Safety Board, Mar. 1986. (Available from NTIS as PB86 910 403.)

8. *Report on the Accident to Boeing 737-400 G-OBME Near Kegworth, Leicestershire on 8 January 1989*. Aircraft Accident Report 4/90, Her Majesty's Stationary Office (London), 1990.

9. *Boeing 757-200 Systems*. D6-76549R5, Boeing Commercial Airplane Co., Aug. 1983.

10. Allen, Donald Mark: Investigation of Display Issues Relevant to the Presentation of Aircraft Fault Information. Ph.D. Diss., Old Dominion Univ., Apr. 1988.

11. *Aircraft Accident Report—Eastern Airlines Flight 935, Lockheed L-1011-384, N309EA, Near Colts Neck, New Jersey, September 22, 1981*. NTSB-AAR-82-5, June 1982. National Transportation Board. (Available from NTIS as PB82 910 405.)

12. *Aircraft Accident Report—Eastern Airlines, Inc., Lockheed L-1011, N334EA, Miami International Airport, Miami, Florida, May 5, 1983*. NTSB-AAR-84-04, National Transportation Safety Board, Mar. 1984.

13. *Boeing Model 737-100—Operations Manual, Volume 2*. Boeing Document D5-27370-100, Commercial Airplane Group, Boeing Co., Nov. 20, 1973.

14. Park, Stephen K.; and Miller, Keith W.: Random Number Generators: Good Ones Are Hard To Find. *Commun. ACM*, vol. 31, no. 10, Oct. 1988, pp. 1192–1201.

15. Giffin, W. C.; Rockwell, T. H.; and Smith, P. J.: *Research on Computer Aided Testing of Pilot Response to Critical In-Flight Events*. NASA CR-173871, 1984.

16. Millikin, George A.; and Johnson, Dallas E.: *Analysis of Messy Data. Volume I: Designed Experiments*. Van Nostrand Reinhold, 1984.

17. *SAS/STAT® User's Guide, Version 6, Fourth ed., Volume 2*. SAS Institute, Inc., 1990.

18. Kirk, Roger E.: *Statistics: An Introduction, Third ed.* Holt, Rinehart and Winston, Inc., 1990.

19. Neter, John; Wasserman, William; and Kutner, Michael H.: *Applied Linear Statistical Models—Regression, Analysis of Variance, and Experimental Designs, Third ed.* Richard D. Irwin, Inc., 1990.

20. Kieras, David E.: Diagrammatic Display for Engineered Systems: Effects on Human Performance in Interacting With Malfunctioning System. *Int. J. Man–Machine Studies*, vol. 36, 1992, pp. 861–895.

21. Summers, Leland G.: *System Status Display Evaluation*. NASA CR-4181, 1988.

Figure 21. Mean number of system identification errors per trial versus scenario for each level of display by exception for Parameter Format Experiment.

Figure 29. Responses to question 6 of subjective evaluation (concerning expected-value-range information): "How much easier or more difficult was it to remember which parameters were out of tolerance when expected-value ranges were added?"

Figure 36. Responses for circular gauges without expected-value-range information to questions 13 and 14 of subjective evaluation: "Please select the three combinations of display format, expected-value range, and presentation style [level of display by exception] that you like the least/best." Total number of responses across all four levels for this condition was 11, and combined average response was -1.364.

Figure 37. Responses for bow tie bargraphs without expected-value-range information to questions 13 and 14 of subjective evaluation: "Please select the three combinations of display format, expected-value range, and presentation style [level of display by exception] that you like the least/best." Total number of responses across all four levels for this condition was 14, and combined average response was -1.571.

Figure 38. Responses for circular gauges with expected-value-range information to questions 13 and 14 of subjective evaluation: "Please select the three combinations of display format, expected-value range, and presentation style [level of display by exception] that you like the least/best." Total number of responses across all four levels for this condition was 15, and combined average response was 0.733.

Figure 39. Responses for bow tie bargraphs with expected-value-range information to questions 13 and 14 of subjective evaluation: "Please select the three combinations of display format, expected-value range, and presentation style [level of display by exception] that you like the least/best." Total number of responses across all four levels for this condition was 11, and combined average response was 0.273.

Figure 40. Responses for column deviations to questions 13 and 14 of subjective evaluation: "Please select the three combinations of display format, expected-value range, and presentation style [level of display by exception] that you like the least/best." Total number of responses across all four levels for this condition was 21, and combined average response was 1.095.

Figure 1. Example of circular gauge parameter format.

Figure 2. Example of bow tie bargraph parameter format.

Figure 3. Example of circular gauge parameter format with expected-value-range information included.

Figure 4. Example of bow tie bargraph parameter format with expected-value-range information included.

Figure 5. Example of column deviation format with explanatory labels.

Figure 6. Level 1 of display by exception (in which all parameters for all systems are always displayed); circular gauge parameter format is used.

Figure 7. Level 2 of display by exception (in which primary engine parameters are always displayed, but parameters for other systems are only displayed when system is affected by fault); circular gauge parameter format is used.

Figure 8. Level 3 of display by exception (in which all parameters of same type are displayed if one or more is out of tolerance); circular gauge parameter format is used.

Figure 9. Level 4 of display by exception (in which only out-of-tolerance parameters are displayed); circular gauge parameter format is used.

Figure 10. Mean visual response time versus level of scenario complexity for each parameter format for Expected Value Experiment.

Figure 11. Mean verbal response time versus level of scenario complexity for each parameter format for Expected Value Experiment.

Figure 12. Mean total response time versus level of scenario complexity for each parameter format for Expected Value Experiment.

Figure 13. Mean number of system identification errors per trial versus level of scenario complexity for each level of display by exception for Expected Value Experiment.

Figure 14. Mean visual response time versus level of scenario complexity for each level of display by exception for Expected Value Experiment.

Figure 15. Mean verbal response time versus level of scenario complexity for each level of display by exception for Expected Value Experiment.

Figure 16. Mean total response time versus level of scenario complexity for each level of display by exception for Expected Value Experiment.

Figure 17. Mean number of system identification errors per trial versus level of scenario complexity for each parameter format for Parameter Format Experiment.

Figure 18. Mean visual response time versus level of scenario complexity for each parameter format for Parameter Format Experiment.

Figure 19. Mean verbal response time versus level of scenario complexity for each parameter format for Parameter Format Experiment.

Figure 20. Mean total response time versus level of scenario complexity for each parameter format for Parameter Format Experiment.

Figure 37. Responses for bow tie bargraphs without expected-value-range information to questions 13 and 14 of subjective evaluation: "Please select the three combinations of display format, expected-value range, and presentation style [level of display by exception] that you like the least/best." Total number of responses across all four levels for this condition was 14, and combined average response was −1.571.

Figure 38. Responses for circular gauges with expected-value-range information to questions 13 and 14 of subjective evaluation: "Please select the three combinations of display format, expected-value range, and presentation style [level of display by exception] that you like the least/best." Total number of responses across all four levels for this condition was 15, and combined average response was 0.733.

Figure 39. Responses for bow tie bargraphs with expected-value-range information to questions 13 and 14 of subjective evaluation: "Please select the three combinations of display format, expected-value range, and presentation style [level of display by exception] that you like the least/best." Total number of responses across all four levels for this condition was 11, and combined average response was 0.273.

Figure 40. Responses for column deviations to questions 13 and 14 of subjective evaluation: "Please select the three combinations of display format, expected-value range, and presentation style [level of display by exception] that you like the least/best." Total number of responses across all four levels for this condition was 21, and combined average response was 1.095.

Figure A1. Mean number of parameter identification errors per trial versus level of scenario complexity for Expected Value Experiment.

Figure A2. Mean number of parameter status errors per trial versus level of scenario complexity for Expected Value Experiment.

Figure A3. Mean total number of errors per trial versus level of scenario complexity for Expected Value Experiment.

Figure A4. Mean number of parameter identification errors per trial versus level of scenario complexity for Parameter Format Experiment.

Figure A5. Mean number of parameter status errors per trial versus level of scenario complexity for Parameter Format Experiment.

Figure A6. Mean total number of errors per trial versus level of scenario complexity for Parameter Format Experiment.

Figure B1. Circular gauges.

Figure B2. [Bow tie] bargraphs.

Figure B3. Circular gauges with expected-value range.

Figure B4. [Bow tie] bargraphs with expected-value range.

Figure B5. Column deviations.

Figure B6. Presentation style 2 using circular gauges.

Figure B7. Presentation style 3 using circular gauges.

Figure B8. Presentation style 4 using circular gauges.

4

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE March 1994	3. REPORT TYPE AND DATES COVERED Technical Paper

4. TITLE AND SUBTITLE
Effects of Expected-Value Information and Display Format on Recognition of Aircraft Subsystem Abnormalities

5. FUNDING NUMBERS
WU 505-64-13-22

6. AUTHOR(S)
Michael T. Palmer and Kathy H. Abbott

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NASA Langley Research Center
Hampton, VA 23681-0001

8. PERFORMING ORGANIZATION REPORT NUMBER
L-17262

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
NASA TP-3395

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified–Unlimited

Subject Category 06

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
This study identifies improved methods to present system parameter information for detecting abnormal conditions and to identify system status. Two workstation experiments were conducted. The first experiment determined if including expected-value-range information in traditional parameter display formats affected subject performance. The second experiment determined if using a nontraditional parameter display format, which presented relative deviation from expected value, was better than traditional formats with expected-value ranges included. The inclusion of expected-value-range information onto traditional parameter formats was found to have essentially no effect. However, subjective results indicated support for including this information. The nontraditional column deviation parameter display format resulted in significantly fewer errors compared with traditional formats with expected-value-ranges included. In addition, error rates for the column deviation parameter display format remained stable as the scenario complexity increased, whereas error rates for the traditional parameter display formats with expected-value ranges increased. Subjective results also indicated that the subjects preferred this new format and thought that their performance was better with it. The column deviation parameter display format is recommended for display applications that require rapid recognition of out-of-tolerance conditions, especially for a large number of parameters.

14. SUBJECT TERMS
Parameter formats; Status displays; Display monitoring; Expected-value information; Column deviation formats

15. NUMBER OF PAGES
75

16. PRICE CODE
A04

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT