Winter 2004

Crisp and Fuzzy Signal Detection Theory and Pilot Weather Judgment: Implications for VFR Flights Into IMC

Joseph T. Coyne
Old Dominion University

CRISP AND FUZZY SIGNAL DETECTION THEORY AND PILOT WEATHER JUDGMENT: IMPLICATIONS FOR VFR FLIGHTS INTO IMC

by

Joseph T. Coyne
B.A. May 1999, State University of New York at Albany
M.S. May 2002, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

PSYCHOLOGY

OLD DOMINION UNIVERSITY
December, 2004

Approved by:

Carryl L. Baldwin (Director)

James P. Bliss (Member)

J. Raymond Comstock Jr. (Member)

Kara A. Latorella (Member)

Bryan E. Porter (Member)

UMI Number: 3160634

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI®

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

CRISP AND FUZZY SIGNAL DETECTION THEORY AND PILOT WEATHER JUDGMENT: IMPLICATIONS FOR VFR FLIGHTS INTO IMC

Joseph T. Coyne
Old Dominion University, 2004
Director: Dr. Carryl L. Baldwin

Weather represents one of the greatest hazards to general aviation (GA), accounting for 15% of the GA accident fatalities. Of the fatal weather accidents 90% are attributed to visual flight rules (VFR) flight into instrument meteorological conditions (IMC). The situation assessment hypothesis suggests that pilots may inadvertently enter IMC because they lack the sensitivity needed to distinguish between visual meteorological conditions (VMC) and IMC. An alternative hypothesis is that pilots recognize conditions have deteriorated but are motivated by some other factor, such as pressure from passengers. The present study uses Jensen's Pilot Judgment Model and Signal Detection Theory to explain pilot judgment. The impact of Graphical Weather Information Systems (GWIS), particularly graphical METARs on pilot judgment was also assessed. Twenty-four general aviation pilots were shown simulated video images of different weather conditions. Several of these trials contained GWIS surface data of varying accuracy. Results indicated that pilots had trouble in distinguishing between VFR and IFR conditions, especially determining ceiling. Overall, pilots had a low sensitivity in determining whether the ceiling was VMC or IMC and tended to overestimate ceilings. This problem was amplified by an interaction with visibility. Pilots' estimates of IMC ceilings actually increased as the visibility increased. A similar

effect of ceiling influencing visibility judgments was also found. Pilots' judgments were additionally influenced by the inaccurate METAR information presented in the GWIS. GWIS data that suggested conditions were worse than those seen out-the-window caused a liberal shift in response bias, and conditions that were better than those out-the-window had a corresponding conservative shift in response bias. Overall the experiment found evidence to suggest both situation assessment and motivation could contribute to a decision to continue into IMC. The interaction of ceiling and visibility also suggests a new potential factor in inadvertent VFR flight into IMC. The improper evaluation of one weather dimension based upon a bias from the other weather dimension needs to be further examined for its role in pilots' decision to continue into deteriorating weather conditions.

This dissertation is dedicated to my parents and grandmother, who have always been supportive.

ACKNOWLEDGMENTS

This work was supported by the NASA Langley Research Center under grant number NGT-01-01030. I am grateful for the direction and support from all of the members of my dissertation committee. Special thanks to Regina Johns of Lockheed Martin for her dedicated work in participant procurement. Thanks also to Paul Stough, Jim Chamberlain, and several others at NASA Langley for their input and expertise.

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES xi

INTRODUCTION 1
- General Aviation Flying and Weather 1
- Jensen's Pilot Judgment Model 5
- Situation Assessment 10
- Motivational Judgment 24
- Signal Detection Theory 29
- Graphical Weather Information Systems 36

PRESENT STUDY 43
- Hypotheses 46

METHODOLOGY. 49
- Participants 49
- Apparatus and Materials 50
- Pre-experimental Questionnaire 50
- Experimental Instructions 51
- Weather Generation Program 51
- Post Experiment Task 53
- Procedure 53
- Dependent Measures 57
- Design 61

RESULTS 65
- Baseline SDT Results 65
- Baseline Accuracy, RMSE, and Comfort Data 67
- GWIS SDT Results 77
- GWIS Accuracy, RMSE, and Comfort Data 85
- Correlations 90

DISCUSSION 95
- Weather Estimation 96
- Pilot Sensitivity and Response Bias 98
- Pilot Rating 102
- Window Based Decision Making Strategies 104
- Card Sort Task 106
- Influence of the Graphical METARs 107

TABLE OF CONTENTS (continued)

Page

Fuzzy Signal Detection Theory 110
Ceiling and Visibility Interaction 112
2-D Simulation of 3-D 115
Limitations 116
Summary and Conclusions 117

REFERENCES 122

APPENDIXES
A. Weather Classification Based Upon FARs 127
B. Glossary 128
C. Signal Detection Formulas 130
D. Questionnaire 132
E. ANOVA and Means Tables 141
F. Individual SDT Metrics 156
G. Exploratory Ceiling by Visibility Analyses 158

VITA 166

LIST OF TABLES

Table Page

1. Nine Weather Cues Identified by Wiggins and O'Hare (2003) 21
2. Weather Response Categorization for Crisp SDT 30
3. Mean (and Standard Deviation) Data for Pilot Experience 49
4. GWIS Information Consistency Manipulations 57
5. Mapping Functions for Ceiling and Visibility 59
6. Fuzzy Response Categorization 61
7. Sources of Variation for the Baseline Analyses 63
8. Sources of Variation for the GWIS Analyses 64
9. Computation of Hit Rates and False Alarm Rates Before and After the Loglinear Correction 66
10. Crisp and Fuzzy SDT Measures for the Baseline Trials 66
11. Number of Extreme Scores in the Crisp and Fuzzy Ceiling and Visibility Analyses 68
12. Ceiling and Visibility Response Rates 68
13. Nonparametric SDT Measures for Ceiling and Visibility 69
14. Number of Cases with Perfect Performance within the GWIS SDT Data 78
15. Number of Extreme Scores within the GWIS Ceiling SDT Data 81
16. Mean Ceiling Crisp A' and Post Hoc Groupings for the GWIS Conditions 82
17. Mean Fuzzy and Crisp Ceiling c and Post Hoc Groupings for the GWIS Conditions 82
18. Number of Extreme Scores within the GWIS Visibility SDT Data 83
19. Mean Crisp Visibility A' and Post Hoc Groups for the GWIS Conditions 84

LIST OF TABLES (continued)

Table Page

20. Mean Fuzzy and Crisp Visibility c and Post Hoc Groupings for the GWIS Conditions 84

21. Visibility Response Criterion across Rating 85

22. Ceiling RMSE Data and Post Hoc Groupings across GWIS Conditions 88

23. Comparison of Compensatory and Non-Compensatory Models 93

24. Decision Models with the GWIS and Restricted Range 94

LIST OF FIGURES

Figure Page

1. Jensen's Pilot Judgment Model 9
2. Graphical representation of the signal detection metrics 32
3. Signal detection sensitivity and bias combinations....32
4. Graphical Weather Information System 52
5. Screen shots of the experimental scenario questions 56
6. Application of signal detection to weather 58
7. Fuzzy mapping function for ceiling....60
8. Mapping function for visibility....60
9. Accuracy data across the baseline visibilities....69
10. Accuracy for the different cloud heights across visibility 70
11. Mean comfort across the visibility conditions 71
12. Mean comfort across the ceiling conditions 71
13. Interaction of ceiling and visibility on mean comfort....72
14. Ceiling RMSE data at each of the four ceilings....73
15. Ceiling RMSE data at each of the four visibilities....73
16. Ceiling RMSE data for the ceiling by visibility interaction 74
17. Visibility RMSE across the visibility conditions....75
18. Visibility RMSE across the ceiling conditions 75
19. The interaction of visibility and rating on visibility RMSE....76
20. The interaction of ceiling and visibility on visibility RMSE....77
21. Sensitivity data for the GWIS conditions 79

LIST OF FIGURES (continued)

Figure Page

22. Combined response bias data for the GWIS conditions 80

23. Accuracy data for the GWIS conditions 86

24. Accuracy data for the ceiling and GWIS interaction 87

25. Accuracy data for the GWIS and visibility interaction..87

26. The interaction of GWIS by ceiling by rating on ceiling RMSE..........................89

27. Mean comfort data for the GWIS conditions..90

28. Safety ratings of ceiling adapted from Driskill et al. (1997) 92

29. Safety ratings of visibility adapted from Driskill et al. (1997) 92

INTRODUCTION

General Aviation Flying and Weather

Weather affects piloting more than any other physical factor. An understanding of the current and forecasted weather is a critical aspect of general aviation (GA) flight safety, particularly for non-instrument rated pilots who are restricted to flights under visual flight rules (VFR) conditions. Despite its importance for flight safety, many pilots believe weather is the most difficult and least understood subject in their pilot training (Lankford, 2001). Gathering, analyzing and making decisions based upon weather is ultimately the responsibility of the pilot in command (PIC) (Buck, 1998). Remaining vigilant about weather is particularly important in GA where there is seldom a co-pilot. Forecasts based on computer models and satellite imagery may be outdated or inaccurate. Weather is a dynamic environment, where conditions can change rapidly and unexpectedly. Pilots must constantly appraise and interpret the information available. Failure to recognize deteriorating weather is a potentially fatal error for GA pilots.

Pilots' concern with weather begins when they decide to fly and does not end until they arrive safely at their destination. The first component of pilots' weather decision making is the evaluation of the weather prior to flight. This process can begin days before the flight, particularly for cross-country trips. It is prior to flight that pilots should learn the overall weather systems. The big picture includes knowing what the weather is along the flight path, where fronts are, and what the conditions are expected to be throughout the flight. The pilot has a number of different weather information sources available for consideration. Some information such as radar and satellite imagery is typically available only prior to flight while others such as Flight Watch can be acquired only enroute.

The pilot's first weather decision is the "go/no go" decision. This decision is driven by the weather conditions at the airport at takeoff and the forecasted conditions for the planned flight route. Unlike the decision at takeoff, in-flight weather decision making is a continuous process. It is the pilot's responsibility to be aware of the current weather situation and to see if it matches the forecast. If a forecast is incorrect or "busted" the pilot must be prepared to respond appropriately. However, as evidenced by pilot reports and as indicated by accident reports, weather-related decisions are sometimes incorrect. Examination of the factors associated with inaccurate or risky weather-related decisions is critical to aviation safety.

Weather can be one of the most dangerous variables for GA pilots. Weather related accidents are more likely to be fatal than any other type of GA accident (Goh & Wiegmann, 2001a). According to the 2002 Nall Report (Aircraft Owners and Pilots Association [AOPA] Air Safety Foundation, 2002), weather was a causal factor in 4.1% of all GA accidents; however, these accidents represented 15.2 % of the total fatal GA accidents in 2001. Attempted VFR into instrument meteorological conditions (IMC) represents the largest weather related hazard, accounting for 90% of all fatal GA weather accidents.

Inadvertent flight into IMC while operating under VFR is most hazardous for GA pilots, whose intent is to navigate and maintain separation by visual cues (i.e., not having to rely on instruments). VFR flights are categorized by rules related to cloud base (ceiling) and visibility. Federal Aviation Regulations (FARs) categorize VFR minimums. These minimums differ by airspace class, although generally the ceiling must be 1000 ft or more and the visibility must be at least 3 statute miles (sm). Appendix A provides a breakdown of the different FAR weather categories. FARs prohibit non-instrument rated pilots from flying when conditions are below these VFR minimums.

The fatality rate for VFR into IMC accidents in the U.S. was approximately 80% between 1990 and 1997 (Goh & Wiegmann, 2001a). Although VFR into IMC accidents represented approximately 2.5% of the total GA accidents between 1996 and 2000, they accounted for over 10% of the fatal GA accidents (AOPA Air Safety Foundation, 2001). Earlier studies found similar statistics (AOPA Air Safety Foundation, 1996; Transportation Safety Board, 1990). In Canada, VFR to IMC accidents accounted for only 6% of the total accidents between 1976 and 1985; however, these accidents represented 26% of the total GA fatalities (or 418 persons) for the same time period (Transportation Safety Board, 1990).

The decision to continue into IMC conditions can have severe consequences, particularly when a pilot does not have an instrument rating. Although 48% of pilots hold an instrument rating, approximately 75% of pilots who are involved in VFR into IMC accidents are not instrument rated (AOPA Air Safety Foundation, 1996). The majority of these accidents occurred during the cruise phase of flight. Frequently difficulties arise when a pilot has taken off while conditions were VFR, but then encounters bad weather enroute. Many of the accidents (209 of 580) involve non-instrument pilots continuing or initiating IMC. Another frequent problem within the VFR into IMC accidents (95 of 580) involves flying VFR under an overcast ceiling in areas with rising terrain. Upon entering IMC, pilots typically become disoriented, which can result in a controlled flight into terrain (CFIT). There are a number of reasons pilots may make an improper weather related flight decision. Due to the potentially fatal consequences of an incorrect weather decision, identifying these reasons is imperative.

Several researchers have investigated the factors that influence pilots' decisions to continue a VFR flight into IMC (Goh & Wiegmann, 2001a, 2001b, 2002; Griffin & Rockwell, 1987; Hunter, Martinussen, & Wiggins, 2003; David O'Hare, Owen, & Wiegmann, 2001; David O'Hare & Smitheram, 1995; Wiegmann, Goh, & O'Hare, 2002; Wilson & Fallshore, 2001). Many variables have been identified that may potentially influence the decision to "press on" into IMC. These can be roughly categorized into two groups: poor situation assessment, and improper motivation. Poor situation assessment suggests that the pilots do not recognize that weather conditions have deteriorated. This category corresponds to an inability to accurately assess or diagnose the situation, or inadequate sensitivity in signal detection theory (SDT) terms. Within the context of this study a signal is the presence of instrument conditions. Inadequate situation assessment or sensitivity suggests that had they perceived the information correctly, they would have decided to divert.

The second major classification involves improper motivation and in SDT terms can be associated with the pilot's response criterion and response bias. Improper motivation is evidenced when a pilot recognizes that the situation has deteriorated, but chooses to continue anyway. One factor that may affect a pilot's response bias is a phenomenon referred to as sunk cost (Arkes & Blumer, 1985). The sunk cost phenomenon refers to a desire to continue in an effort to preserve the resources and costs that have already been invested in the task. For example, pilots who are farther along in the flight plan may be more motivated to not "waste" the time invested in the flight and therefore may be motivated to continue on despite adverse weather.

Investigations that provide initial evidence supporting both situation assessment and motivational influences as contributing factors in decision making errors will be discussed further. Previous investigations provide valuable but limited information on the factors involved in continuing a VFR flight into IMC. Before discussing the previous research in any detail, a unifying framework or model that incorporates the various elements of pilot judgment is provided.

Jensen's Pilot Judgment Model

Jensen's Pilot Judgment Model is a dynamic model well suited for examination of the weather-related decision making process (Goh & Wiegmann, 2002). Jensen's model uses a normative foundation that incorporates both sensory/cognitive factors and motivational factors and allows separate metrics of each to be constructed through the use of SDT.

The majority of previous research on VFR into IMC has focused only on single aspects of decision making in isolation. A number of different models and heuristics have been used to describe the decision making process. Weather related decision making is a process characterized by uncertainty (D. O'Hare, 2003). If pilots were fully aware of the information and the outcomes the process would be a simple act of choice. However, pilots do not know exactly what the weather is, what it will be, or what all the consequences of a decision to continue or divert may be. The pilot is therefore tasked with weighing the information available, considering the possible outcomes associated with each choice and selecting the best option. The concept of evaluating the value and likelihood of an outcome is the basis for normative decision making models.

A more recent trend in the literature is a naturalistic decision making approach. Naturalistic decision making focuses on expert decision makers in more "realistic" tasks. The more realistic tasks are those that involve constraints such as time pressure and questionable information. Naturalistic models such as the recognition primed decision model explain how experts are able to use cues from the environment to quickly identify and diagnose a situation (Lipshitz, Klein, Orasanu, & Salas, 2001). Once the initial situation assessment has occurred a decision can be made, often without comparing possible outcomes. Naturalistic decision models have been useful in real world tasks and could be applicable to VFR into IMC decision making. However, the majority of pilots involved in VFR into IMC accidents are the less experienced pilots (AOPA Air Safety Foundation, 1996). Less experienced pilots, and novice operators in general, may lack the sensitivity for accurate situation assessment. Naturalistic decision models are useful when considering expert decision makers. However, normative models are more useful in predicting novice performance (Kaempf & Klein, 1994).

It is therefore necessary to use a normative model that recognizes decision making as an iterative process that involves time constraints and questionable information. It is also important to have a model that incorporates factors external to the situation such as social pressures. Jensen's Pilot Judgment Model is a dynamic model that incorporates a number of different interdependent processes. Although the model has a normative foundation it also allows for external factors such as social pressure to come into the decision making process. The model incorporates the concept of signal detection theory, and benefits from the ability to separate the sensory/cognitive and motivational aspects of decision making into separate metrics. Jensen's model has been used in previous investigations as an organizing framework to provide a better understanding of the VFR into IMC problem (Goh & Wiegmann, 2001b). Jensen's detailed model of decision making includes eight steps (Figure 1).

Problem Vigil. The first stage of the judgment process refers to the constant stage of vigilance a pilot attempts to maintain at all times. Pilots use their senses to look for changes that can affect the safety of flight and their progress. Attention, defined by Jensen as the "mental faculty that controls the subject matter chosen for information processing" is of critical importance in this stage. Pilots are trained to focus their attention on the important safety related aspects of flight. Problems can occur here when a pilot simply does not attend to flight critical information such as weather.

Recognition. The second stage is problem recognition. In this stage, perception and expectancy are two important factors. During the recognition stage the pilot realizes a problem has developed that may affect the safety of flight. As a whole the perceptual system reduces the information in the environment to a more manageable size, allowing pilots to become aware of the different objects in their environment. The second key aspect is expectancy. As pilots gain more experience they learn which patterns of events can be grouped together. These expectancies drive which aspects of the environment pilots may focus on and become increasingly important when there is limited time available.

Problem Diagnosis. The third step is the problem diagnosis. It is here the pilot attempts to discover the nature of the problem. Although this step is more important for mechanical problems it has some application to weather. It is here that the pilot asks what is causing the problem. Knowledge and experience play key roles in this component. For example if pilots have accurate knowledge of the weather systems, when a problem occurs they will be able to properly diagnosis the problem (i.e., a front advancing from a particular location faster then predicted). Accurately diagnosing the situation would allow them to divert in the appropriate direction. However, an improper diagnosis could lead to their decision to continue further into IMC.

Alternative Identification. The fourth element of pilot judgment is to identify a set of possible alternatives that would allow the pilot to solve or avoid the problem. For example, after diagnosing the situation as IMC, they may continue towards the destination, return to the departure airport, select another alternate, or try and fly around the weather.

Risk Assessment. The fifth element is an assessment of the risk involved in each of the identified alternatives. This requires that the pilot estimate the probability of success of each alternative. There are a number of factors that can effect the estimation of risk, including skill of the pilot, amount of fuel, and facilities available at nearby airports.

Background Factor. The sixth factor is the background factor. Jensen (1995) sums this up as "the motivational forces that keep us from following purely rational decisions" (p. 46). These are the non-flight related factors that can have an effect on every decision made by the pilot. These come from a variety of sources such as ego, social pressure, and decision framing.

Decision Making. The pilot applies the previous elements and arrives at his or her decision.

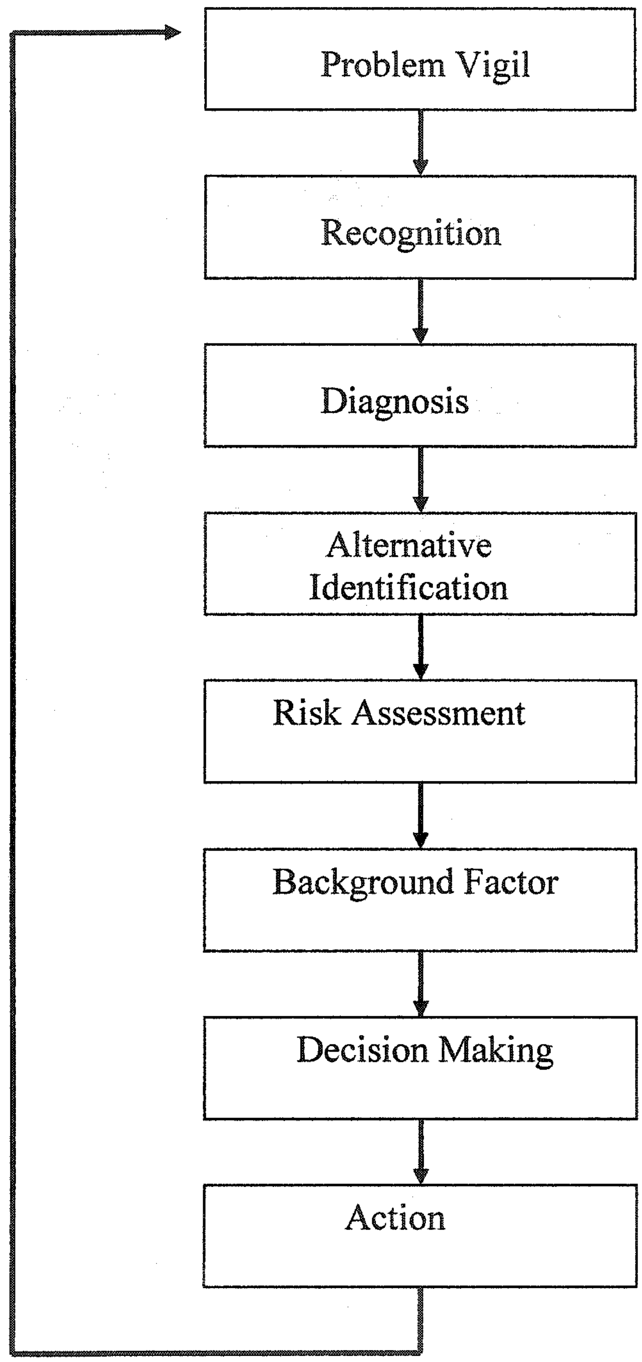


Figure 1. Jensen's Pilot Judgment Model.

Action. The final component is the execution of the pilot's decision.

Jensen's framework provides an important tool for addressing the causal factors associated with pilots' decisions to continue a VFR flight into IMC conditions. Problems can arise in any of the eight stages of Jensen's judgment model. However there is a fundamental difference in problems that arise early in Jensen's framework and those that arise later in the process. Failures in the early stages are the inability to recognize cues, interpret, integrate and diagnose information, and understand the risk associated with each decision. These stages together represent rational judgment. Failures in the first three stages are represented in the situation assessment hypothesis (Goh & Wiegmann, 2001b). Ultimately pilots who make situation assessment errors do not realize they are flying into IMC. However, pilots may make a proper diagnosis of the situation but still be pressured due to background factors to make an incorrect decision. Background or motivational factors would be reflected in changes in response bias. The framework provides a necessary understanding of the judgment process that is needed before reviewing the VFR into IMC literature that emphasizes the different situation assessment and motivation aspects of this process in isolation. The model allows for both situation assessment and motivational components to play a role in the final decision.

Situation Assessment

The situation assessment hypothesis states that pilots continue into IMC because they do not completely understand that the conditions no longer support a VFR flight (Goh & Wiegmann, 2001b). For example the pilot's attention may not have been on the relevant weather cues (problem vigil), there may have been a failure to recognize the changes in weather (recognition stage) or the pilot may have underestimated the severity of the changing weather or where it was developing (diagnosis stage). According to this hypothesis a poor decision would be the result of errors in the early, rational stages of the judgment process, specifically, the stages of problem vigil, recognition, and/or diagnosis. Although alternative identification and risk assessment are both rational components of judgment, the situation assessment hypothesis emphasis focuses primarily on problems that lead to the diagnosis of the situation.

Ultimately the situation assessment hypothesis assumes that if pilots are able to recognize the weather cues or have accurate information indicating they are flying into IMC, they will divert (Wiegmann et al., 2002). If this hypothesis is incorrect, then a pilot's decision to continue into IMC is a willful disregard for the FARs and the weather cues that would have indicated a safer course. Support for the situation assessment hypothesis comes from both accident data (AOPA Air Safety Foundation, 1996) and empirical investigations (Goh & Wiegmann, 2001b; Wiegmann et al., 2002).

In addition to its relationship with the early stages in Jensen's model, the situation assessment hypothesis also has ties to situation awareness (SA). According to Endsley (1995); situation awareness is the portion of the person's knowledge that refers to the state of a dynamic environment. It is the perception of elements, the comprehension of their meaning, and the projection of their status into the future. A simpler definition offered by Vidulich (2003) is that SA is the "momentary understanding of the current situation and its implications" (p. 116). There is considerable debate as to exactly what situation awareness is and how it should be represented as a model (Vidulich, 2003). Some researchers use it as a structural model of information processing (Endsley, 1995) where as others keep more with a "black box" approach, not considering the specific inputs or outputs of SA but instead defining it as an emergent property of information processing (Wickens, 2001).

Some authors have also argued for a distinction between the process and product of situation awareness (Adams, Tenney, & Pew, 1995). The process of SA reflects the various perceptual and cognitive aspects involved in the construction and updating of the state of awareness. The product refers to the state of awareness. However the process and product are interdependent and cyclical. The process creates the state of awareness, which in turn determines what is expected and important and therefore also drives the perceptual processes. This relationship is also a driving mechanism in Jensen's judgment model. Specifically, it is evidenced in the relationship between expectancy and perception in the recognition stage.

Regardless of the mechanisms behind situation assessment and situation awareness, pilots have a mental representation of the dynamic weather situation. This representation includes their model for their current situation and what they expect the weather to be along the flight path. When the pilots' mental model of the weather situation does not match the actual weather conditions, pilots are susceptible to situation assessment errors. Within the framework of Jensen's model, failures in SA can result from errors at the stages of perception of the elements (problem vigil and recognition), comprehension of their meaning (diagnosis) or the projection of their future status (alternative identification and risk assessment). SA is a major input to the decision making process (Endsley, 1995) and faulty SA can lead to an incorrect decision. Simulator studies provide evidence that faulty situation assessment may be the cause of VFR into IMC accidents (Goh & Wiegmann, 2001b; Wiegmann et al., 2002). Additional evidence for the role of SA, particularly as it relates to pilot experience, can be found in accident data (AOPA Air Safety Foundation, 1996; David O'Hare et al., 2001).

Simulator Research. One source of support for the situation assessment hypothesis comes from flight simulation research. Wiegmann and colleagues had pilots of varying experience levels fly a 120 NM cross-country flight. Twenty-five of the 36 pilots had instrument ratings. Pilots were allowed to review weather information from Terminal Area Forecasts (TAFs), meteorological repots (METARs), and winds aloft information prior to takeoff. The destination airport did not support IFR landings thus even instrument rated pilots should not have continued into IMC. All of the pilots took off under VFR conditions. However, these conditions began to deteriorate either at 30 NM (short group) or 90 NM (long group). The experimenters measured the amount of time the pilots remained in IMC conditions before deciding to divert. Only 1 of the 36 pilots attempted to continue, however ultimately this pilot failed after crashing the aircraft simulation. The distance pilots flew into IMC ranged from .91 NM to 13.72 NM. Due to the wide variation of the values, nonparametric statistics were used in their analysis.

Pilots who encountered adverse weather early were more likely to proceed further into the IMC conditions than those who experienced IMC later in the flight (Median = 5.94 vs. Median = 2.75 NM respectively). This corresponds to 2.86 minutes for the early weather group and 1.48 minutes for the late weather group.

Previously the aircraft's location along the flight route had been used as support for a motivational (background) factor known as the sunk cost hypothesis (David O'Hare et al., 2001). This will be discussed in more detail along with other motivational factors later in this paper. Sunk cost suggests that as goals become closer there is a shift in situation framing from gains to losses. As more resources are invested (i.e., time) in a particular goal it is less likely that the goal will be discarded than if fewer resources were invested. However, pilots who encountered the weather earlier (i.e., those that had invested the least amount of resources) were more likely to press on into the deteriorating weather conditions. This contradicts the sunk-cost hypothesis and the anecdotal "get-home-itis."

The results are easily explained using situation assessment as an explanation of weather decision making. Pilots received a weather briefing prior to departure allowing them to create a mental model of the weather system. As pilots take off and begin their flight they are provided with weather cues from out the window that either match their current mental model of the system, or cause them to update their model. Pilots who encountered the adverse conditions earlier had just received their weather briefing and predictions. Because of the recency of the reports pilots may have been more likely to trust their original model and would be more likely to "take a look" to update their model and gain a more accurate weather awareness. Knowledge that the departure airport is close behind and provides a safe haven would also encourage this behavior. The later the weather is encountered in the flight, the older the original information from the report has become, and therefore the less reliable it is. The pilots in the early group had more reason to trust the weather reports due to their recency and therefore were more likely to question their interpretation of the out-the-window conditions. Having older weather data, the group experiencing adverse conditions late in flight was more likely to trust their own interpretation.

This interpretation follows an explanation of VFR into IMC flight described by Smith (2001). Smith uses anchoring and adjustment biases to explain pilots' decisions to continue into IMC. If the initial weather briefing received by the pilot states "good VFR conditions" the pilot will use this information as an anchor. Their interpretation of the weather would be based upon a good weather VFR anchor. However this anchor can negatively influence the pilot's judgment. Pilots encountering bad weather after having a good weather anchor would be more inclined to rate the weather as less severe. In this situation their diagnosis of the situation would be less accurate than pilots without an anchor. Within the framework of Jensen's model the failure has occurred in one of the earlier stages. The failure may be at recognition, because the pilot's expectations are influencing them not to focus on weather since they believe it should be VFR. Alternatively, the pilots may be including the incorrect weather reports in their diagnosis and arriving at an incorrect assessment. Regardless of the precise stage at which the error occurs, it represents a failure on the rational side of the judgment process, and ultimately a situation assessment error.

Wiegmann et al. (2002) also offered another perceptual explanation of their results. Pilots in the short weather group experienced the adverse weather almost immediately after take off. This contrasts with the long group that experienced a long "baseline" of steady weather conditions prior to encountering the adverse weather. It may have been easier for the pilots to detect the change because it would have represented a sudden change from the baseline, whereas in the long weather condition the change in weather would have appeared to be a continuous change.

Stronger support for a situation assessment explanation comes from another simulator study conducted at the University of Illinois (Goh & Wiegmann, 2001b). Thirty-two VFR only pilots first completed a preflight questionnaire to ascertain different motivational biases that may have been present. Following the questionnaire, participants flew a simulated flight. The first leg of the flight was used to familiarize the pilots with the simulator. During the experimental route the pilots flew an 85 NM cross-country flight. Participants were given a map of the route, alternate airports, TAFs and winds aloft information. Weather conditions at takeoff were VFR. However, 45 minutes into the flight conditions deteriorated into below VFR conditions (ceilings were 1500 Mean Sea Level (MSL) and visibility was 2 sm). Participants had a 5-minute window from when the conditions deteriorated below VFR minimums until the experiment was terminated. If a decision was not made to divert during this time participants were considered to have made a decision to continue. Pilots were asked to estimate ceiling, visibility, and distance to the airport at the time the simulation ended.

Of the 32 pilots only 10 made the decision to divert using the experimenter's cut-off. The pilots who decided to divert were compared with those who continued. The most important factor in predicting this dichotomy was the pilot's estimation of visibility at the time the scenario ended. The pilots who decided to continue overestimated the visibility (conditions were worse than the pilots believed). Pilots who decided to divert had a mean error of 0 for visibility estimates compared to a mean error of 1.4 sm for the pilots who decided to continue. Pilots in both groups tended to overestimate the ceilings by about 2200 ft. The study revealed both the importance of situation assessment in predicting VFR flight into IMC and demonstrated pilot's difficulty with interpreting out-the-window weather cues. If pilots make an error in the recognition stage of pilot judgment they would have no reason to consider a need for diverting. These two studies suggest that failures in the recognition and diagnosis stages may be why pilots inappropriately continue into IMC.

In both of the University of Illinois (UI) studies (Goh & Wiegmann, 2001b; Wiegmann et al., 2002) the ceiling and visibility both dropped below VFR minima. In the UI studies either ceiling or visibility could have led to an IMC classification using FARs. Would the pilots' decisions have been the same if only ceiling or visibility fell below VFR minima? Using the guidelines provided by the FARs the answer would be yes. If pilots interpreted one of the two components as being below VFR minima then they should have diverted. However research conducted by the FAA (Driskill et al., 1997; Hunter et al., 2003) suggests that pilots may be incorrectly combining weather information and overestimating weather conditions. How pilots combine weather is an important step in situation assessment, and may be a problem for pilots.

Information Integration and the Diagnosis Stage. The pilot is tasked with evaluating weather as a whole. However, there are different components that play a role in this overall assessment. Ceilings and visibility vary independent of each other. A low ceiling or poor visibility alone can categorize conditions as IMC. Research sponsored by the FAA addressed the question of how pilots mentally combine different ceilings and visibilities (Driskill et al., 1997; Hunter et al., 2003). These FAA studies utilized the same methodology, however Hunter and colleagues looked at samples from several countries.

Pilots were provided with textual information about cloud ceiling, visibility and precipitation. This information was given within 81 textual scenarios (3 sets of 27). Sets were divided by terrain (water, mountainous, and non-mountainous). Each set was comprised of 3 levels of ceiling, visibility, and precipitation (each containing a high, medium, and low). Participants were instructed to rank order the 27 scenarios for each terrain route. After the ranking they were instructed to provide a rating ranging from 1 to 100 to describe their level of comfort with the flight.

Equations were developed for both compensatory and non-compensatory decision models using the comfort rating as the dependent variable and the three different weather attributes as the predictor variables. Within a compensatory decision making framework the positive and negative attributes of each option are considered and the selection is based upon the greatest number of positive attributes. If pilots use this type of strategy, good ceilings would "compensate" for poor visibility. Compensatory models of decision making are an efficient use of the available information, however they are not the optimal decision making strategy. In fact compensatory models can place inexperienced pilots at a greater risk of being involved in a weather accident (Hunter et al., 2003). For example inexperienced pilots cannot allow for good ceilings to compensate for poor visibility. If a non-instrument pilot cannot see where they are flying then they should not be flying. Indeed for conditions to be VFR they must meet both ceiling and visibility requirements stipulated in the FARs.

Non-compensatory models such as the multiple hurdle model offer a safer alternative. This non-compensatory model compares each aspect against a criterion. For example, a pilot would continue a flight only if the ceiling met their ceiling criterion and the visibility met their visibility criterion. When either the ceiling or visibility falls below the pilot's criterion they will opt not to fly. Failure of any aspect to be above the pilot's criteria should result in a decision to abort. FARs provide ceiling and visibility criteria for VFR flight. However, a better strategy would be for pilots to establish personal minimums for flying VMC. Despite the FARs, GA pilots are for the most part unsupervised. This lack of supervision may lead to a weak commitment to follow the FARs. Pilots should have a stronger commitment if they establish a set of their own personal minimums (Jensen, Guilkey, & Hunter, 1998). These personal minimums are generally more conservative than the FARs.

Results from both FAA studies (Driskill et al., 1997; Hunter et al., 2003) indicate that pilots used a compensatory decision making strategy. Models using either the sum or product of the three weather attributes had the largest correlations with the pilot's reported level of comfort. Correlations for these two models were above .8 across three samples from different countries (Hunter et al., 2003). How the pilots integrate the different weather components is the critical component of Jensen's diagnosis stage of judgment. Averaging all of the weather components instead of using the worst condition can result in an improper diagnosis. The evidence from the FAA research suggests that pilots' problems in situation assessment may result from improper diagnosis.

The evidence from information integration (Driskill et al., 1997; Hunter et al., 2003) and recognition of deteriorating conditions (Goh & Wiegmann, 2001b; Wiegmann et al., 2002) all provide direct support for a situation assessment hypothesis. Additional support is provided by research examining the influence of flight experience on VFR into IMC decisions.

Experience. Experience plays a critical role in the early rational stages of Jensen's model, and therefore also in the situation assessment hypothesis. Within the problem vigil stage, more experienced pilots may simply have more resources available to focus their attention on the different safety critical aspects of flight. Experience can also affect the recognition stage of judgment. Particularly, experience influences which cues pilots believe are important and drives their expectations of the situation. Additionally, failures may occur at the diagnosis stage as pilots may simply lack the experience necessary to interpret the real time weather (Wiegmann et al., 2002).

The importance of experience in decision making is illustrated in its defining role in Recognition Primed Decision making (RPD) (Lipshitz et al., 2001). According to RPD the key to decision making is having the relevant experience or knowledge necessary to properly recognize and assess situations. Fair weather pilots may simply lack the experience and knowledge of weather cues to recognize IMC conditions. Indeed 75% of the pilots involved in VFR into IMC accidents did not have instrument ratings (AOPA Air Safety Foundation, 1996). This is a substantial overrepresentation considering that non-instrument rated pilots account for only half of the GA population.

Data from accident reports (David O'Hare et al., 2001) and VFR to IMC simulations (Wiegmann et al., 2002) have demonstrated an important relationship between pilot experience (based upon cross country hours) and an inappropriate decision to continue. The more experienced pilots were, the more likely they were to divert when they encountered bad weather. This suggests that experienced pilots can more accurately recognize the cues of deteriorating weather conditions and/or make better decisions regarding the continuation of the flight. Survey data reveal that expert pilots (over 1,000 cross country hours) rely on different weather cues than novice pilots when making decisions to continue a flight (Wiggins & O'Hare, 2003). Inexperienced pilots may not understand the importance of the different weather cues and therefore ignore them when assessing the weather situation.

Wiggins and O'Hare (2003) compared the ability of experts and novices to determine if they could remain in VMC conditions. Participants were classified as experts and novices based on their cross-country hours. Experts were defined as those who had accumulated over 1000 cross-country hours. The study was conducted on the web using static in-flight images. Pilots rated the importance of nine different cues for determining whether remaining in VMC was possible. The nine different cues were determined after interviewing a group of experts. These cues are presented below in Table 1. Experts rated horizontal visibility and increasing cloud concentration as more important than novices, whereas novices rated wind strength as more important than experts.

TABLE 1: Nine Weather Cues Identified by Wiggins and O'Hare (2003)
A change in the type of cloud formation
An increase in cloud density (concentration)
A darkening of the cloud
A lack of adequate terrain clearance
A lowering cloud base
Rain showers
A change in wind direction
A change in wind speed
A reduction in horizontal visibility (loss of horizon)

In the final section, pilots viewed ten static out-the-window images. After viewing the images the pilots were asked to state whether it was possible to remain in

VMC on their current track and altitude. They then reported the confidence in their answer and reported what cues they used to make their decision. Experts more frequently used cloud type, cloud base and cloud concentration during their assessment process. Pilots who judged it possible to remain in VMC more frequently relied upon cloud base and cloud type to make their decisions.

Experts (i.e., pilots with more than 1000 cross country hours) were more confident in their decisions than novice pilots. Pilots who decided to remain in VMC were more confident with their decision than those who believed VMC flight was no longer possible. The appropriateness of the pilots' decisions was not discussed within the context of their paper. The authors appeared to make the assumption that the decisions made by the experts were correct. However it is apparent that expert and novice pilots weight different weather cues differently. The cues that pilots use could influence not only what aspects of weather they pay attention to and perceive, but also affect what information they use to diagnose the situation.

Wiggins and O'Hare (1995) used multiple flight decision scenarios to compare experts (pilots' with over 1,000 cross country hours), intermediates (pilots' with 101 – 1,000 cross country hours), and novices (pilots' with less than 101 cross country hours). Participants were given an image of the aircraft and its position along the flight path. Pilots had access to data screens containing aircraft information (current aircraft state and performance), weather information, and terrain and airport information. The analysis revealed that expert pilots used fewer information screens and returned to the same information screen less often than novice pilots. Pilots with less experience were more likely to return to their original point of departure; which due to terrain, weather and fuel limitations represented an incorrect decision. Expert pilots were more likely to make the correct decision and continue to their original destination. However this evidence is indirect. Expert decisions to continue on with the flight may have actually been a result of experts having a greater confidence in their abilities. It is unclear if experts would be more likely to divert when conditions become IMC.

Utilizing a full mission flight scenario Wiegmann, Goh, and O'Hare (2002) found that experience, particularly recent flight experience (within the last 90 days), was negatively correlated with the distance pilots traveled into IMC conditions. More experienced pilots recognized the deteriorating conditions and decided to divert earlier than less experienced pilots. Recent flight experience had a stronger impact on the distance traveled into IMC conditions when the conditions were experienced later in flight. Although total flight hours, solo hours, IFR hours and cross country hours were also negatively correlated with distance flown, none of these measures reached significance.

Although the evidence to support the role of experience in weather decision making has been demonstrated, it is still not clear which components of situation assessment expertise impacts. In the Wiegmann, Goh and O'Hare (2002) study, although flight experience was a significant variable for flight decisions, there was no significant relationship between experience and the ability to estimate ceiling or visibility conditions. Previous research had identified the ability to estimate visibility as the best predictor of the pilots' decision to divert (Goh & Wiegmann, 2001b). However, the role of expertise was not considered in this investigation. A strong link between expertise and situation assessment in weather decision making has yet to be established. Experienced pilots may still have difficulties with sensitivity, or accurately recognizing weather information.

Situation assessment is not the only explanation for VFR into IMC accidents. Indeed a pilot may have a perfect mental representation of the weather system and still continue into IMC. The alternative explanation is that the pilots are influenced by motivational factors not accounted for by the current weather situation. Jensen's model classifies motivational influences as background factors that take place after an assessment of the situation has been made.

Motivational Judgment

Many motivational factors have been investigated within the VFR into IMC literature. A popular motivational explanation is that some pilots may have an attitude of invulnerability. On the average, GA pilots are overconfident in their abilities and do not fully appreciate the risks associated with weather (Wilson & Fallshore, 2001). Other motivational factors include how the decision is framed (David O'Hare & Smitheram, 1995), how the pilots rate their abilities (Goh & Wiegmann, 2001b; Wilson & Fallshore, 2001), how hazardous pilots believe weather to be (Goh & Wiegmann, 2001b), and potential influence from passengers (Goh & Wiegmann, 2002).

Decision Framing. O'Hare and Smitheram (1995) conducted a laboratory study investigating how prospect theory could be applied to weather decision making. The most notable feature of prospect theory is that people make decisions differently based upon risks involving perceived losses versus perceived gains (Kahneman & Tversky, 1984). According to prospect theory, people view outcomes not as end states or total assets, but instead view outcomes in terms of gains or losses from a reference point. A value function for losses and gains shows a different value being placed upon losses and gains of equal amounts. The function is considerably steeper for losses. As a result, people are more risk averse when considering gains; however they become more risk seeking when considering losses. For example, over 80% of people surveyed will choose a sure gain of $240 over a gamble involving a 25% chance to win $1000 and a 75% chance to win nothing (Kahneman & Tversky, 1984). This is a clear demonstration of risk aversion when considering gains. However, only 13% of people surveyed chose a sure loss of $750 over a gamble involving a 75% chance to lose $1000 and a 25% chance to lose nothing. This in turn demonstrates people's tendency to be risk seeking when considering losses.

Manipulating the reference point from which decisions are evaluated may change preferences for the same option. The manipulation of the reference point is known as decision framing. With respect to flight, pilots typically set their reference point as either the departure airport or their current position. The selection of the reference point may impact their decisions (David O'Hare & Smitheram, 1995).

Prospect theory has several implications for pilots' choices in weather decision making. Pilots who frame the decision of diverting in questionable weather conditions as potential losses (i.e., increased fuel consumption, delays, and increased cost) could be expected to be more likely to risk continuing into IMC. Whereas, pilots who frame the decision to divert as potential gains (i.e., safety of passenger and aircraft, maintenance of an untarnished flight record) should be more likely to divert when encountering IMC.

O'Hare and Smitheram (1995) tested prospect theory in GA weather decision making scenarios. Their focus was on the critical role of the reference point in how pilots will frame the decision. Pilots were placed in one of two groups and given identical weather scenarios. The scenarios gave pilots access to information regarding weather (textual information), airports, topographic maps, and the current state of the aircraft. Pilots reviewed this information at several points along the flight path and made the decision to continue or divert. Weather conditions along the flight path deteriorated to Marginal VFR. Information for one group was worded to emphasize the departure airport as a reference point and reminded them of the time and money invested so far. Information for the other group used the current position as the reference point thus encouraging pilots to ignore these previous "losses."

The decision framing manipulation had a significant impact on pilots' decisions. When the decision to continue or divert was framed as gains using the current position as the reference point 75% of the pilots decided to divert. Only 33% of pilots from the group that used the departure airport as the reference decided to divert. The risk seeking decision to continue in the latter group is attributed to the framing of their decision in terms of losses.

Sunk Cost. The sunk cost phenomenon is another potential motivating factor in the decision to continue into IMC. This effect is demonstrated in the tendency for people to continue with an endeavor after an investment has been made (Arkes & Blumer, 1985). This investment can be anything from time to money. The effect is attributed to the desire not to appear wasteful. An analysis of GA accident data from New Zealand from the 1988 to 2000 (David O'Hare et al., 2001) provides some evidence that the sunk cost effect may play a role in weather related decision making. The data showed that weather accidents occurred significantly further from the departure airport than other types of GA accidents, such as mechanical failures. This finding supports the sunk cost hypothesis. Pilots who had invested more time and money into the flight were less likely to abandon their original goal or destination.

Self-judgment. Goh and Wiegmann (2001b) investigated several motivational factors, testing 32 non-instrument rated pilots in a simulated weather decision making flight. Pilots were given a modified version of the Aeronautical Risk Judgment Questionnaire (David O'Hare, 1990). The questionnaire contained items regarding pilots' background, self judgment, hazard awareness, and risk awareness. The self-judgment questions used a 7-point Likert scale to assess pilots own skill compared to others, their willingness to take risks, and the frequency in which they take risks. Hazard awareness questions asked the pilots to estimate the percentage of accidents due to six broad causal factors (e.g., weather, pilot error). They also ranked seven specific factors (e.g., fatigue, and flying into adverse weather) on how likely they would be to cause an accident. The participants answered the question in two ways, one was how likely the factors would be in general to contribute to an accident, and the other was how likely the factor would be to contribute to an accident in which they might be involved.

Self-ratings indicated that the pilots who opted to continue into IMC had significantly higher ratings of skill and were more likely to take risks. This suggests that greater confidence in their piloting abilities led these pilots to be more willing to risk a flight into IMC conditions. Questionnaire data from other investigators also suggests that pilots are subject to this same optimism bias and an ability bias (Wichman & Ball, 1983; Wilson & Fallshore, 2001).

A survey of 57 student pilots and 103 GA pilots revealed that pilots believe they are less likely to inadvertently fly into IMC, and more likely to successfully fly out of IMC than other pilots of equal experience (Wilson & Fallshore, 2001). Unfortunately the survey did not provide data on the actual experience levels of the participants.

In addition to an assessment of their own skills, decisions made under uncertainty involve the perception of risk. Pilots must be able to determine the likelihood of suffering a loss. Pilots in the Goh and Wiegmann study (2001b) who continued the flight believed themselves to be less likely to be involved in an accident than those in the divert group. According to these biases, pilots underestimate the risk associated with a hazard and overestimate their abilities to overcome a hazard. In combination, these biases can lead to poor decisions, particularly on the part of non-instrument rated pilots when facing instrument conditions. Although the assessment of risk is part of the rational judgment process, the combination of overestimating skill and underestimating risk is demonstrative of the hazardous attitude of invulnerability some pilots may have.

Social pressure. An investigation of NTSB accident data from January 1990 through December 1997 revealed several differences between VFR into IMC accidents and a random selection of other GA accidents (Goh & Wiegmann, 2002). VFR into IMC accidents totaled 409 during the reviewed period. Results of the analysis indicate that social pressure may have been a factor in the decision to continue. Compared to the random selection of accidents, VFR into IMC accidents were more likely to involve passengers. The addition of passengers may have caused the pilots to try and impress their passengers, or the passengers may have pushed for the pilot to continue to the original destination.

Despite support for both situation assessment and motivation in pilot judgment there has been no systematic attempt to separate the relative contribution of each. Pilot judgment is a process of collecting and weighing the available evidence. Once the evidence has been evaluated regarding weather conditions the pilot is left with two possible choices, conditions are IMC or VMC. The decision faced by the pilot is similar to many other types of diagnostic decisions that have benefited from SDT. Further, identifying the relative influence of situation assessment and motivational factors will assist in the development of strategies aimed at improving weather-related decision making and overall aviation safety.

Signal Detection Theory

Jensen's full model of pilot judgment provides a valuable tool for decomposing the different stages of pilot judgment. However, the principal advantage of the model is that it can be reduced into a two-factor model based upon the metrics of SDT. These two factors are rational judgment, which encompass the stages from problem vigil through risk assessment, and motivational judgment, which represents the background factor. Within the SDT framework sensitivity is the metric for accuracy of the diagnosis. There are both parametric (i.e., d') and non-parametric (i.e., A') metrics for sensitivity. The response criterion, which is representative of directional bias, also has a parametric (i.e., β) and non-parametric (i.e., c) metric. The parametric statistics are used when the signal and noise distributions are both normal and have equal variance. If the assumptions were not met d' would vary with response criterion and therefore the non-parametric A' would be applied (Stanislaw & Todorov, 1999).

Crisp Signal Detection Theory. Signal detection theory was originally used in psychology as a psychophysical tool for characterizing human performance in detecting weak signals (Green & Swets, 1966). The theory has been used to explain both how accurate an operator is and what type of bias is present. Since its inception in psychology the theory has seen widespread use, particularly in diagnostic fields (Swets, Dawes, & Monahan, 2000). In its traditional, or "crisp" form it has been applied to situations where there are two distinct states of the world (i.e., a signal is either present or it is not).

In terms of weather decision making, a signal can be thought of as the presence of IMC or the absence of IMC, (i.e., VMC). To each of these two states of the world the observer can make two distinct responses; either a signal is present (IMC) or it is not (VMC). Based on the two states of the world and the two responses there are four possible outcomes (these are provided in Table 2 below). SDT assumes that there are two stages to information processing. First, evidence is collected regarding the presence or absence of the signal. Second, the evidence is compared to a cutoff or criterion and used to make a decision.

TABLE 2: Weather Response Categorization for Crisp SDT

	Actual Weather Conditions	
Pilot's Response	IMC (target)	VMC (noise)
IMC	Hit	False Alarm
VMC	Miss	Correct Rejection

The advantage of SDT is that it allows for the separation of sensitivity and response bias. The sensitivity of the system represents the ability to distinguish between signal and noise. The greater the sensitivity the more accurate the system or operator will be in distinguishing between the two. The sensitivity metric is calculated from the *z* transformation of both the hit rate (HR) and false alarm (FA) rate. Different hit -false alarm rate pairs can generate the same sensitivity. The second component is the response criterion. This component is involved in the decision making stage. When two different hit-false alarm pairs produce the same sensitivity they differ in the amount of response bias. The criterion represents how a stimulus that is not clearly classified as signal or noise will be classified. An individual can have a liberal bias such that they require less evidence to believe a signal was present. As a result they will detect the signal more often (i.e., more hits), but also demonstrate a tendency to respond positively when no signal is present (i.e., more false alarms). Alternatively a conservative bias would result in a reduction in both hits and false alarms. The criterion is a measure of the bias relative to the halfway point between the means of the hit rate and false alarm rate. The different SDT metrics are graphically depicted in Figure 2 and variations of sensitivity and bias are provided in Figure 3. In terms of weather decisions the response criterion will distinguish between pilots who are more likely to consider questionable conditions IMC and those who consider the same conditions VMC.

Fuzzy Signal Detection. The original or crisp SDT is based upon classical set theory. Conditions either are IMC or they are not. FARs provide a clear binary categorization of weather conditions as either VFR or IMC. However despite this categorization real world signals can be fuzzy (Parasuraman, Masalonis, & Hancock, 2000). According to fuzzy logic, an event can exist somewhere between one state and the other.

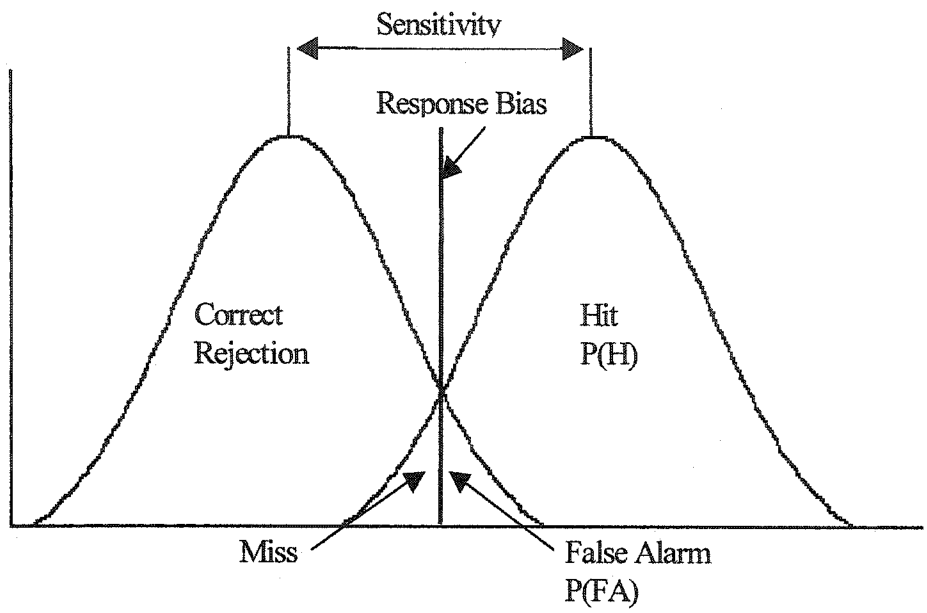


Figure 2. Graphical representation of the signal detection metrics.

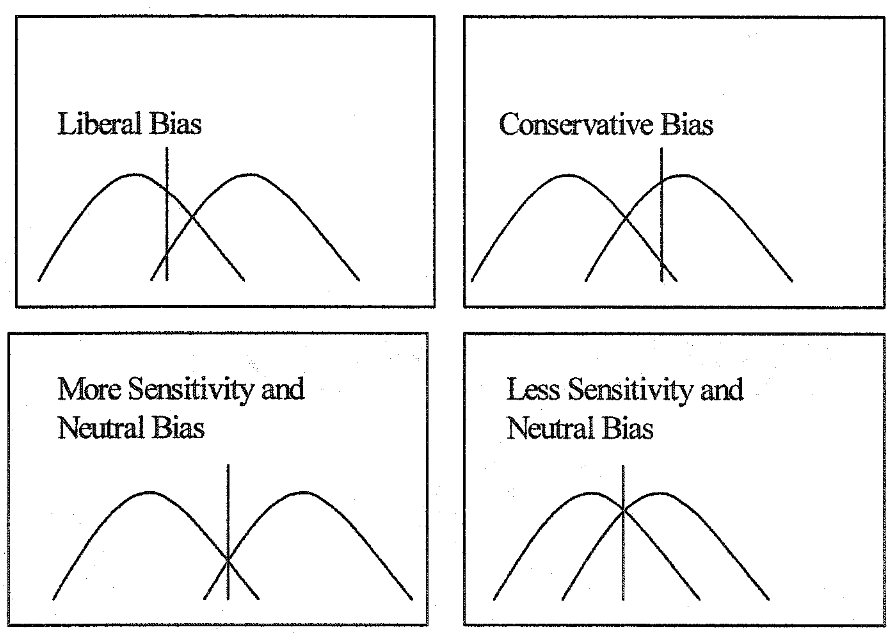


Figure 3. Signal detection sensitivity and bias combinations.

Fuzzy signal detection theory represents a useful extension of signal detection theory using fuzzy logic. Instead of the binary classification of signals used in crisp SDT, fuzzy SDT allows the signal to take on an infinite range of values. Because an event can lie somewhere in between, forcing a classification of the event into one of two categories can result in loss of important information. The advantage to fuzzy SDT is that it improves the precision of crisp SDT by "retaining the information provided by the middle ground, rather than by rounding it into oblivion" (Parasuraman et al., 2000).

Weather conditions have a binary classification as either VFR or IMC conditions based upon FARs. However ceiling and visibility clearly vary along a continuum. This variation of weather along a continuous dimension makes it amenable to fuzzy SDT. The first step in using fuzzy SDT is to create a fuzzy mapping of signal strength to the state of the world. Although it can be argued that a fuzzy mapping function is arbitrarily created, the same argument can be applied to the categorization point used in crisp SDT. Once a fuzzy mapping function has been created, fuzzy SDT works essentially the same as crisp SDT. The major difference after the mappings have been created is that an individual's response can fall into multiple cells.

Fuzzy SDT has clear applicability to real world contexts such as weather decision making. However, results obtained using a fuzzy analysis can be quite different than those using a crisp analysis (Masalonis & Parasuraman, 2003). Masalonis and Parasuraman (2003) reanalyzed data sets from two previous studies using both crisp and fuzzy SDT. The reanalysis involved conflict detection in air traffic control (ATC). Similar to weather an ATC conflict has a legal definition in the FARs, however despite this set point conflicts can vary along a continuum. The legal definition of a conflict is two aircraft separated by less than 5 miles. Thus a mapping function of aircraft separation and signal was created. Analyses were then done with both crisp and fuzzy SDT.

The results of analyses indicated that the fuzzy SDT reduced both the hit rate and false alarm rate. This had different effects across the two studies. The first study reanalyzed data from an ATC conflict detection system (machine). The hit rate dropped from .84 to .75 after using a fuzzy analysis. However, false alarm rate only dropped from .0052 to .0022. This resulted in the fuzzy analysis yielding a lower sensitivity. Both the fuzzy and crisp analysis revealed a liberal response criterion (participants were more willing to respond that a conflict was present resulting in higher hit rates and higher false alarm rates). The fuzzy analysis shifted the criterion so that it was less liberal (i.e., more conservative).

The second data reanalysis investigated differences in automation-aided controller conflict detection using data collected from four controllers. The same decrease in P(H) and P(FA) was found as in the first analysis. However the decrease in the P(H) was typically smaller than the decrease in P(FA) resulting in higher sensitivities when using a fuzzy SDT analysis. The decrease in P(H) and P(FA) in a fuzzy analysis also results in a more conservative response criterion. The authors attributed this to the assignment of slightly lower signal strengths to more severe conflicts.

Fuzzy SDT may indeed be a more realistic representation of sensitivity. However, it may be at the cost of the response criterion. When using SDT as a tool for decomposing diagnostic decisions it is typically assumed that evidence is ambiguous. Swets et al. (2000) (p 2) point out that we should consider "the degree of evidence as being represented by a value along a single dimension with high values tending to be associated with the positive diagnostic alternative and low values tending to be associated with the negative alternative." The response criterion or decision threshold is then associated with the amount of evidence necessary to make either a positive or negative decision. It is actually the variability of the evidence or "fuzziness" of the signal that allows the response criterion to take on meaning.

It becomes unclear as to what meaning the response criterion takes on in fuzzy SDT. If the operator is allowed to specify where on a continuum a signal falls then there is no cutoff for what they say is a signal and a non signal. Additionally, in crisp SDT using the operators' report of confidence has been used to generate multiple points (i.e., multiple response criterions) on the Receiver Operator Characteristic (ROC) curve. However, Hancock, Masalonis, and Parasuraman (2000) suggest that the operator's degree of confidence could be used as a fuzzy response. In effect fuzzy SDT may be adding precision to the sensitivity metric at the price of removing some of the variability found in the response criterion.

The use of both fuzzy and crisp analysis allows a more complete understanding of decision making. Masalonis and Parasuraman (2003) suggest that fuzzy SDT be used to compliment crisp SDT and not to replace it, particularly in exploratory analyses. Although fuzzy SDT may be a more accurate representation of operator sensitivity it is likely that it does so at the expense of a meaningful response criterion. The present study will compare pilots using both SDT methodologies.

SDT has a long standing history in psychological research particularly with respect to diagnostic decisions (Swets et al., 2000). However, due to methodological constraints previous researchers have not used it as a tool in VFR into IMC decision making. The studies conducted at the University of Illinois (Goh & Wiegmann, 2001b; Wiegmann et al., 2002) were full flight simulations that allowed pilots to diagnose the conditions at only one point in the entire experiment. One of the main objectives of the present study is to use SDT as a tool to measure the relative contribution of situation assessment and motivational judgment in VFR into IMC decision making. SDT will ultimately lead to a better understanding of why pilots take a VFR flight into IMC. In addition to the ability to decompose pilot judgment based solely upon their skills in interpreting out the window conditions, SDT can also be valuable for investigating how graphical weather information systems impact pilot decisions.

Graphical Weather Information Systems

Current weather decision making research provides an incomplete picture. Decisions are not based only upon weather forecasts and the "out-the-window" view. Pilots also have the ability to access different weather information sources enroute. A number of different weather services can be requested via the radio. Some GA aircraft also include some single sensor based weather avionics to detect a single weather hazard (e.g. Stormscope ™, Strike Finder™). However, these systems typically have a limited range.

If pilots are aware of changing weather conditions they can use these different automated sources to aid in their diagnosis, or use the sources in advance to determine if conditions are following the forecast.

Reliance on auditory information for acquiring weather information enroute contains a number of problems. One of the biggest drawbacks is that enroute services such as Flight Watch (FW) become congested precisely when they are needed most. There are only a limited number of specialists at FW and when conditions change, particularly when they change unexpectedly, pilots will typically turn to this service at the same time. Automated systems such as Automated Weather Observing System (AWOS) and Automated Surface Observation System (ASOS) provide only a static picture of specific airports and paint only a limited picture of the weather situation to the pilot. Because of the limited nature of the available systems pilots may not be able to use them to their full advantage in the diagnosis of the situation.

New technology has made it possible to provide pilots with data linked weather information enroute. This information can be conveyed through a graphical weather information system (GWIS) that provides pilots with aircraft location on a moving map display as well as the location of weather with respect to the aircraft. A number of these systems have been developed all of which are capable of providing graphical METAR and NEXRAD data. These products provide surface weather observations and precipitation information.

To date, few studies have examined graphical weather displays and pilot decision making. Some initial research conducted at Massachusetts Institute of Technology's (MIT) Lincoln Laboratory (Lind, Dershowitz, & Bussolari, 1994) investigated the effect NEXRAD data presented on a GWIS systems had on pilots' decisions. The MIT study utilized twenty instrument rated pilots of two different experience levels (moderate experience 40-150 instrument flight hours, and extensive experience over 150 instrument flight hours). There were 10 pilots in each of the experience groups. Participants completed 4 test flight scenarios with half of the scenarios being completed with a GWIS display.

The methodology utilized in the experiment was similar to previous research conducted at Ohio State (Griffin & Rockwell, 1987; Layton & McCoy, 1989; Potter, Rockwell, & McCoy, 1989). Pilots were presented with a verbal description of the weather situation at three points along the flight route. After the description pilots were asked what they would do next. Pilots could make calls to FW to obtain weather information if desired. All four of the scenarios could have been completed. Subjects in the GWIS group were provided graphical precipitation data. These data could be requested and presented at various ranges.

There were significant differences between the decisions made with and without a GWIS display. When using the GWIS displays, pilots made significantly fewer calls to ATC and FW for weather information. Consistent with other GWIS studies (Latorella & Chamberlain, 2002; Novacek, Burgess, Heck, & Stokes, 2001; Yuchnovicz, Novacek, Burgess, Heck, & Stokes, 2001) pilots in the MIT study were significantly more confident in their decisions when they had a GWIS display available. Decisions made with the GWIS display were such that in certain scenarios pilots were more likely to continue and in other scenarios pilots were more likely to divert. Overall the GWIS display provided a better global understanding of the weather situation. Pilots could see where weather was localized and make decisions about how to divert based on this information. In one scenario, half of the pilots without the GWIS display decided not to take off because of predicted thunderstorms. All of the pilots with the GWIS display decided to take off because based on the display it was apparent that none of these storms had manifested. In another scenario pilots with the GWIS display saw that an area of precipitation was localized and 80% made the decision to divert around the weather as opposed to 40% of the pilots without the display.

In an exit interview, pilots were asked how they would like to receive weather information in flight. None of the pilots responded that they would prefer only voice information, where as 85% preferred a combination of both. The remaining 15% claimed they would prefer only the graphical information. Of the 85% preferring a combination of both, 55% wanted an even split and 30% wanted a split with more information being obtained from a graphical display. This investigation demonstrated that a GWIS display could provide IFR pilots with a beneficial tool for avoiding weather and making more informed decisions. The MIT study also demonstrates pilots' willingness to trust and use GWIS information.

GWISs and Convective Weather. Other research on GWIS displays has focused on decision making regarding convective activity (Beringer & Ball, 2003; Chamberlain & Latorella, 2001; Latorella & Chamberlain, 2002; Novacek et al., 2001; Yuchnovicz et al., 2001) and workload associated with graphical displays.

Researchers at RTI International (Novacek et al., 2001) and the FAA (Beringer & Ball, 2003) independently investigated the impact of a graphical weather display on instrument rated pilots' decision making around areas of convective activity in a simulated flight. In the RTI study, pilots flew a mission from Newport News-Williamsburg International Airport stopping at Richmond International Airport and then continuing on to NASA-Wallops Flight Facility. Pilots were divided into groups of 12 with one group using a high resolution (4x4 km cells) NEXRAD image and the other group having a lower resolution (8x8 km cells) NEXRAD image. NEXRAD images provide pilots with graphical information about precipitation. There were six levels of precipitation graphically displayed in the NEXRAD image (the colors ranged from light green to magenta, corresponding to changes from light to severe precipitation). Due to technological limitations this information was always at least six minutes old when it reaches the aircraft. Thus it was not a real time depiction of the current weather situation. All of the pilots in the experiment had a graphical weather display available to them in addition to normal enroute sources of weather (ATC, FW, ASOS, and ATIS). The graphical weather display used in the experiment provided NEXRAD images, textual METARs and graphical METAR information (ceiling and visibility categories).

One objective of the experiment was to determine the propensity for pilots to misuse a GWIS display. The proper use of a GWIS display is to use the information in the display "strategically." That is, to use the information to plan routes around areas of potential hazards. An inappropriate use of a GWIS display would be to use the information "tactically." Pilots may try to use the information as if it were a real time depiction of the current weather information and possibly attempt to navigate through areas of bad weather. Because of the time delay present in the current system tactical use of the information is inappropriate.

Pilots should have abandoned the approach at Richmond because of potential hazardous convective activity and altered the course to Wallops Island to avoid that activity. Pilots' decisions were judged on a 4-point scale. With respect to the Richmond decision, judgments were made using pilots' decisions to continue the approach and their ability to remain at least 5 NM away from hazardous weather. Before reaching Wallops the GWIS showed pilots two thunderstorms. One storm was north and one storm was south of the pilot's most direct route. The NEXRAD displayed an "enticing corridor that tempted the pilots to fly between them" (p 29). A good decision was one in which the pilot diverted south to avoid the thunderstorms. A poor decision was one in which the pilot attempted to navigate through them.

Findings from the RTI experiment indicated the potential benefits of GWIS displays and also provided some recommendations for their design. The data suggested that the lower resolution NEXRAD images led to better decisions. These results were also reproduced in an experiment conducted by the FAA (Beringer & Ball, 2003). Five of the twelve pilots with the high-resolution display made good decisions at Richmond compared with nine of twelve pilots in the low-resolution display. There are two potential explanations for this. One is that the lower resolution display covered a larger area on the display and thus provided a more salient warning cue. A second explanation is that the lower resolution led to greater uncertainty and thus more cautious behavior on the part of the pilots. The FAA study found a benefit for the low resolution NEXRAD display compared to pilots without any display.

Research conducted at NASA Langley indicated that pilots who had a GWIS had better weather SA than pilots who only had conventional auditory information (HIWAS, FW, and ATC) or information from out the window. SA was demonstrated in better detection of convective weather cells as well as better estimates of distance to those cells.

The research available concerning GWIS displays still leaves many unanswered questions. Most importantly, how will having a display impact pilot's decision making regarding continuing a flight into IMC? Ideally the display provides pilots with the information they need to make the most appropriate decision regarding continuing the flight or diverting from the original course. However, other studies using GWIS displays (Beringer & Ball, 2003; Novacek et al., 2001) as well as research with GPS (David O'Hare et al., 2001) suggests that the information may be used inappropriately and may actually lead to poorer decisions.

The majority of GWIS research has investigated the use of NEXRAD (precipitation) data (Beringer & Ball, 2003; Chamberlain & Latorella, 2001; Latorella & Chamberlain, 2002; Novacek et al., 2001). This information can be between 12 and 14 minutes old before it is updated (Beringer & Ball, 2003). METAR information on the other hand is only updated every hour. During this time frame weather can obviously change for better or worse. It is not certain if pilots will use this potentially old information when evaluating and diagnosing weather conditions. It is critical to understand how pilots use GWIS information in their weather judgments before these displays are widely used in GA.

PRESENT STUDY

Previous research suggests that both situation assessment (Goh & Wiegmann, 2001b; Wiegmann et al., 2002) and motivational factors (David O'Hare & Smitheram, 1995) affect pilot judgment. Unfortunately these factors are typically investigated separately and in such a way that the relative contribution of each is not understood. Following Jensen's model of judgment it is clear that both the situation assessment and motivation can affect pilot judgment within the same situation. However, without a tool such as SDT to measure both factors it is impossible to understand how each impacts the pilot's decision.

The primary objective of the proposed investigation was to model how pilots classify weather categories following Jensen's ADM which utilizes a crisp SDT analysis. Additionally, the proposed study examined the discrimination ability of fuzzy SDT, relative to crisp SDT, in an effort to further increase understanding of pilot's ability to accurately assess weather conditions. Weather-related decision making appears particularly suited to analysis with fuzzy SDT methodologies. Although fuzzy SDT is not meant to replace crisp SDT it can provide some additional insight into the pilot's sensitivity (Masalonis & Parasuraman, 2003). The SDT approaches allowed for the separation of rational (i.e., situation assessment) and motivational biases. Both rational judgment and motivational bias have been shown to influence pilots' decisions and ultimately their classification of the weather conditions. Identifying the extent to which each component influences the overall judgment is a critical step in addressing the VFR into IMC problem and evaluating new GWIS technology.

Using SDT the principal aim of the current experiment was to determine how well GA pilots could assess weather based on an "out-the-window" visual. Anecdotally, pilots cite difficulty with accurately estimating distances or visibility while in flight (Boatman, 2001). GA pilots, specifically non-instrument rated pilots, may lack the necessary weather experience to accurately assess weather cues available out-the-window. Although empirical evidence is limited (Goh & Wiegmann, 2001b), pilots appear to be very poor at estimating ceiling and visibility based upon the out-the-window conditions. This inaccurate situation assessment may cause a pilot to inadvertently continue into IMC.

Results from previous VFR into IMC research have demonstrated a relationship between pilot experience and pilot judgment (Wiegmann et al., 2002; Wiggins & O'Hare, 1995). Additionally it has been found that non-instrument rated pilots are more likely to be involved in VFR into IMC accidents (AOPA Air Safety Foundation, 1996). Despite a link between experience and pilots weather judgment, no relationship between experience and the ability to estimate weather conditions has been found. Using SDT, the present study therefore compared instrument and non-instrument rated pilots in both their ability to distinguish VFR and IMC conditions as well as identify any difference in bias between the two groups.

The second major objective of the present study was to examine the weather-related decision making strategies used by pilots to determine if pilots use compensatory or non-compensatory strategies. Previous work, sponsored by the FAA, investigated how ceiling and visibility information is combined to make a decision regarding the pilot's comfort in continuing a flight (Driskill et al., 1997; Hunter et al., 2003). The results of these studies revealed pilots used a compensatory strategy when integrating the information. Specifically an average of the two pieces of information was used when making judgments about their comfort in continuing the flight. This strategy is inappropriate, and can lead to an incorrect situation assessment. For example if the ceilings are below VFR minimums but visibility is unlimited pilots should not consider VFR flight. Pilots' judgments should use the worst of the two conditions (ceiling and visibility) when making judgments about continuing under VFR, and not a combination of the two. How pilots use the two conditions to form a diagnosis is an essential component of understanding situation assessment. However the results of these FAA studies are limited because pilots were provided with a textual representation of the conditions without the opportunity to view graphical scenes resembling what they would see out the window. Pilots' strategies for integrating the information may differ if the ceiling and visibility must be derived from a single image. It is not clear if a compensatory strategy will be maintained if pilots have to make judgments based upon the out the window view. The determination of the most used decision making strategy followed the analysis used by Driskill et al. (1997) and Hunter et al. (2003).

The third objective was to evaluate how GWIS information is used. These systems are ultimately being designed to provide pilots with an image of the weather with respect to their aircraft. Several studies have looked at the use of a GWIS in weather decision making with respect to convective weather (Beringer & Ball, 2003; Yuchnovicz et al., 2001). However, it still remains unclear as to how the GWIS will affect pilots' categorization of conditions as IMC or VMC. Surface weather observations, specifically graphical METARs, display airport ceiling and visibility information along four categories (LIFR, IFR, MVFR, VFR). However, it is unknown if pilots will generalize this surface information to the area around the airport. This information is specific to the airport and is only updated every hour. Since the information has the potential for being old, data displayed on the GWIS may indicate weather conditions that are either slightly better, slightly worse or equivalent to the conditions out the window.

The current investigation was designed to assess whether or not pilots consider graphical METAR information in their assessment of the situation. Providing pilots used GWIS METAR information, the current investigation sought to examine how this information is integrated with their "out-the-window" assessment. If pilots decide to incorporate the inaccurate GWIS information into their judgment it would influence both their sensitivity and response criterion. The directional shift in criterion would be dependent upon whether the GWIS was over estimating or underestimating conditions. Inaccurate information should reduce pilot sensitivity.

Hypotheses

1.) Pilots have evidenced an inability to estimate weather conditions both anecdotally (Boatman, 2001) and in previous research (Boatman, 2001; Goh & Wiegmann, 2001b). Previous research has shown pilots on average may be inaccurate by about 2200 feet (SD approximately 1000 feet) when judging ceiling and 1.5 miles (SD approximately 1 mile) when judging visibility (Goh & Wiegmann, 2001b). Similar errors and deviations when estimating weather conditions were expected in the present study.

2.) Previous research has suggested that pilots overestimate weather conditions (Goh & Wiegmann, 2001b). If this trend continues in the present study there should be a conservative response criterion for pilots. That is, pilots will be more likely to say conditions are VFR.

3.) Instrument pilots will be more accurate in their assessment of weather conditions. Experience has been shown to improve weather judgments (Wiggins & O'Hare, 1995). Therefore, since instrument rated pilots are not limited to flying in visual conditions, their additional experience training with different weather conditions should provide them with an advantage over non-instrument pilots who can only fly in VMC.

4.) Pilots' overall weather judgment will follow FARs. That is, weather categorization will be based upon a non-compensatory decision model where categorization is driven by the lesser of the two conditions. Specifically, pilots' comfort and categorization should be based only on the worst condition. This represents the safest strategy and also the correct strategy as outlined by FARs. Previous research (Driskill et al., 1997; Hunter et al., 2003) suggests that pilots do not follow FARs when combining weather information. According to the FAA's research, pilots use a compensatory model (combine both categories) when making weather judgments. However, these results were based upon textual weather information and the results might have been due to aspects of the experimental design. The multiple points of data used in the original experiments may have enticed pilots to use all of the available information. The present study compares the decision making strategy used in evaluating both textual information regarding the current conditions and an out-the-window depiction of the conditions. The out-the-window representation, which provides a holistic representation of the conditions, should demonstrate a non-compensatory decision making strategy as is outlined by FARs.

5.) When provided with only textual representations of the conditions, pilots are expected to demonstrate use of compensatory decision making strategy as found by previous FAA research (Driskill et al., 1997; Hunter et al., 2003).

6.) The pilots will use the GWIS when it is available. Pilots have been reported as being poor in their assessments of "out-the-window" weather conditions (Boatman, 2001; Goh & Wiegmann, 2001b), and may therefore rely on the GWIS to estimate the weather conditions. Pilots' assessments of the weather situation will vary with the consistency of the GWIS METAR information. That is, when METAR information does not precisely coincide with the out-the-window conditions pilots will show a lower sensitivity compared to the accurate and no-display conditions, and reveal a shift in the response criterion. The shift in the criterion would be in the direction of the inconsistency (i.e., when the ceiling is displayed as worse than the actual there would be a shift towards a more liberal response criterion).

METHODOLOGY

Participants

Twenty-four general aviation pilots participated in the experiment. One of the non-instrument rated pilots was female. Pilots were recruited, scheduled and compensated through a contract with Lockheed-Martin. The contractor at Lockheed-Martin provided a database of pilots without-military or commercial experience from the Mid-Atlantic region. Pilots were then selected from the database based upon cross-country hours and instrument ratings. Half of the pilots selected had instrument ratings and the other half were non-instrument rated. Pilots also supplied information on experience during the experiment at NASA Langley Research Center (Appendix D). There was some inconsistency in the information provided by pilots to Lockheed-Martin in their applications, and that obtained at the time of the experiment. Several pilots cited that they used their logbooks to provide the information to Lockheed-Martin, but did not have the exact information available at the time of the experiment. The mean experience data from both the study and Lockheed-Martin database are provided in Table 3.

TABLE 3: Mean (and Standard Deviation) Data for Pilot Experience

	Study				Lockheed-Martin		
	Age	Total Flight Hours	Last 90 Days	Cross-Country Hours	Total Flight Hours	Last 90 Days	Cross-Country Hours
IFR Rated	33.75 (10.89)	440.04 (259.00)	17.45 (19.12)	196.17 (156.48)	404.4 (221.46)	13.74 (13.66)	155.34 (107.40)
VFR	47.92 (13.81)	364.5 (159.75)	19.42 (16.76)	128 (65.04)	339.59 (151.82)	25.63 (26.67)	97.33 (40.11)

Apparatus and Materials

All experimental data were collected in a small experimental chamber at NASA Langley Research Center. Two computers linked via a local area network were used to run the experiment. The first computer drove the out-the-window weather depiction. This computer has a Pentium 4 3.0 GHz Processor, 1 GB of RAM, and a GeForce FX 5950 Ultra video card with 256 MB of video RAM. The computer was linked to a Dell 3300 MP projector with a maximum brightness of 1500 ANSI lumens, a native resolution of 1024 x 768 pixels, and a contrast ratio of 1700:1 (full on/full off). The projected image was 34.75 inches x 26 inches. The second computer displayed a static representation of the flight instruments and collected the participant's responses. The second computer had a Pentium Xeon 1.8 GHz Processor, and 512 MB of RAM. The information was presented on a 17 inch Dell flat panel display.

Pre-experimental Questionnaire

A pre-experimental questionnaire (See Appendix D) was administered to collect information from several different categories. The first part of the questionnaire acquired background and demographic information. Specifically information on ratings, total flight hours, cross-country hours, flight hours in the last 90 days, endorsements, age and gender were collected. The second part of the questionnaire surveyed pilots' weather and aviation operation knowledge. This section specifically addressed pilots' knowledge of FARs regarding VFR flight. A third section collected data about personal VFR minimums. The final section surveyed pilots' attitudes towards risk, their own piloting abilities, their involvement in hazardous flight activities, and FARs regarding VFR flight.

Experimental Instructions

Following the pre-experimental questionnaire pilots were given instruction about the experiment. This instruction was in the format of a presentation. Pilots where told that their objective was to view a series of simulated out-the-window video clips and answer several questions regarding the scene. The questions were to be answered as if there was preflight information obtained 2 hours ago that indicated that a warm front may be coming through the area and there was a possibility of reduced ceilings and visibility.

Pilots were told that graphical METAR information for an airport approximately 1 mile from their position would be available on some of the trials. They were informed that this METAR information could be between 5 – 60 minutes old and they would not be given any indication of the actual age of the report. They were told that how they used the information in the METAR to answer the questions was at their discretion. The system used was representative of EchoMap™. A sample METAR screen is presented in Figure 4.

Weather Generation Program

Several pilots at NASA Langley Research Center served as pretest pilots. The initial out-the-window imagery was static representations of different weather conditions rendered in a flat area without physical features. Feedback from the pretest pilots indicated that lack of terrain features made it difficulty to determine altitude and distance. The pretest pilots also indicated the need for motion to help judge their distance from the clouds. Pretest pilots also suggested the use of landmarks to help judge distance. The pilots also provided feedback on images rendered with different flight simulator software.

Based upon the feedback from the pretest pilots, the out-the-window weather conditions were generated through Microsoft's Flight Simulator 2004. The videos were augmented with satellite imagery and terrain overlays from MegaScenery™ and displayed via an overhead projector. Due to a limitation of the software, all of the clouds rendered were stratus clouds with a height of 900 feet. All of the videos taken from Flight Simulator 2004 used the same location over Brookhaven Airport (KHWV) in Shirley, NY.

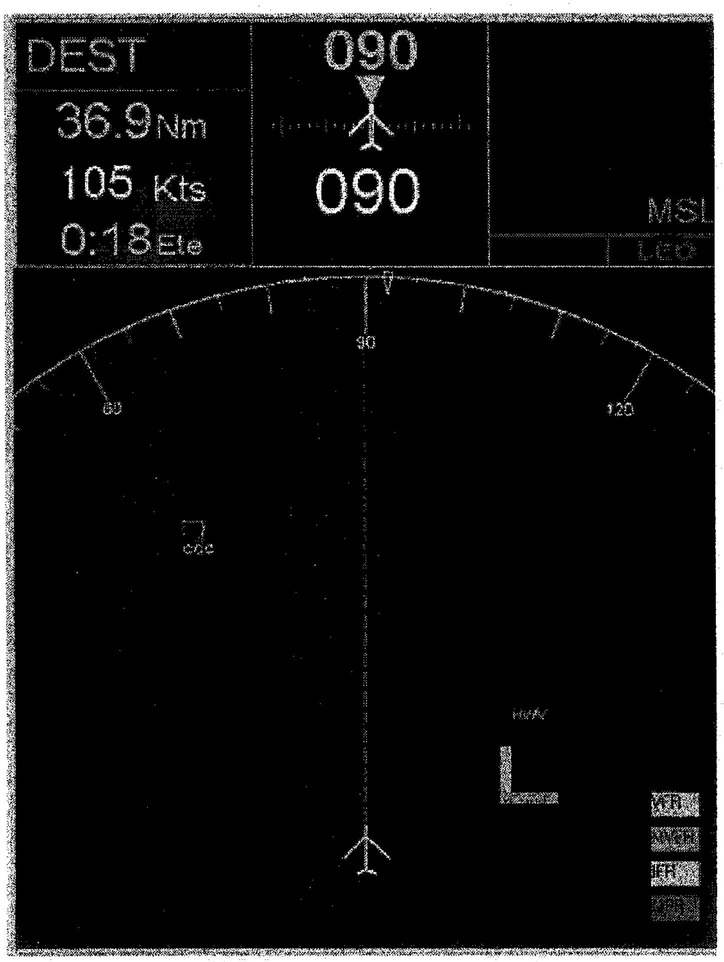


Figure 4. Graphical Weather Information System.

This location was selected based upon several criteria. First the airport is approximately at sea level. This alleviates any confusion between MSL and AGL when pilots estimated cloud ceiling. Second, the area had satellite imagery available to provide a more realistic depiction. Third, the specific region had several features that could help pilots estimate distance including an interstate (approximately 2 miles from the aircraft's position), and Calverton, a former naval base and operational very-high-frequency omni-directional range (VOR) (approximately 5 miles away). All of the videos collected were five seconds in duration and started from the same location just east of KHWV. The video resolution was 1024 by 768 pixels.

Post Experiment Task

The post experiment task followed the methodology previously used by FAA sponsored research (Driskill et al., 1997; Hunter et al., 2003). Pilots were given 16 cards with different weather scenarios. The cards gave the pilots ceiling and visibility information. Ceiling was always listed before visibility. Pilots were asked to sort the cards according to how comfortable they would be flying given the conditions represented on the cards. After sorting the cards from least comfortable to most comfortable pilots were then asked to assign comfort ratings for each card ranging from 0 to 100.

Procedure

Pilots first read and signed a written informed consent document. They were then briefed about the schedule for the day. Participants were then asked to complete the pre-experiment questionnaire on the second computer. Upon completing the questionnaire, pilots were given training on interpreting the data on a GWIS. The training included a familiarization with the test area including images taken from sectional charts and a depiction of the area without any weather. Landmarks and their approximate distances were given to the pilots. Upon completing the training, pilots began the main experimental trials.

The main experimental trials procedure presented pilots with a five-second out-the-window video. The video was displayed using an overhead projector. A second display provided pilots with the primary flight instruments and the moving map display that also served as a platform for the graphical METARs. The primary instruments include the altimeter, attitude indicator and airspeed indicator.

The videos looped until the pilots answered a set of questions regarding the current situation. Using the available information, pilots were required to answer several questions (see Figure 2). Questions appeared one at a time, and pilots were given an unlimited amount of time to answer each question. After completing each question the next trial was presented. There was a two-second pause between the presentation of each trial. Pilots completed two blocks of 46 trials. Each block consisted of all of the different ceiling, visibility, and GWIS manipulations. The order of presentation was randomized within each block. A break was provided in between the two blocks. After completing the second block of trials the participant completed the post experiment task.

Baseline (No GWIS conditions). The different out-the-window videos were presented via an overhead projector. There were a total of 16 different videos in the baseline conditions. These videos represented a combination of four different ceilings and four different visibilities. Ceiling was manipulated by adjusting the base layer of the clouds. The definition of ceiling followed the definition of ceiling provided in the FARs, which is also used in aviation weather reports. Specifically, a ceiling is the lowest layer of clouds or obscuring phenomenon that is reported as broken, overcast or an obscuration. Ceiling is reported as above ground level. Cloud cover was presented at four different base levels (400, 900, 2900, and 4500 ft). VFR conditions were those above 3,000 feet, MVFR ceilings were between 1,000 and 3,000 feet, IFR conditions were between 500 and 1000 feet, and LIFR conditions were below 500 feet. Visibility took on one of four distances (2, 3, 5, and 10 miles). These four levels represented one VFR, two MVFR and one IFR.

The aircraft's altitude was always at 2400 feet AGL. This allowed both clearances above the low clouds and below the higher clouds. Pilots must maintain at least 500 feet below the cloud deck or 1000 feet above the cloud deck to remain in VFR. Since pilots cannot control the aircraft (i.e., climb or descend) they were asked to evaluate the conditions from the perspective of a ground station. That is, are the ceilings and visibility above VFR minima with respect to that area? When asked about their comfort in continuing VFR, pilots were instructed to answer this considering the fact that they could climb or descend if necessary. The 36 baseline conditions were thus broken down across four factors, ceiling (4), visibility (4), and block replication (2).

GWIS conditions. Weather information on the GWIS was manipulated so that there were five levels of weather information consistency. The weather conditions are presented below in Table 4. The GWIS provided graphical METAR information for an airport less than 1 mile from the aircraft. Graphical METARs provide ceiling and visibility information for equipped airports. METARs are issued every hour and graphically present weather conditions in four categories (LIFR, IFR, MVFR, VFR).

When information in the GWIS was better or worse than the actual conditions the difference was in the magnitude of one category. For example when the ceiling information on the GWIS was worse than reported, the GWIS ceiling was IFR when out the window conditions were MVFR. To allow for a full manipulation of information consistency, out the window ceiling and visibility was constrained to either IFR or MVFR when flying with a GWIS. The 60 GWIS trials can be broken down across 5 factors, ceiling (2), visibility (3), GWIS consistency (5), and block replication (2).

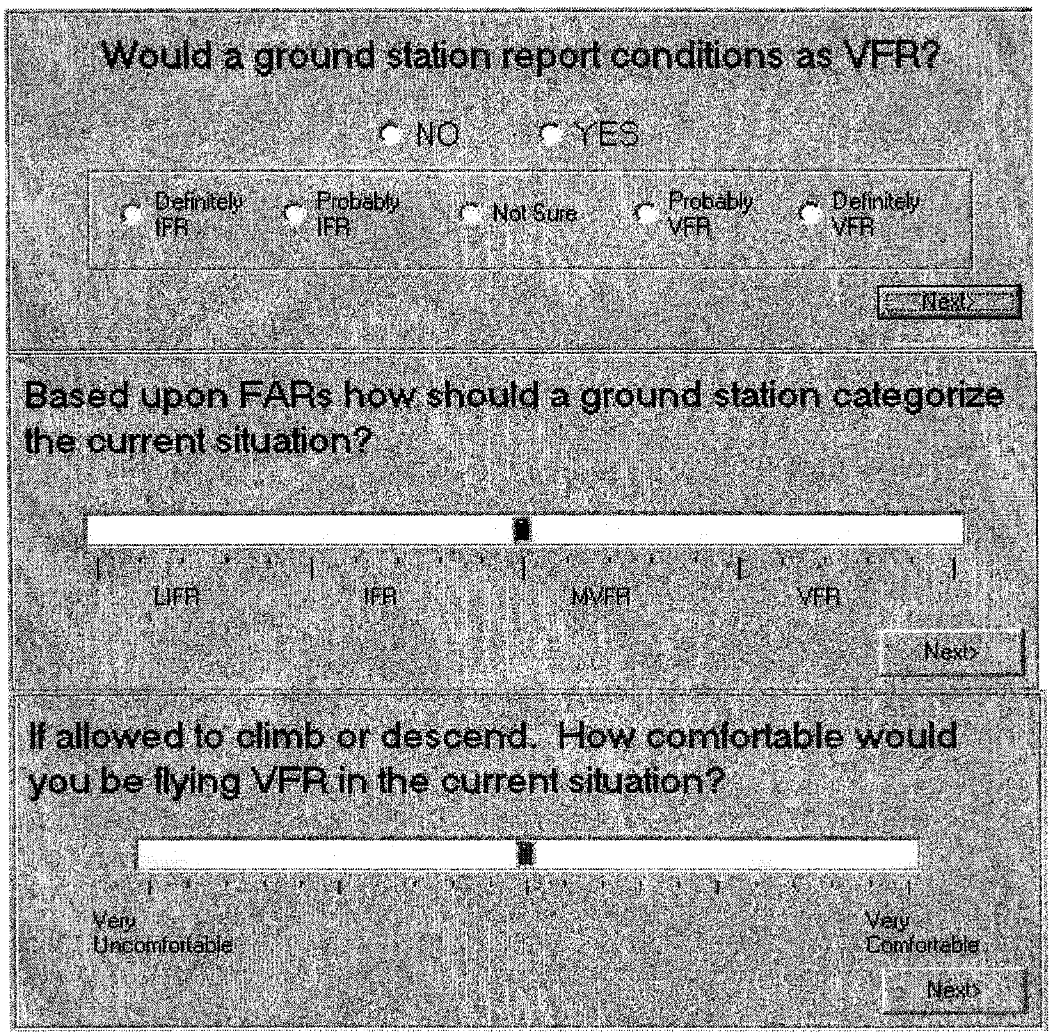


Figure 5. Screen shots of the experimental scenario questions.

There were a total of 46 experimental scenarios that were replicated twice for a total of 92 experimental trials. The experimental trials were divided into two blocks of 46 trials. There were a total of 32 trials without any GWIS information, and 60 trials with GWIS information of varying levels of consistency.

TABLE 4: GWIS Information Consistency Manipulations

	Out the window	
Condition	Ceiling	Visibility
GWIS Consistent	Same	Same
GWIS Higher Ceiling	Better	Same
GWIS Lower Ceiling	Worse	Same
GWIS Greater Visibility	Same	Better
GWIS Poorer Visibility	Same	Worse
No GWIS	NA	NA

Dependent Measures

Crisp SDT (sensitivity and response criterion). The crisp SDT analysis followed the formulas originally developed by Green and Swets (1966). The different SDT formulas can be found in the Appendix C. The response data used for the crisp SDT analysis were based upon the question "Is the current situation VMC or IMC?" The crisp SDT response matrix is provided in Table 2. Actual categorization of the conditions was supposed to follow FARs.

The application of SDT to weather has the potential for confusions particularly with regard to response bias. Figure 3 is provided to help clarify the SDT terms, principally the response bias. The sensitivity metric in SDT represents the ability to distinguish between IMC and VFR conditions. The lower the sensitivity the more overlap there would be between the signal and noise distributions depicted below. The response bias represented by the vertical bar in Figure 6 is the participant's criterion for responding that conditions are either VFR or IFR. The participant's response bias depicted in Figure 6 represents a neutral response criterion or a c of 0. However, if pilots are more likely to respond that conditions are VFR the vertical bar would shift to the right and c would increase. This would then be a conservative shift in response bias resulting in fewer false alarms at the cost of more misses. From a safety perspective a conservative bias (i.e., pilots tending to respond conditions are VFR) could potentially result in more accidents. Alternatively, a decrease in c (a shift to the left) would result in more hits at the cost of more false alarms. A c below 0 would be a liberal response bias and from a safety perspective this could potentially result in fewer accidents.

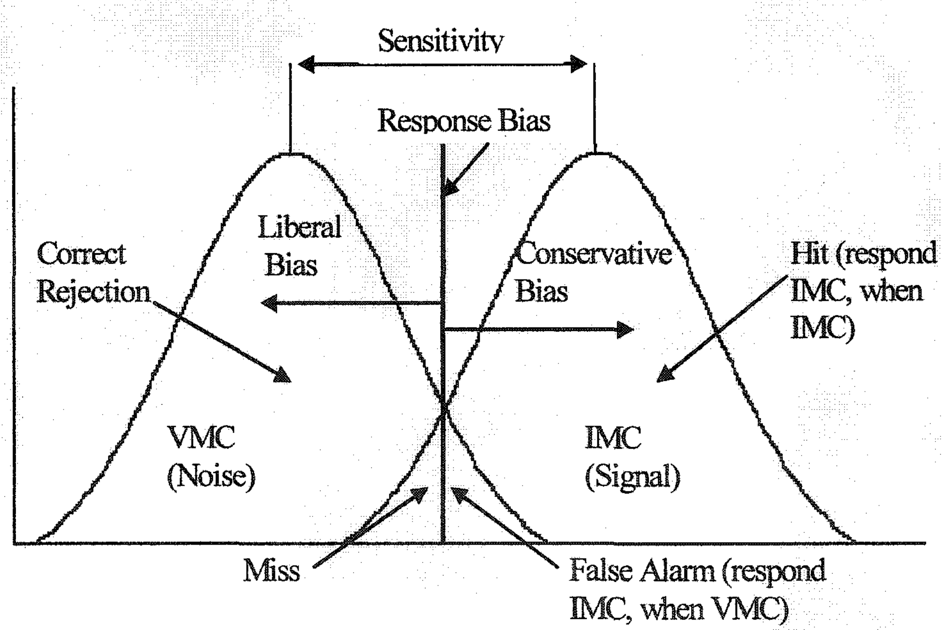


Figure 6. Application of signal detection to weather.

Fuzzy SDT (sensitivity and response criterion). Use of fuzzy SDT requires that either the signal or response vary along a continuum. A mapping function for the signal uses the objective properties of the world to determine the signal's position along a continuum. The mapping functions for both ceiling and visibility consist of four linear transformations. These transformations are based upon the categorization of weather across four levels, i.e., LIFR, IFR, MVFR and VFR. The LIFR conditions take on a signal value between 1 and .95, IFR values fall between .85 and .95, and MVFR values fall between 0 and .85, VFR conditions are always categorized as 0. The functions for ceiling and visibility across the two conditions are provided in Table 5 below. Figures 7 and 8 are a graphical representation of these functions for ceiling and visibility respectively.

TABLE 5: Mapping Functions for Ceiling and Visibility

	Mapping Functions	
Category	Ceiling	Visibility
LIFR	If ceiling < 500 then s = (-.0001 * ceiling) + 1	If visibility < 1 then s = (-.05 * visibility) + 1
IFR	If ceiling ≥ 500 and < 1000 then s = (-.0002 * (ceiling – 500) + .95)	If visibility ≥ 1 and < 3 then s = (-.05 * (visibility – 1) + .95)
MVFR	If ceiling ≥ 1000 and < 3000 then s = (-.000425 * (ceiling – 1000) + .85)	If visibility ≥ 3 and < 5 then s = (-.425 * (visibility – 3) + .85)
VFR	If ceiling ≥ 3000 then s = 0	If visibility ≥ 5 then s = 0

The mapping functions were created such that the relative position within each of the four categories was equivalent for both ceiling and visibility. Specifically, the midpoint of the ceiling MVFR category is 2000 ft. This value generated the same signal strength as the midpoint of the visibility MVFR category, which is 4 sm. All VFR conditions took on signal strength of 0.

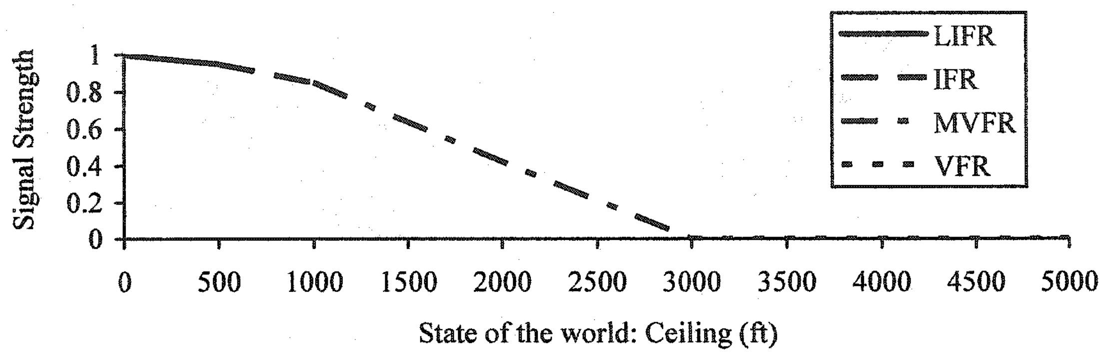


Figure 7. Fuzzy mapping function for ceiling.

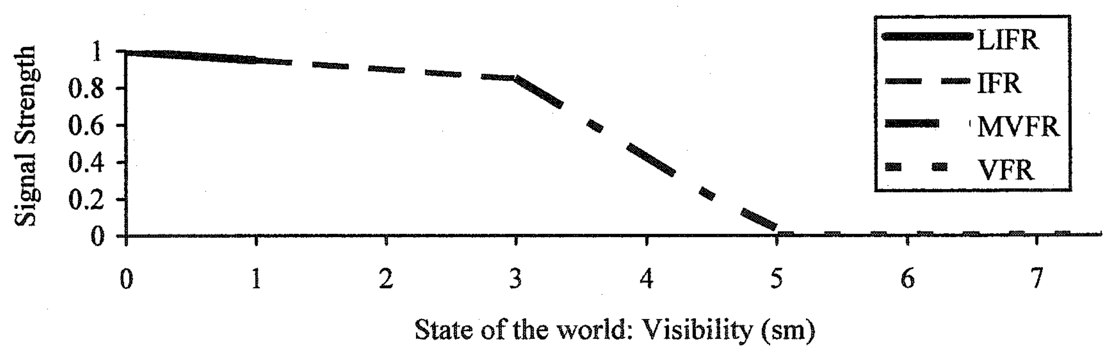


Figure 8. Mapping function for visibility.

Weather categorization is based upon both ceiling and visibility. Two separate mapping functions are depicted below. The state of the world is based upon the condition closest to 1, or closest to IFR conditions. The fuzzy response was generated from the category slide bar (see Figure 2, question number 2). The response strength was generated from the position of the bar. The fuzzy mapping for the response was the same as the mapping for the signal. The position within each category was determined by the strength of the response.

The mapping functions for the signal and noise can be used to compute hit, miss, false alarm, and correct rejection rates. The major difference is that a response can fall into more than one of the four categories used in crisp SDT. For example if the weather is computed as .75 from the mapping function and the pilot's response maps to .85 the resulting response would be equivalent to a .75 hit, .10 false alarm, and .15 correct rejection. Memberships in the four-decision response outcome matrix were defined by functions derived from Parasuraman et al. (2000). Once the hit rate and false alarm rate have been determined computation of d' and β are the same as crisp SDT.

TABLE 6: Fuzzy Response Categorization

	Function
Hit	H = minimum (s, r)
Miss	M = maximum (s-r, 0)
False Alarm	FA = maximum (r-s, 0)
Correct Rejection	CR = minimum (1-s, 1-r)

Comfort. Pilots were asked to rate their comfort in continuing the flight under VFR (see Figure 5 question 3).

Ceiling and Visibility Absolute Error. Participants provided absolute ceiling information in feet and visibility information in statue miles. The error was calculated as the absolute difference between the actual and reported conditions.

Design

A series of multiple univariate ANOVAs were used instead of a single MANOVA. The multivariate and univariate analyses are intended to be used to address separate research questions. According to Huberty and Morris (1989) it is inappropriate to follow a MANOVA with multiple univariate analyses. A MANOVA should be used when the researcher is interested in determining if any overall (including main effects and interactions) differences exist. The analysis forms an underlying construct based upon all of the dependent variables. In addition if significance is found, the researcher can also identify the relative contribution of each of the dependent variables. However, for the purposes of this study the authors consider the dependent variables to be "conceptually independent." There is no interest in creating a linear composite of all of the dependent variables. Instead, the interest is in seeing where differences lie in each of the dependent variables. This is particularly the case for the crisp and fuzzy SDT metrics. Therefore, for the purposes outlined in the current experiment, multiple univariate analyses were performed.

ANOVAs. A series of separate univariate ANOVAs were performed on the dependent measures listed above. Two separate series of ANOVAs correspond to the baseline conditions and the GWIS conditions plus their no-GWIS equivalents. Additionally, these two series were broken down based upon the dependent measures involved. The SDT metrics were not analyzed with ceiling and visibility as an independent variable (IV). For a SDT analysis there needs to be a signal and noise. Since signal and noise are defined by the IVs ceiling and visibility they were not included in the ANOVAS with the SDT metrics.

The no GWIS conditions was analyzed with a three way, Ratings (2) x Ceiling (4) x Visibility (4), mixed repeated measures ANOVA. Pilot instrument ratings were the only between subject variable (i.e., instrument rated vs. non-instrument rated). The remaining variables were all within subjects. Ceiling had four different levels (400 ft, 900 ft, 2900 ft, and 4500 ft). Visibility had four levels (2, 3, 5, and 10 miles). The first ANOVA series involved the comfort ratings, and ceiling and visibility estimates. The different sources of variation and the error terms used in the baseline analyses are provided in Table 7.

TABLE 7: Sources of Variation for the Baseline Analyses

Effect	df source	Error term	df error
Rating (R)	1	Rating (Subject)	22
Ceiling (C)	3	Rating (Subject) x Ceiling	66
Visibility (V)	3	Rating (Subject) x Visibility	66
R x C	3	Rating (Subject) x Ceiling	66
R x V	3	Rating (Subject) x Visibility	66
C x V	9	Rating (Subject) x C x V	198
R x C x V	9	Rating (Subject) x C x V	198

A second series of ANOVAs compared the pilot ratings on the four different SDT metrics.

The third ANOVA series compares four factors, Ratings (2) x Ceiling (2) x Visibility (3) x GWIS (6). There were six levels for the weather information display. These include the baseline moving map only condition, and five levels of information consistency (See Table 5). Ceiling has only two levels (900 ft, 2900 ft) and visibility has three levels (2, 3, and 5 miles). The reduced number of out the window conditions resulted because only IFR and MVFR conditions can receive the full manipulation of information consistency. This series of ANOVAs was performed on the comfort ratings, and ceiling and visibility estimates. The different sources of variation and the error terms used in the baseline analyses are provided in Table 8.

TABLE 8: Sources of Variation for the GWIS Analyses

Effect	df source	Error term	df error
Rating (R)	1	Rating (Subject)	22
Ceiling (C)	1	Rating (Subject) x Ceiling	22
Visibility (V)	2	Rating (Subject) x Visibility	44
GWIS (G)	5	Rating (Subject) x GWIS	110
R x C	1	Rating (Subject) x Ceiling	22
R x V	2	Rating (Subject) x Visibility	44
R x G	5	Rating (Subject) x GWIS	110
C x V	2	Rating (Subject) x C x V	44
C x G	5	Rating (Subject) x C x G	110
V x G	10	Rating (Subject) x V x G	220
R x C x G	5	Rating (Subject) x C x G	110
R x V x G	10	Rating (Subject) x V x G	220
R x C x V	2	Rating (Subject) x C x V	44
C x V x G	10	Rating (Subject) x C x V x G	220
R x C x V x G	10	Rating (Subject) x C x V x G	220

The fourth ANOVA series compared two factors, Ratings (2) x GWIS (6). This ANOVA series was performed on the four SDT metrics.

RESULTS

Baseline SDT Results

Signal detection metrics were computed for each participant for all baseline trials. In several instances crisp metrics indicated that participants had either a hit rate (HR) of 1 (2 VFR pilots, 0 IFR pilots), or false alarm rate (FAR) of 0 (4 VFR pilots, 5 IFR pilots). These extreme values prevent the calculation of d', B, and c. This problem did not occur for the fuzzy computations. Several solutions have been proposed to address this problem. The hit and false alarm rate metrics can be pooled across participants, however this technique can be problematic and is only recommended if participants have similar response biases and sensitivities (Macmillan & Kaplan, 1985). For example, if participants with equal sensitivity but different biases were pooled, their pooled sensitivity would be an underestimate. The most common solution is to correct the extreme scores by replacing rates of 0 with (0.5 / n), and rates of 1 with (n – 0.5)/ n, where n is the number of trials (Stanislaw & Todorov, 1999). However this approach can also lead to biased measures of sensitivity. Hautus (1995) recommends the loglinear approach, where .5 is added to the number of hits and false alarms and 1 is added to the number of signals and noise trials. This correction is applied irrespective of whether the data is extreme or not. Although only changing the extreme scores is the most common approach (Stanislaw & Todorov, 1999) the loglinear approach is less susceptible to bias and was therefore used in the computation of SDT metrics. Table 9 below provides the corrected and uncorrected hit rates and false alarm rates for both the crisp and fuzzy responses.

TABLE 9: Computation of Hit Rates and False Alarm Rates Before and After the Loglinear Correction

	Crisp				Fuzzy	
	HR	Corrected HR	FAR	Corrected FAR	HR	FAR
IFR	.808	.794	.111	.141	.770	.057
VFR	.833	.817	.118	.147	.834	.118

Both parametric and nonparametric SDT measures were calculated. However since response bias was not manipulated nonparametric measures are more appropriate. There was no significant difference between instrument and non-instrument rated pilots on any of the SDT metrics. ANOVA Tables for all interactions and means are available in Appendix E. The detailed computation of SDT metrics for each observer is provided in the Appendix F.

TABLE 10: Crisp and Fuzzy SDT Measures for the Baseline Trials

	Crisp SDT				Fuzzy SDT			
	d'	Beta	A'	c	d'	Beta	A'	c
IFR	2.114	2.106	.916	.180	2.542	4.860	.922	.486
VFR	2.206	1.550	.918	.067	2.191	2.719	.922	.178

Ceiling and Visibility SDT. The pilots' estimations of ceiling and visibility allowed for additional SDT analyses. In the original SDT calculations conditions were, normatively, based upon the worst conditions, however in the current analysis separate metrics were calculated for ceiling and visibility. Pilots' estimations of ceiling and visibility were used to determine crisp and fuzzy hit and false alarm rates to look at both ceiling and visibility independently. The participants' response was considered IFR for ceiling when estimates were below 1000 ft and IFR for visibility when estimates were below 3 miles. The fuzzy calculation was based upon the same formulas used to classify the conditions.

The SDT data generated from the ceiling and visibility estimates was checked with the overall SDT data for consistency. The estimate data had to be compared to the overall SDT data because there were no SDT data for ceiling and visibility. The overall SDT data created from the estimate were based upon the lowest response (e.g., for the overall crisp if either estimate was IFR then the overall response was IFR). The correlation between the overall crisp SDT data generated from the pilot estimates and the overall crisp SDT response was .51. This correlation is based upon two dichotomous variables, and the lack of variability in both variables has the effect of lowering the correlation. Comparing the response generated from the estimates and the crisp SDT response, 70 percent of the responses matched. The correlation between the fuzzy SDT data generated from the estimates and the overall fuzzy SDT response was .85.

The data generated from the ceiling and visibility estimates had a large number of extreme scores and therefore all of the calculations are derived from numbers based upon a loglinear correction. The number of obtained extreme scores is presented in Table 11. Hit rate and false alarm rates are provided for both instrument and non-instrument pilots in Table 12. Data for the non-parametric SDT measures are displayed in Table 13.

Baseline Accuracy, RMSE and Comfort Data

The accuracy, RMSE and comfort data were analyzed using a three-way, ratings (2) x ceiling (4) x visibility (4), mixed repeated measures ANOVA. All significant main effects were followed up by Tukey post hoc analyses.

Accuracy. Analysis of the accuracy data revealed no main effects for pilot rating, $F(1,22) = 0.212, p > .05$; or ceiling, $F(3,66) = 2.181, p > .05$. A significant main effect was found for visibility, $F(3,66) = 5.848, p < .05$. Post hoc analysis revealed that accuracy for the 10-mile visibility conditions was significantly lower than accuracy for the 2 and 5 mile conditions and there was no significant difference between the 10 and 3 mile visibility conditions. The data for the main effect of visibility on accuracy are presented in Figure 9.

TABLE 11: Number of Extreme Scores in the Crisp and Fuzzy Ceiling and Visibility Analyses

	Crisp							
	Ceiling				Visibility			
	HR = 1	HR = 0	FAR = 1	FAR = 0	HR = 1	HR = 0	FAR = 1	FAR = 0
IFR	0	2	0	10	1	0	0	5
VFR	0	2	0	9	4	0	0	0
	Fuzzy							
	Ceiling				Visibility			
	HR = 1	HR = 0	FAR = 1	FAR = 0	HR = 1	HR = 0	FAR = 1	FAR = 0
IFR	0	0	0	2	0	0	0	1
VFR	0	0	0	2	1	0	0	0

TABLE 12: Ceiling and Visibility Response Rates

	Ceiling				Visibility			
	Crisp HR	Crisp FAR	Fuzzy HR	Fuzzy FAR	Crisp HR	Crisp FAR	Fuzzy HR	Fuzzy FAR
IFR	.324	.044	.744	.081	.630	.120	.742	.094
VFR	.363	.054	.763	.114	.769	.220	.817	.224

TABLE 13: Nonparametric SDT Measures for Ceiling and Visibility

	Ceiling				Visibility			
	Crisp A'	Crisp c	Fuzzy A'	Fuzzy c	Crisp A'	Crisp c	Fuzzy A'	Fuzzy c
IFR	.752	1.185	.903	.395	.856	.493	.726	.116
VFR	.739	1.108	.891	.262	.850	-.014	.796	.239

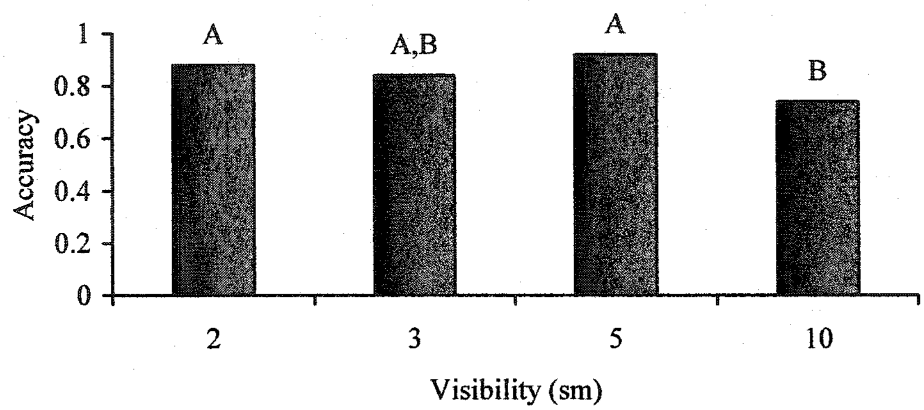


Figure 9. Accuracy data across the baseline visibilities. Means with different letters are significantly different at $p < .05$.

A significant interaction of ceiling and visibility was also found, $F(9,198) = 16.647$, $p < .05$. This interaction is depicted in Figure 10 below. Separate analyses inspected the effect of visibility at each of the 4 different ceiling conditions. Visibility had a significant simple main effect when analyzed at the 400, 900 and 4500 ft ceiling conditions, but not at the 2900 ft ceiling condition. When analyzed at the two IFR (400 ft and 900 ft) conditions there was a significant difference between the 10-mile visibility condition and all of the other visibilities. The 10-mile visibility condition had a significantly lower accuracy than the 2, 3 and 5 miles conditions within the 400 and 900 ft ceiling conditions. No significant differences were found for the different visibilities at the 2900 ft ceiling condition. The analysis at the 4500-foot ceiling conditions revealed the 3-mile visibility condition was significantly lower than both the 5 and 10 miles conditions. The 3-mile condition was no significantly different from 2 miles within the 4500-foot ceiling conditions. There were no other interactions found for the accuracy data.

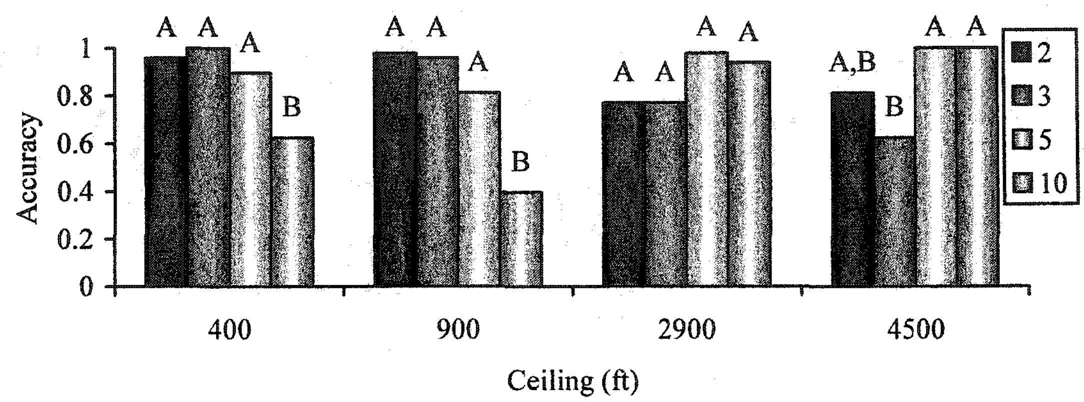


Figure 10. Accuracy for the different cloud heights across visibility. Means within each ceiling and with different letters are significantly different at $p < .05$.

Comfort. Analysis of comfort data revealed no main effect for pilot type. Significant main effects were found for both ceiling, $F(3,66) = 101.977, p < .05$ and visibility, $F(3,66) = 166.269, p < .05$. Mean comfort data across the different visibilities is reported in Figure 11. Post hoc analysis found significant increases in mean comfort with each increase in visibility.

Mean comfort data for ceiling is reported in Figure 12. There was a significant increase in comfort rating between the low ceilings (400 ft and 900 ft) and the higher ceilings (2900 ft and 4500 ft). There were no differences within the two low and two high ceilings.

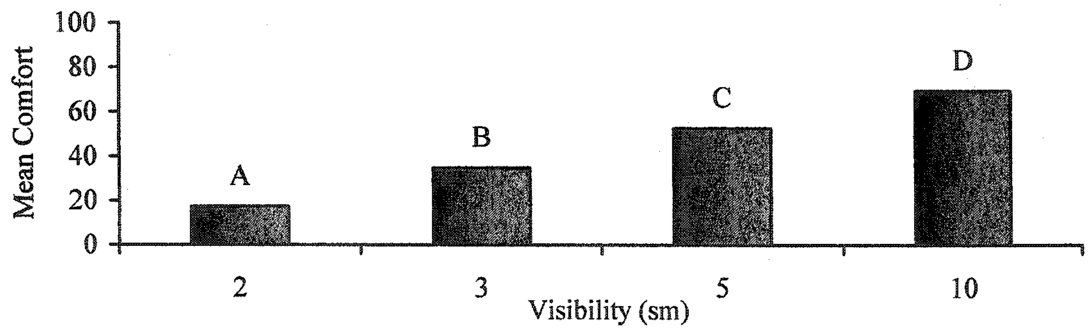


Figure 11. Mean comfort across the visibility conditions. Means within each ceiling and with different letters are significantly different at $p < .05$.

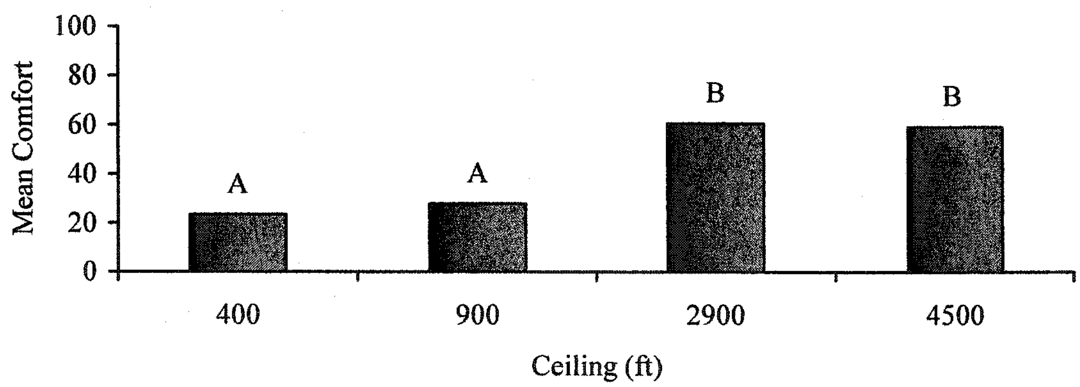


Figure 12. Mean comfort across the ceiling conditions. Means with different letters are significantly different at $p < .05$.

There was a significant interaction between ceiling and visibility, $F(9,198) = 8.495, p < .05$. The data for the interaction is presented in Figure 13. Four separate analyses were conducted at each level of ceiling. There was a simple main effect for visibility at each level of ceiling. At each level of ceiling comfort increased with increasing levels of visibility. At both 400 ft and 900 ft there were no significant differences between 2 and 3 mile visibility. However the 2 and 3 mile visibility conditions were significantly lower than the 5 mile visibility. Additionally the 5 mile visibility was significantly lower than the 10 mile visibility. Data for the 2900 ft and 4500 ft ceilings had significant increase in comfort with each increase in visibility. There were no other significant interactions for the comfort data.

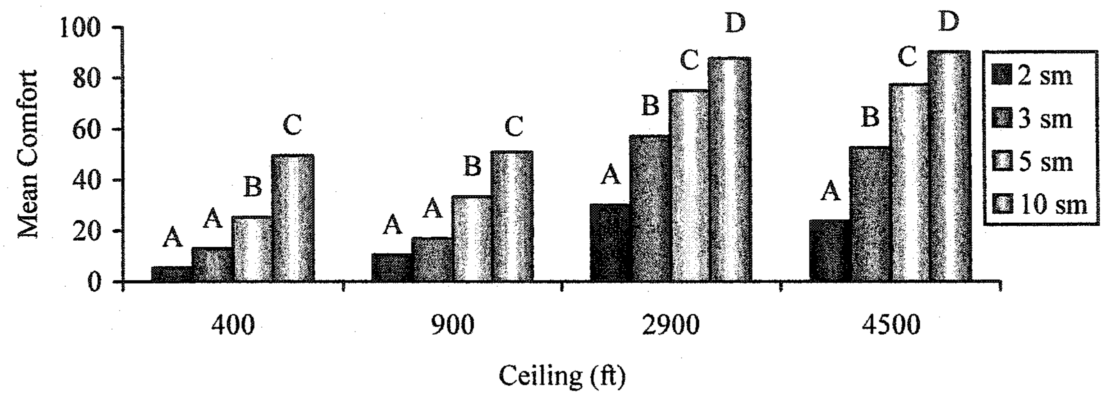


Figure 13. Interaction of ceiling and visibility on mean comfort. Means within each ceiling and with different letters are significantly different at $p < .05$.

Ceiling RMSE. Significant main effects were found for both ceiling, $F(3,66) = 18.99, p < .05$ and visibility, $F(3,66) = 3.14, p < .05$ in the ceiling RMSE data. Post hoc analysis of ceiling revealed that ceiling RMSE was significantly larger at 4500 feet than at any other level. There were no differences within the 400, 900, and 2900 ft ceilings. Ceiling RMSE data for each of the four ceilings is provided in Figure 14.

Post hoc analysis at the different visibility levels revealed no significant differences within 3, 5, and 10-mile visibility. No significant differences were found within the 2, 5, and 10-mile visibility conditions. However ceiling RMSE was significantly larger at 2 miles than it is at 3 miles.

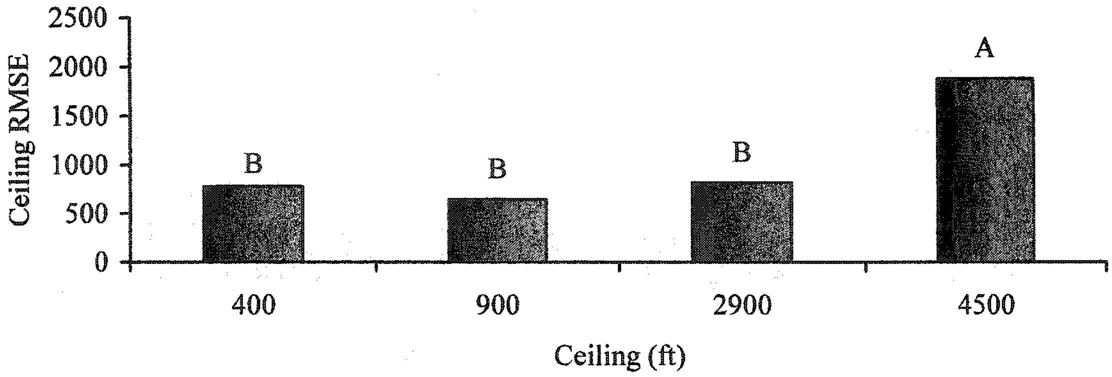


Figure 14. Ceiling RMSE data at each of the four ceilings. Means with different letters are significantly different at $p < .05$.

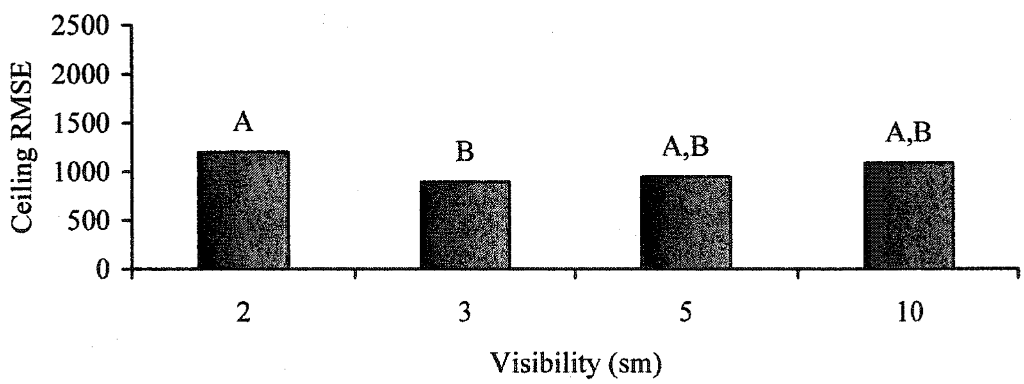


Figure 15. Ceiling RMSE data at each of the four visibilities. Means with different letters are significantly different at $p < .05$.

There was also a significant interaction of ceiling and visibility, $F(9,198) = 6.34$, $p < .05$. A series of analyses were done on visibility at each level of ceiling. There was a simple main effect at the 400, 900, and 4500 ft ceilings. The 2900 ft ceiling did not have a significant simple effect for visibility. At 400 ft there was a significant difference between the two lowest visibilities (2 and 3 miles) and the 10 mile condition. The 5 mile condition was not significantly different from any of the other visibilities. At 900 ft the only significant difference was between the 3 mile condition and the 10 mile condition. There were no differences among any of the visibilities with a 2900 ft ceiling. At the 4500 ft ceiling the 2 mile visibility had significantly higher error than any other visibility condition. No differences existed among the 3, 5, and 10 mile conditions. No other interactions were present within the ceiling RMSE data.

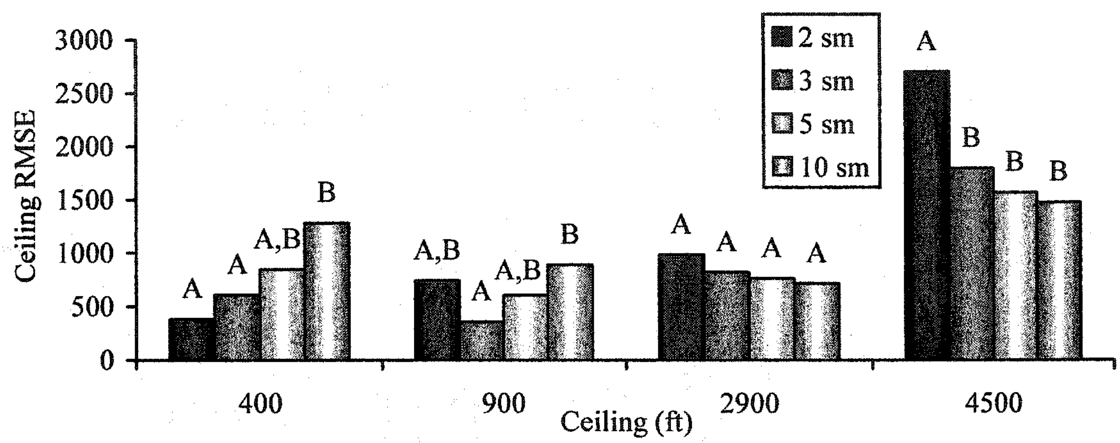


Figure 16. Ceiling RMSE data for the ceiling by visibility interaction. Means within each ceiling category and with different letters are significantly different at $p < .05$.

Visibility RMSE. Significant main effects were found for both ceiling, $F(3,66) = 6.387, p < .05$ and visibility, $F(3,66) = 23.283, p < .05$ in the visibility RMSE data. Post hoc analysis revealed that the RMSE at the 2 mile visibility was significantly smaller than all other visibilities. Additionally RMSE was smaller for 3 and 5 mile visibilities than RMSE data for the 10 mile visibility. RMSE data for visibility is presented below in Figure 17.

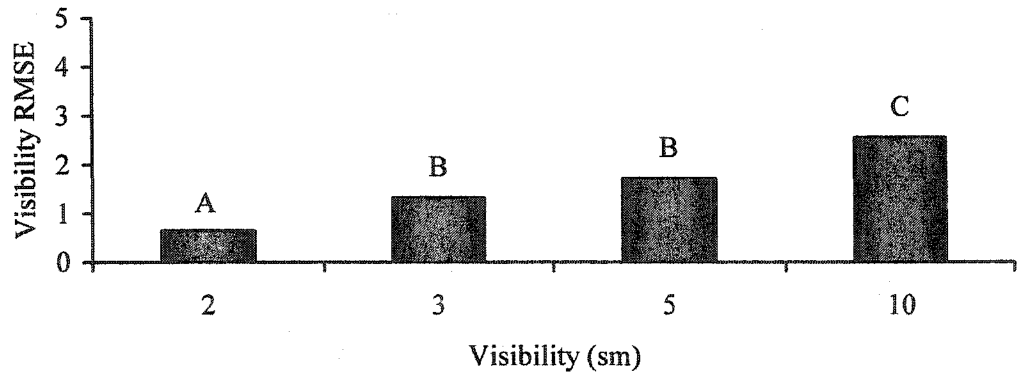


Figure 17. Visibility RMSE across the visibility conditions. Means with different letters are significantly different at $p < .05$.

Post hoc analysis on the visibility RMSE across the four levels of ceiling revealed that RMSE with the 400 ft ceiling conditions was significantly higher than conditions with 2900 ft ceilings and 4500 ft ceilings. RMSE with the 900 ft ceilings was significantly greater than the 4500 ft ceilings but not the 2900 ft ceilings. There were no differences between the 2900 ft ceilings and 4500 ft ceilings.

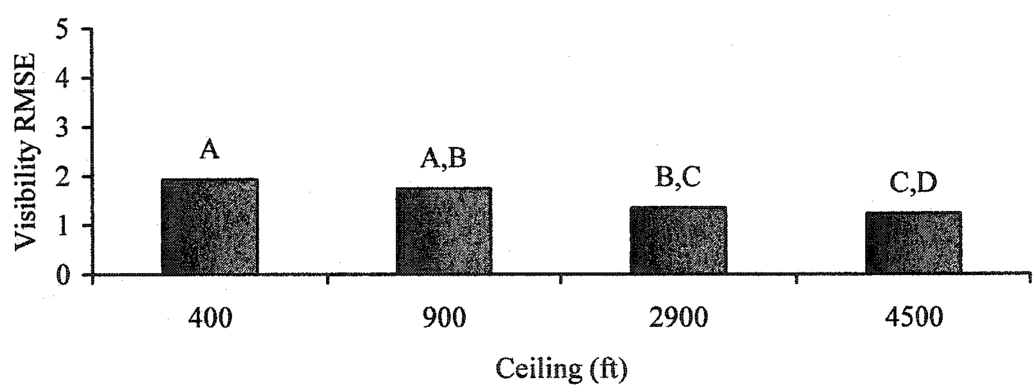


Figure 18. Visibility RMSE across the ceiling conditions. Means with different letters are significantly different at $p < .05$.

A significant interaction of rating and visibility was found for the visibility RMSE, $F(3,66) = 6.121$, $p < .05$. A series of analyses on the simple main effect of rating at each level of visibility were conducted. At 2, 3, and 5 mile visibility there was no simple main effect of rating. At 10 miles visibility the instrument pilots had a significantly lower visibility RMSE than the non-instrument pilots. The data for the visibility by rating interaction is presented in Figure 19.

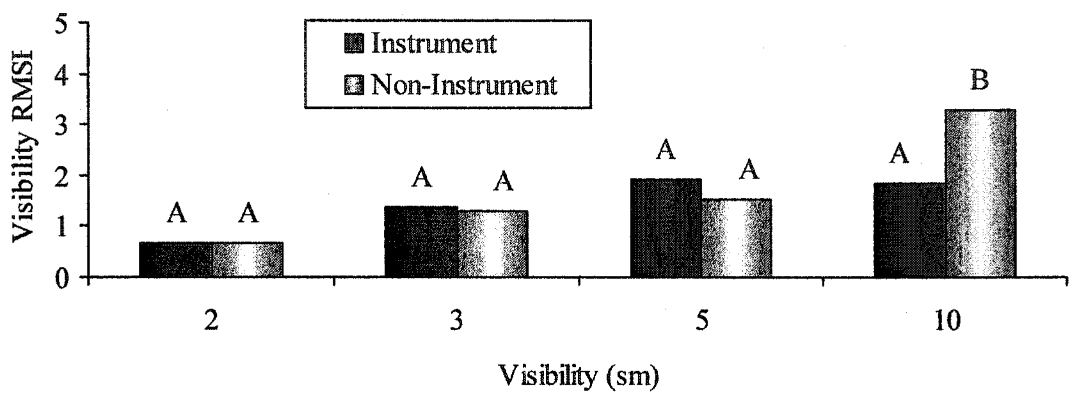


Figure 19. The interaction of visibility and rating on visibility RMSE. Means within each visibility and with different letters are significantly different at $p < .05$.

There was also a significant interaction of ceiling and visibility on visibility RMSE, $F(9,198) = 24.706$, $p < .05$. A series of analyses on the simple main effect of ceiling were performed at each level of visibility. A simple main effect of ceiling was present at 2, 3, and 10 mile visibility conditions. The analyses at 2 miles visibility found a significant difference between error at 400 ft and 900 ft, with the error at 900 ft being significantly smaller than error at 400 ft. There were no other significant differences at 2 miles. At 3 miles visibility, RMSE was significantly lower at 400 ft and 900 ft compared to 2900 ft. The 4500 ft condition at 3 miles was not significantly different from any of the other ceilings. There was no effect of ceiling in the 5 mile conditions. At 10 miles the two high ceilings (2900 and 4500 ft) had a significantly lower visibility RMSE than the low ceilings (400 and 900 ft). There were no differences within the high and low ceilings. Data for the interaction is plotted in Figure 20. There were no three-way interactions or any other significant interactions in the visibility RMSE data.

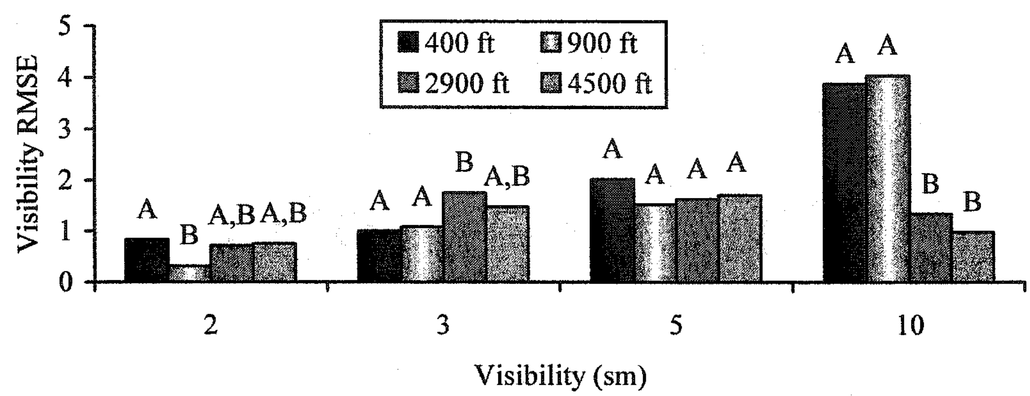


Figure 20. The interaction of ceiling and visibility on visibility RMSE. Means within each visibility and with different letters are significantly different at $p < .05$.

GWIS SDT Results

Non-parametric signal detection metrics were calculated for each participant. As with the baseline data there were several instances of extreme scores. However, unlike the baseline analysis this also extended to the fuzzy SDT data. A loglinear correction was performed on the GWIS SDT data. After the correction the SDT was analyzed in a two way, 2 (ratings) x 6 (GWIS), mixed repeated measures ANOVA. A table of the extreme scores from perfect performance is presented in Table 14 below. Additionally within the overall crisp SDT data there were three cases of a false alarm rate of 1. All three cases were non-instrument pilots. One was located in ceiling worse condition and the remaining two were within the visibility worse condition.

TABLE 14: Number of Cases with Perfect Performance within the GWIS SDT Data

	Overall Crisp											
	None		Accurate		Ceiling Better		Ceiling Worse		Visibility Better		Visibility Worse	
	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0
IFR	3	8	4	9	2	8	1	5	4	8	5	5
VFR	6	7	8	9	5	10	6	6	5	7	8	5
	Overall Fuzzy											
	None		Accurate		Ceiling Better		Ceiling Worse		Visibility Better		Visibility Worse	
	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0
IFR	0	0	0	0	0	0	1	0	0	0	1	0
VFR	0	0	0	0	0	0	2	0	0	0	2	0

Sensitivity. The ANOVA on the crisp A' found a significant main effect for GWIS, $F(5,110) = 4.647$, $p < .05$. The visibility worse was not significantly different from the ceiling worse condition, however it was significantly lower than every other GWIS condition. The ANOVA on the fuzzy A' also had a significant main effect of GWIS, $F(5,110) = 4.002$, $p < .05$. The ceiling worse condition had significantly lower sensitivity than the visibility better, accurate, and no GWIS conditions. There was no significant difference between ceiling worse, ceiling better and visibility worse. Data for both sensitivity metrics is provided in Figure 21.

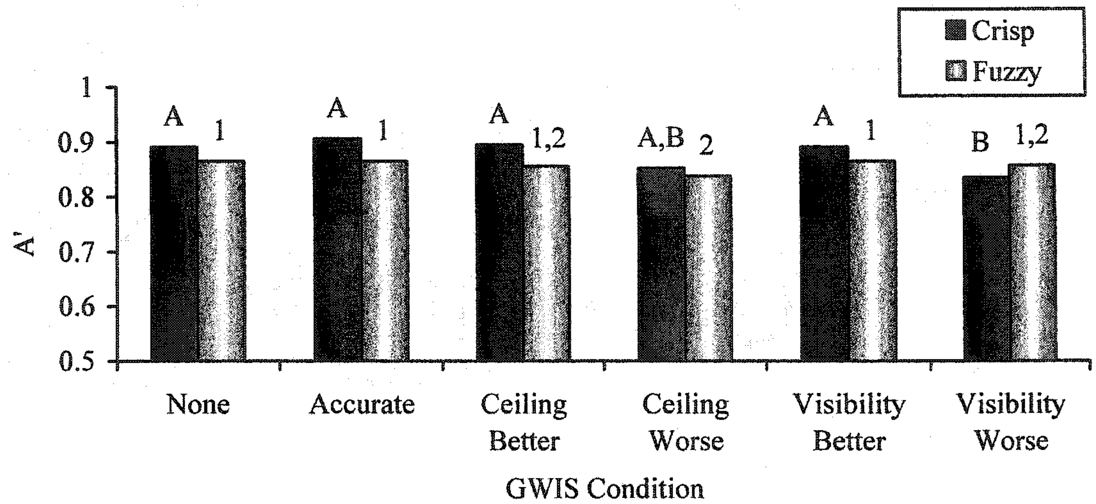


Figure 21. Sensitivity data for the GWIS conditions. Comparisons should only be made within each analysis technique (i.e., crisp or fuzzy). Means with different numbers or letters are significantly different at $p < .05$.

Bias The analysis of the crisp response bias data revealed a significant effect of GWIS on the crisp c, $F(5,110) = 4.705, p < .05$. Post hoc analysis revealed that the visibility worse condition had a significantly lower (more liberal, i.e., more likely to respond that conditions are IFR) c than any other GWIS condition with the exception of the ceiling worse condition. There were no other significant differences within the other GWIS conditions. Analysis of the crisp c revealed no other main effects or interactions. There was a significant main effect of GWIS on the fuzzy response bias, $F(5,110) = 12.227$, $p < .05$. Post hoc analysis revealed that the visibility worse and ceiling worse conditions had a significantly lower fuzzy c than the four other GWIS conditions. There were no significant differences between the visibility worse or ceiling worse conditions. There were also no differences within the four other GWIS conditions. Response bias data for both the fuzzy and crisp c is provided in Figure 22.

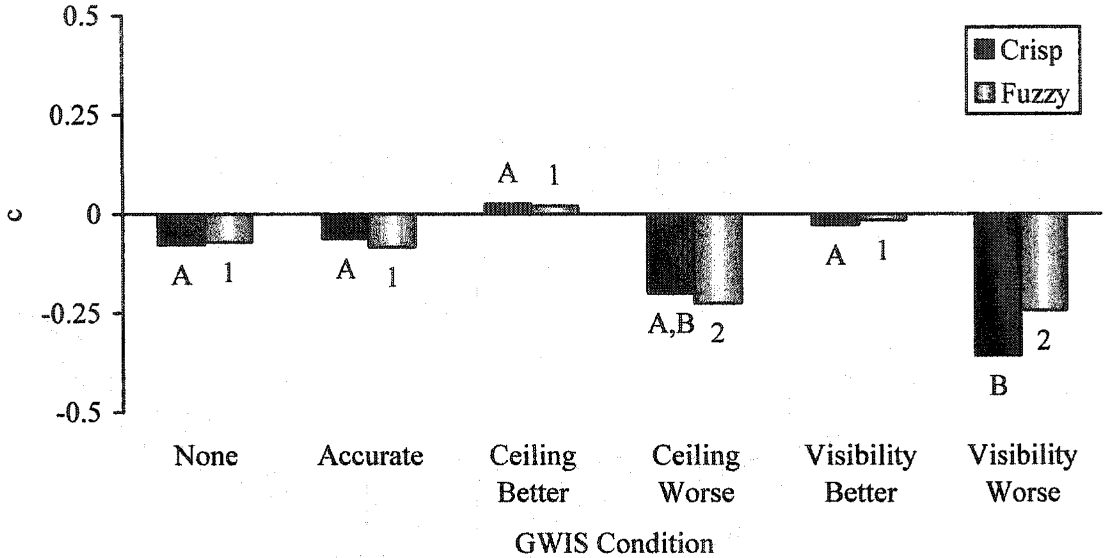


Figure 22. Combined response bias data for the GWIS conditions. Comparisons should only be made within each analysis technique (i.e., crisp or fuzzy). Means with different numbers or letters are significantly different at $p < .05$.

GWIS Ceiling SDT. As with the baseline trials, pilots' estimates of ceiling and visibility were used to create crisp and fuzzy responses to both ceiling and visibility. The incidence of extreme scores in the crisp ceiling analysis was high. Every pilot in every GWIS condition had a crisp ceiling false alarm rate of 0. Additionally there were a number of cases where the hit rate was equal to 0. There were also several instances of perfect ceiling scores within the fuzzy data, but no instances where the scores reflected all wrong responses. The data for the extreme GWIS ceiling scores is presented in Table 15. A loglinear correction was applied to all of the data.

The analysis was performed only using the nonparametric analysis. The data were analyzed in a two way, 2 (ratings) x 6 (GWIS), mixed repeated measures ANOVA.

TABLE 15: Number of Extreme Scores within the GWIS Ceiling SDT Data

	Ceiling Crisp Perfect Performance											
	None		Accurate		Ceiling Better		Ceiling Worse		Visibility Better		Visibility Worse	
	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0
IFR	0	12	0	12	0	12	1	12	0	12	0	12
VFR	0	12	0	12	0	12	0	12	0	12	1	12

	Ceiling Crisp All Incorrect											
	None		Accurate		Ceiling Better		Ceiling Worse		Visibility Better		Visibility Worse	
	HR = 0	FAR = 1	HR = 0	FAR = 1	HR = 0	FAR = 1	HR = 0	FAR = 1	HR = 0	FAR = 1	HR = 0	FAR = 1
IFR	4	0	6	0	7	0	3	0	5	0	4	0
VFR	3	0	3	0	6	0	1	0	3	0	3	0

	Ceiling Fuzzy Perfect Performance											
	None		Accurate		Ceiling Better		Ceiling Worse		Visibility Better		Visibility Worse	
	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0
IFR	0	4	0	4	0	7	1	2	0	5	0	2
VFR	0	2	0	2	0	3	0	1	0	2	0	1

There was a significant main effect for GWIS on the crisp sensitivity metric, $F(5,110) = 6.676$, $p < .05$, but not the fuzzy sensitivity metric. The highest crisp ceiling sensitivity was obtained within the ceiling worse condition. The ceiling worse condition had a significantly higher sensitivity than the no-GWIS, accurate, and ceiling better GWIS conditions. Additionally, the visibility worse condition had a significantly higher sensitivity compared to the ceiling better condition. The ceiling crisp A' means for the different GWIS conditions and the post hoc groupings are in Table 16. There was no main effect for GWIS on fuzzy ceiling A' data.

Analysis of both the crisp and fuzzy ceiling response bias yielded a significant main effect of GWIS, $F(5,110) = 10.597$, $p < .05$ and, $F(5,110) = 6.786$, $p < .05$ respectively. Within the crisp ceiling analysis, the ceiling worse condition had a significantly lower response bias than the no-GWIS, accurate, and visibility better GWIS conditions. The ceiling better condition had a significantly larger response bias compared to the no-GWIS and visibility worse conditions. Within the fuzzy ceiling response bias analysis the ceiling worse condition had a significantly lower response bias compared with the ceiling better and visibility better GWIS conditions. The ceiling better condition had a significantly higher response bias compared with the no-GWIS and visibility worse GWIS conditions. Means and post hoc groupings for the fuzzy and crisp ceiling response bias are in Table 17.

TABLE 16: Mean Ceiling Crisp A' and Post Hoc Groupings for the GWIS Conditions

	None	Accurate	Ceiling Better	Ceiling Worse	Visibility Better	Visibility Worse
Mean	.666	.679	.621	.778	.687	.729
Post hoc Group	1,2	1,2	1	3	1,2,3	2,3

TABLE 17: Mean Fuzzy and Crisp Ceiling c and Post Hoc Groupings for the GWIS Conditions

	None	Accurate	Ceiling Better	Ceiling Worse	Visibility Better	Visibility Worse
			Crisp			
Mean	.997	1.042	1.227	.735	1.048	.879
Post hoc Group	2	2,3	3	1	2,3	1,2
			Fuzzy			
Mean	.233	.253	.387	.113	.289	.208
Post hoc Group	1,2	1,2,3	3	1	2,3	1,2

GWIS visibility SDT. The crisp and fuzzy visibility SDT data contained several extreme scores. The frequency of the extreme scores within the visibility SDT analysis is presented in Table 18. A loglinear correction was used on both the crisp and fuzzy visibility data.

TABLE 18: Number of Extreme Scores within the GWIS Visibility SDT Data

	Visibility Crisp Perfect Performance											
	None		Accurate		Ceiling Better		Ceiling Worse		Visibility Better		Visibility Worse	
	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0
IFR	2	5	1	7	0	4	2	4	1	8	0	7
VFR	5	0	6	1	4	1	2	1	4	1	6	1

	Visibility Crisp All Incorrect											
	None		Accurate		Ceiling Better		Ceiling Worse		Visibility Better		Visibility Worse	
	HR = 0	FAR = 1	HR = 0	FAR = 1	HR = 0	FAR = 1	HR = 0	FAR = 1	HR = 0	FAR = 1	HR = 0	FAR = 1
IFR	1	0	2	0	3	0	1	0	3	0	2	0
VFR	0	0	0	0	0	0	0	0	2	0	0	0

	Ceiling Fuzzy Perfect Performance											
	None		Accurate		Ceiling Better		Ceiling Worse		Visibility Better		Visibility Worse	
	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0	HR = 1	FAR = 0
IFR	0	2	0	4	0	2	0	3	0	3	0	2
VFR	1	0	1	1	1	0	1	0	1	0	1	0

Analysis of the crisp visibility sensitivity yielded a significant main effect of GWIS, $F(5,110) = 2.432$, $p < .05$. The sensitivity for the visibility better condition was significantly lower than the no-display GWIS condition. No other differences were found within the crisp visibility sensitivity data. Crisp A' means and post hoc groupings for the GWIS conditions are provided in Table 19. There was no significant main effect of GWIS on the fuzzy A' data.

TABLE 19: Mean Crisp Visibility A' and Post Hoc Groups for the GWIS Conditions

	None	Accurate	Ceiling Better	Ceiling Worse	Visibility Better	Visibility Worse
Mean	.791	.783	.759	.782	.694	.769
Post hoc Group	2	1,2	1,2	1,2	1	1,2

Analysis of the visibility crisp and fuzzy response bias revealed significant main effects for GWIS, $F(5,110) = 4.982$, $p < .05$ and, $F(5,110) = 3.497$, $p < .05$ respectively. Means and post hoc groupings for the crisp and fuzzy response bias are available in Table 20.

TABLE 20: Mean Fuzzy and Crisp Visibility c and Post Hoc Groupings for the GWIS Conditions

	None	Accurate	Ceiling Better	Ceiling Worse	Visibility Better	Visibility Worse
	Crisp					
Mean	.183	.276	.323	.239	.563	.306
Post hoc Group	1	1	1,2	1	2	1
	Fuzzy					
Mean	.034	.046	.044	.040	.087	-.085
Post hoc Group	1,2	2	2	1,2	2	1

Additionally both the crisp and fuzzy c revealed a significant main effect of Rating, $F(5,110) = 8.054$, $p < .05$ and, $F(5,110) = 5.941$, $p < .05$ respectively. In both analyses instrument pilots had a significantly higher response criterion (tendency to overestimate conditions) than non-instrument pilots. The mean data for the fuzzy and crisp visibility response criterion is in Table 21.

TABLE 21: Visibility Response Criterion across Rating

	Crisp c	Fuzzy c
Instrument	.558	.190
Non-Instrument	.071	-.134

GWIS Accuracy, RMSE, and Comfort Data

The accuracy, RMSE and comfort data was analyzed using a four way, 2 (ratings) x 2 (ceiling) x 3 (visibility) x 6 (GWIS), mixed repeated measures ANOVA. Due to the previous baseline analysis, only main effects of the GWIS manipulation and interactions involving the GWIS manipulation will be explained.

Accuracy. There was a significant main effect of GWIS on the accuracy data, $F(5,110) = 2.911, p < .05$. Post hoc analysis revealed that the only significant difference was between the visibility worse condition and the accurate condition, with the accurate condition having significantly higher accuracy. The accuracy data for the GWIS conditions are provided in Figure 23.

The GWIS condition had a significant interaction with ceiling level. A test of simple main effect of GWIS at each of the two levels of ceiling was performed. There was no simple main effect for 900 ft, but there was a simple main effect for 2900 ft. Within the 2900 ft ceiling conditions, accuracy for the visibility worse conditions was significantly lower than accuracy for the accurate and ceiling better conditions. The ceiling worse condition was also significantly lower than the ceiling better conditions, but not the accurate display conditions. Data for the ceiling and GWIS interaction are provided in Figure 24.

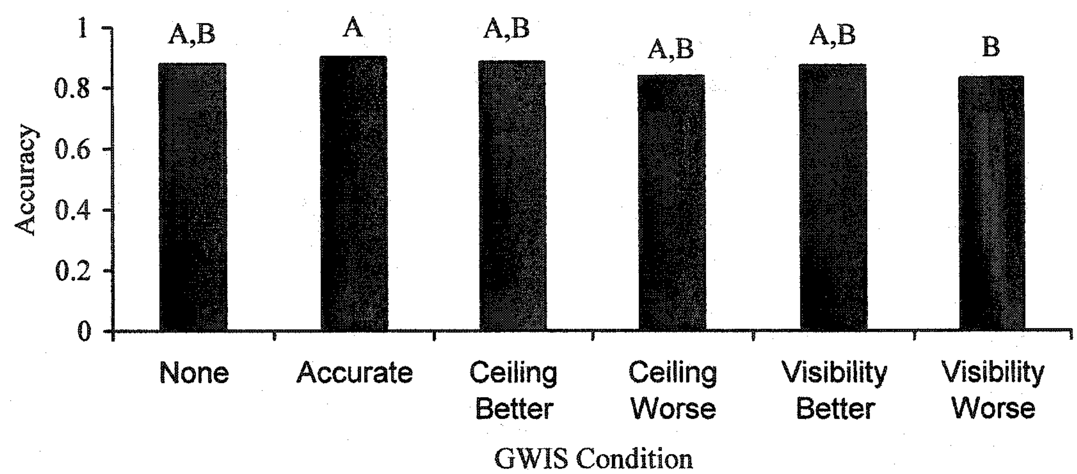


Figure 23. Accuracy data for the GWIS conditions. Means within each visibility and with different letters are significantly different at $p < .05$.

There was a significant interaction of GWIS and visibility on accuracy, $F(10,220) = 2.506$, $p < .05$. A test of simple main effects was done for visibility at each of the GWIS conditions. Only the visibility worse condition had a significant effect of visibility. Accuracy at the visibility worse condition was significantly higher at 2 miles visibility than 3 miles visibility. Neither 2 nor 3 miles were significantly different from the 5 mile condition. Data for the interaction are in Figure 25. The GWIS manipulation was not involved in any other interactions within the accuracy data.

Visibility RMSE. Inspection of the visibility RMSE data for the GWIS manipulation revealed a data entry error with one of the participants. The overall mean for visibility RMSE was 1.16 with a standard deviation of 1.15. The invalid entry had a visibility RMSE of 203, which is 175 standard deviations away from the mean. The entry was excluded from the subsequent analyses, as it is likely an entry error on the part of the participant. The visibility RMSE did not show either a main effect of GWIS or any interactions involving GWIS.

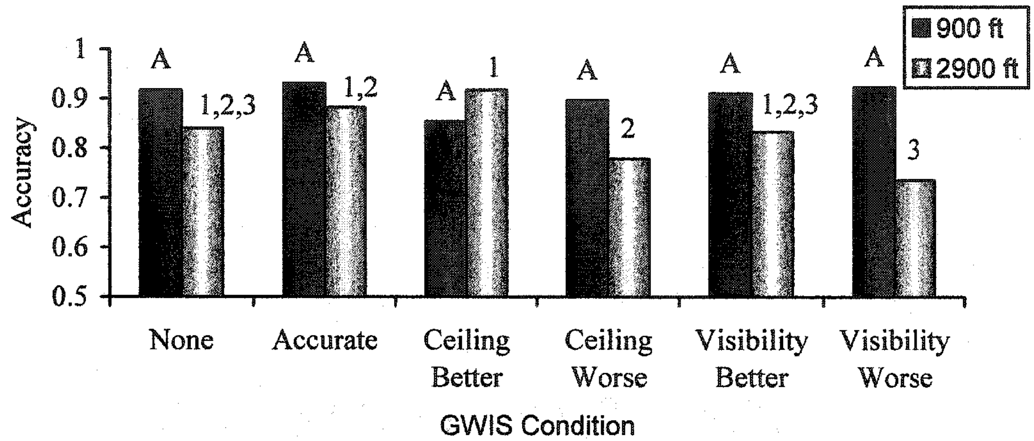


Figure 24. Accuracy data for the ceiling and GWIS interaction. Comparisons should only be made within each ceiling condition (i.e., 900 or 2900). Means with different numbers or letters are significantly different at $p < .05$.

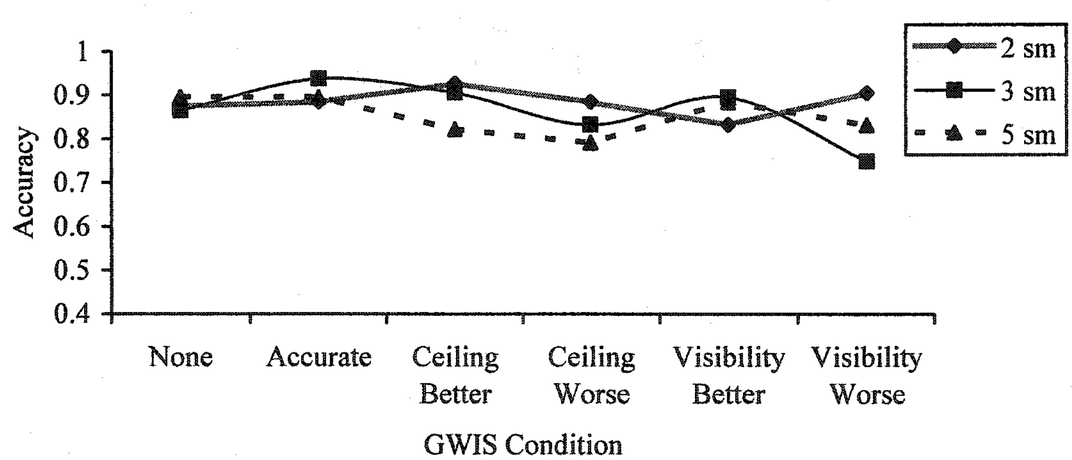


Figure 25. Accuracy data for the GWIS and visibility interaction.

Ceiling RMSE. Analysis of the ceiling RMSE data revealed a significant main effect of GWIS, $F(5,110) = 3.82$, $p < .05$. The ceiling worse condition had a significantly higher ceiling error than the accurate and visibility better GWIS conditions. The ceiling RMSE means and post hoc groupings for GWIS condition are provided in Table 22.

TABLE 22: Ceiling RMSE Data and Post Hoc Groupings across GWIS Conditions

	None	Accurate	Ceiling Better	Ceiling Worse	Visibility Better	Visibility Worse
Mean	709.896	590.625	698.872	839.080	604.688	649.306
Post hoc Group	1,2	1	1,2	2	1	1,2

Additionally the ceiling RMSE was involved in a significant 3-way interaction of rating x GWIS x visibility, $F(10,220) = 2.798$, $p < .05$. A test of simple interactions of rating and visibility was done at each level of GWIS. Only the ceiling better condition had a significant interaction of rating x visibility. A subsequent test of simple main effects of rating was done at each visibility within the ceiling better condition. Rating had a significant simple main effect at the 2 mile visibility condition, however it was not significant at any other level of visibility. Instrument pilots had a significantly higher ceiling RMSE within the ceiling better GWIS condition and at 2 miles visibility. The data for the 3-way interaction is provided in Figure 26. Ceiling RMSE was not involved in any other interactions.

Comfort. The ANOVA on the comfort data revealed a significant main effect of GWIS, $F(5,110) = 4.647$, $p < .05$. Post hoc analysis revealed that the visibility worse condition had a significantly lower level of comfort than the visibility better and ceiling better conditions. The ceiling worse condition was also significantly lower than the ceiling better condition, but not the visibility better condition. Data for mean comfort across each of the GWIS conditions are in Figure 27. Comfort was not involved in any interactions with GWIS.

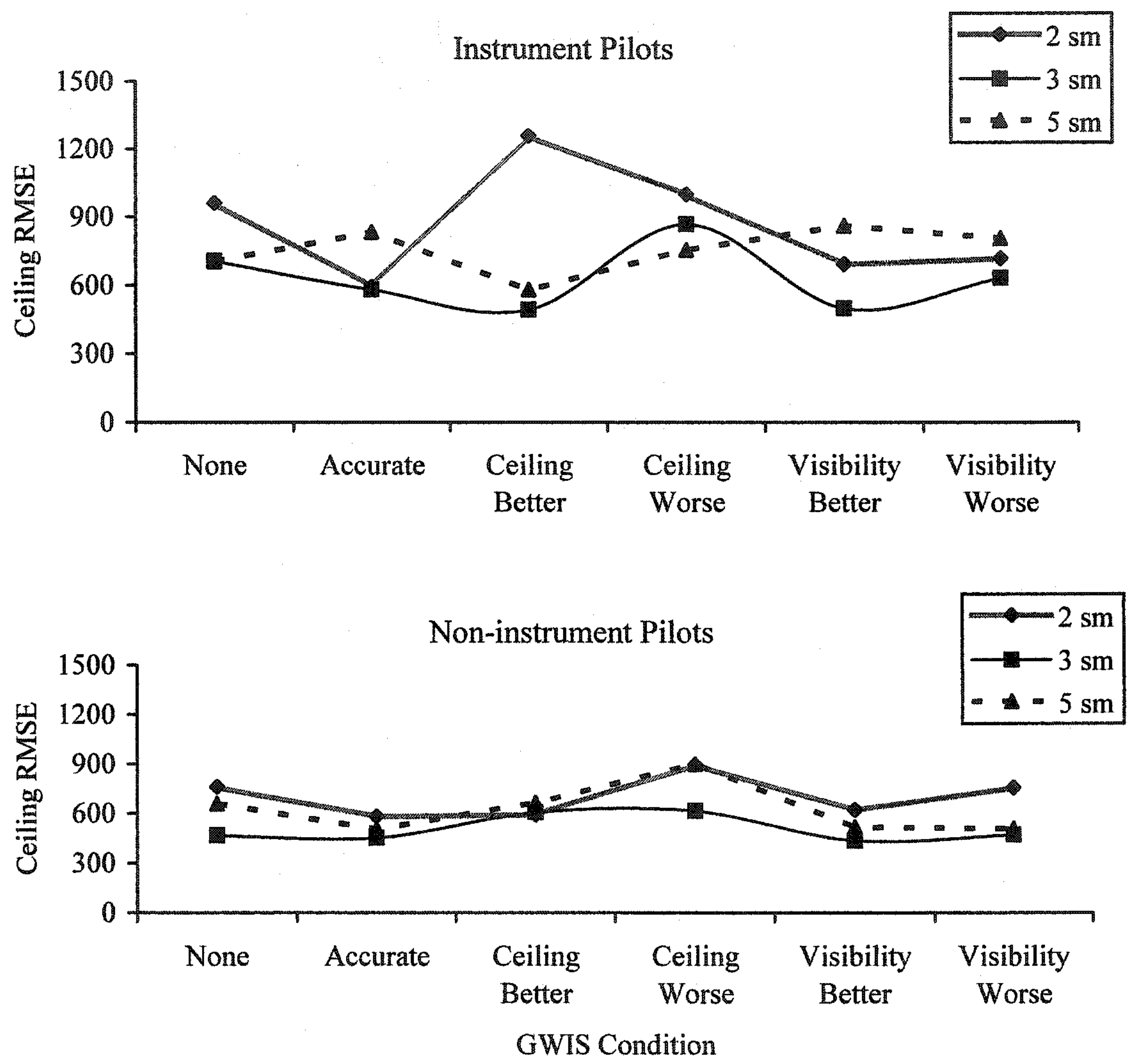


Figure 26. The interaction of GWIS by ceiling by rating on ceiling RMSE.

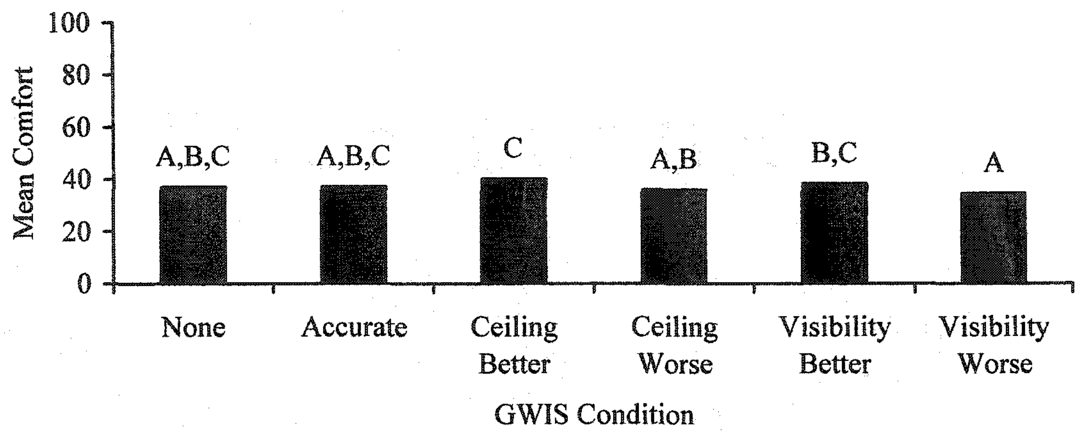


Figure 27. Mean comfort data for the GWIS conditions. Means with different letters are significantly different at $p < .05$.

Correlations

The second part of the analysis was a series of correlations between several compensatory and non-compensatory decision models and mean comfort levels used in previous research (Driskill et al., 1997; Hunter et al., 2003). The comfort levels were those obtained from both the experimental conditions and card sort task.

The different models were constructed from the standardized safety benchmark ratings used in Driskill et al. (1997). These benchmark ratings transform ceiling and visibility into a normal distribution with a mean of 5 and standard deviation of 1. The distributions for ceiling and visibility safety ratings are provided in Figures 24 and 25 respectively.

The values in the original Driskill study did not contain the necessary range of ceilings and visibilities used in the current study. Additional data points had to be estimated for the purposes of the current experiment. For example since no value existed for a ceiling of 4500 ft one was estimated from existing data points to be 6.33 (i.e., the value midway between Driskill's 4000 ft and 5000 ft points). Additionally, the values did not drop below the values established in Driskill et al. (1997) (i.e., the 400 ft ceiling condition took on the value of the lowest point on the original scale, that of a 600 ft ceiling, 3.664).

Additional models were computed incorporating the GWIS categories for ceiling and visibility. The models containing the GWIS conditions used the midpoint of each FAR category range (i.e., the midpoint of IFR ceilings is 750 ft). For ceiling LIFR was the value assigned to 600, 3.664 (i.e., the lowest value on the original scale). Since the VFR category has no maximum the midpoint of the VFR values used by Driskill et al. (1997) was selected as the value for the VFR GWIS. The GWIS VFR value for ceiling was that of a 4000 ft ceiling and the corresponding value for visibility was that of 7 miles visibility. The values assigned to each of the GWIS ceiling and visibility categories are in Figures 28 and 29.

The correlations are divided into those obtained from the card sort task, those obtained from the out the window, and those obtained from the GWIS out the window conditions. Due to the inaccuracy of pilots' estimation of ceiling and visibility both the actual and pilots' estimated ceiling and visibilities were used to construct the different models. The actual conditions were those presented by the computer (e.g., a ceiling of either 400, 900, 2900, or 4500 ft). The estimated conditions were what the pilots believed the conditions to be. The correlations' obtained from the baseline conditions and card sort task are presented in Table 23. All correlations were significant ($p < .05$).

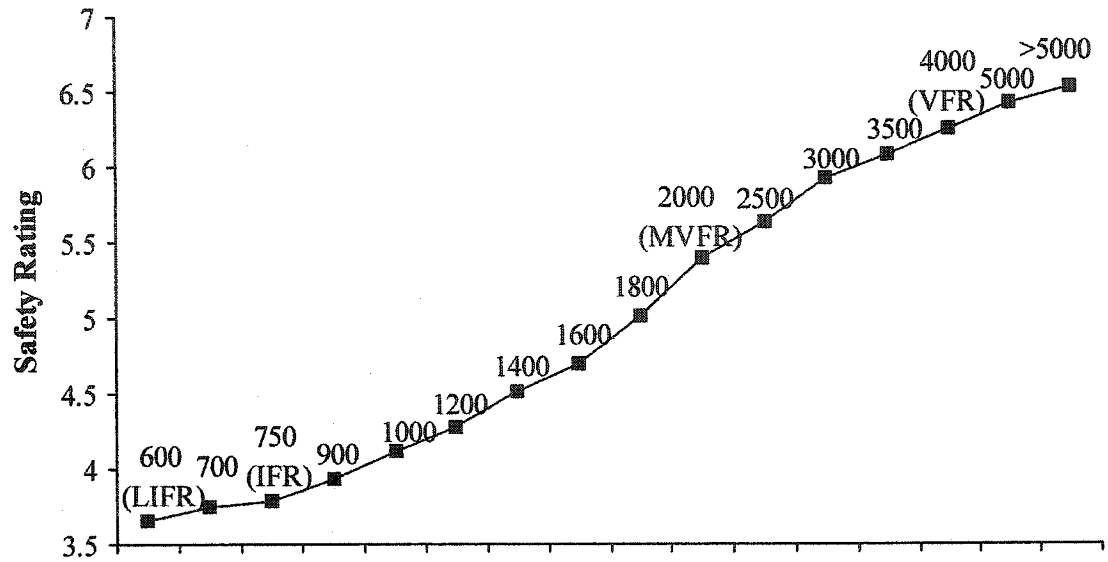


Figure 28. Safety ratings of ceiling adapted from Driskill et al. (1997). Values in parentheses are the GWIS category values.

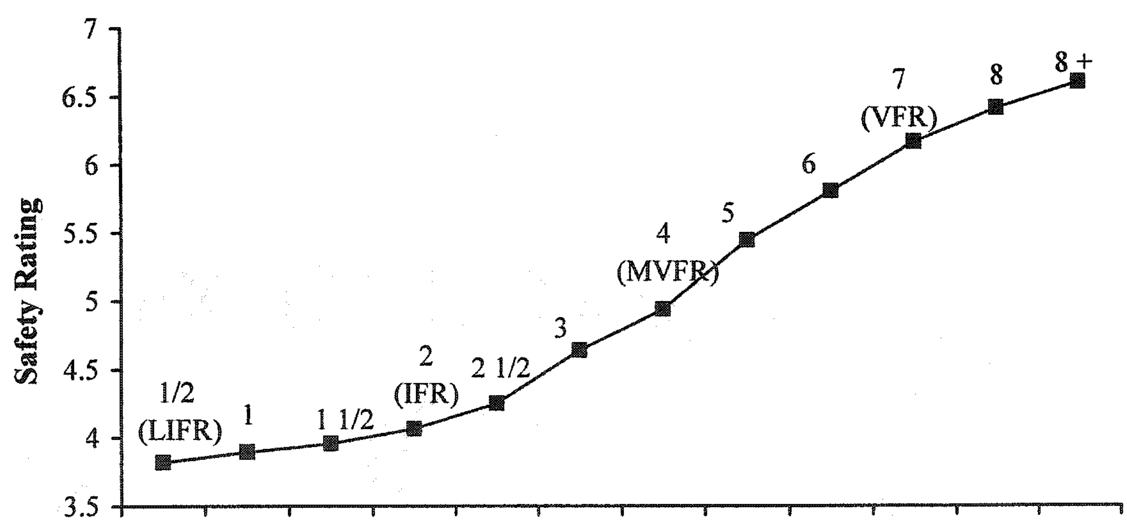


Figure 29. Safety ratings of visibility adapted from Driskill et al. (1997). Values in parentheses are the GWIS category values.

Additionally, the models incorporating the GWIS utilized only the 900 ft ceiling and 2900 ft ceiling, and the 2, 3 and 5-mile visibility conditions. This represents a restricted range of values for both the ceiling and visibility correlations. Such a restriction of range will often result in an attenuation or reduction of the obtained correlations compared with correlations using the full range of ceiling and visibility levels. All correlations were significant ($p < .05$).

TABLE 23: Comparison of Compensatory and Non-Compensatory Models. (All correlations were significant ($p < .05$))

Model Used	Window Conditions	Obtained Correlation (r)
Non-Compensatory Models	Ceiling	.490
	Visibility	.441
	Worst Factor (C or V)	.622
Compensatory Models	Ceiling + Visibility (additive)	.658
	Ceiling * Visibility (multiplicative)	.661
	Pilot Estimates of Window Conditions	
Non-Compensatory Models	Ceiling	.620
	Visibility	.768
	Worst Factor	.788
Compensatory Models	Ceiling + Visibility	.809
	Ceiling * Visibility	.802
	Textual Conditions	
Non-Compensatory Models	Ceiling	.741
	Visibility	.378
	Worst Factor	.840
Compensatory Models	Ceiling + Visibility (additive)	.812
	Ceiling * Visibility (multiplicative)	.839

TABLE 24: Decision Models with the GWIS and Restricted Range. (All correlations were significant ($p < .05$))

	OTW Conditions	Correlation With Comfort
Non-Compensatory Models	Ceiling	.496
	Visibility	.331
	GWIS Ceiling	.358
	GWIS Visibility	.225
	Worst Factor (C or V)	.567
	Worst Factor (GWIS, C, V)	.490
Compensatory Models	Ceiling + Visibility (additive)	.581
	Ceiling * Visibility (multiplicative)	.584

DISCUSSION

Overall, the results of the present study reveal valuable insight into pilot weather judgment. The results provide additional support to the situation assessment hypothesis, as the data revealed that pilots have problems estimating weather conditions. In addition to being fairly inaccurate at their estimation, pilots' also had a tendency to overestimate conditions (i.e., estimate that conditions are better than they are). From a safety perspective this conservative response bias or tendency to respond that conditions are VFR is a potential hazard. Accident data suggested that IFR pilots might be better at estimating weather conditions (AOPA Air Safety Foundation, 1996). However, the problem in estimating conditions does not appear to be mitigated by the additional training IFR pilots have. Overall pilot rating only had a small impact on responses, and there only within an interaction with visibility in the visibility RMSE data.

The present study examined several decision making strategies that pilots might use when making weather judgments. The evidence to suggest the use of a compensatory model, as described in previous research (Driskill et al., 1997; Hunter et al., 2003) over a worst factor model was not substantiated. However, the interaction of ceiling and visibility (see Appendix G for additional analyses on this interaction) that prevailed throughout the data suggests a different type of compensatory model.

The study found mixed results for the use of the graphical METAR information. The overall SDT results suggest that pilots use the GWIS information only when it suggests that conditions are worse. There was no significant impact on the overall GWIS measures when it suggested either ceiling or visibility was better. However, the examination of the ceiling and visibility SDT measures increase pilots' tendency to overestimate conditions when the GWIS contained conditions that were better than those depicted.

Weather Estimation

One of the major objectives of the experiment was to provide some empirical evidence of pilots' ability to estimate weather conditions. Previous research has suggested that pilots are poor at estimating weather in flight (Goh & Wiegmann, 2001b). However, this evidence was taken from a single IFR weather estimation. The present study set out to systematically examine how pilots estimated different weather conditions. Several metrics of pilot performance in estimating weather were created through a number of questions. Two of these performance metrics were the ceiling and visibility root mean square error (RMSE) for both ceiling and visibility.

The first hypothesis was simply to confirm previous findings about pilots' ability to estimate weather conditions within a more systematic experiment. Ceiling RMSE ranged from about 650 feet to 1800 feet across the four ceilings used in the current experiment. The errors were not as large as the 2200 ft error obtained in previous research (Goh & Wiegmann, 2001b). However, the errors are still problematic to flight safety as they often resulted in pilots misinterpreting IFR conditions as VFR. The ceiling SDT analysis (described later in more detail) revealed a strong tendency to overestimate the cloud height. The bias was so strong that it resulted in an average crisp hit rate for ceiling below 35% and several pilots actually had 0% hit rates.

Pilots were also tasked with estimating visibility. Previous research found pilots were on average off by 1.5 miles in their estimations (Goh & Wiegmann, 2001b). Overall the current study yielded similar errors in visibility estimation. Across the four visibility conditions visibility RMSE ranged from .65 miles to 2.5 miles, with the error increasing as visibility increased. The estimate from previous research was generated at low visibility (i.e., 2 miles), and the corresponding estimate at 2 miles in the present study had an error of only .65 miles.

It is not surprising that the estimates for visibility and ceiling were somewhat better in the present study. Pilots in the current experiment had more knowledge of the simulated area, and a greater focus on the weather. Pilots only task during the experiment was to estimate and evaluate the weather conditions. Before the experiment began, they were shown the area on a clear day and were made aware of two landmarks and the distances to those landmarks. Landmark information was given to keep the experiment similar to pilots' knowledge of a familiar area. Another advantage was that pilots had the video clip of the weather repeating in a continuous loop while answering the questions. In the University of Illinois study (Goh & Wiegmann, 2001b) pilots were flying the simulator and were asked questions regarding ceiling and visibility after the simulation was complete. The large errors found in the present study are a testament to the difficulty of estimating weather conditions.

The first hypothesis of obtaining errors similar to prior research (Goh & Wiegmann, 2001b) was only partially supported. Some of the errors generated within the current experiment were smaller than those previously found. However this may again be due to the emphasis on weather estimation within the current experiment, the use of landmarks, and the constant presence of the weather stimulus. Despite being less substantial than errors previously obtained, the errors in the current experiment are still cause for concern since as the SDT analysis revealed they tended to be overestimates of the weather conditions.

Pilot Sensitivity and Response Bias

This experiment was the first to apply SDT to pilot weather estimation. The theory has been applied to a range of settings where an individual would have to distinguish between two possible states of the world, such as VMC and IMC flight conditions. However, a new extension of the theory (Parasuraman et al., 2000) has allowed for analysis of conditions that vary along a continuum. This fuzzy extension of SDT was created to allow for a more precise measure of sensitivity (Hancock et al., 2000; Parasuraman et al., 2000); however, there have been only limited applications of the fuzzy technique to date. The analysis of weather conditions provides a unique application of SDT. In addition to having two distinct states (i.e., IMC and VMC), and varying along a continuum, weather can actually be considered to vary along multiple continuums (i.e., ceiling and visibility). To fully appreciate pilot weather judgment a series of separate SDT analyses were performed.

The first SDT analysis was based upon the traditional or crisp version of the theory (Green & Swets, 1966). The overall weather was classified based upon the worst condition, overall conditions were either VMC or IMC. On average pilots had an A' of .917 and a small conservative response bias. On average pilots missed just under 20% of the IFR conditions. However, pilots varied in their sensitivities and biases. Although overall there was a slight conservative bias (i.e., a positive c), eight of the twenty-four pilots actually had a liberal response bias (i.e., a negative c). A complete breakdown of the 24 pilots and their sensitivities and response biases is provided in Appendix F. When considering flight safety it is best for pilots to respond more frequently that conditions are IFR (i.e., have a liberal response bias), however only eight of the pilots had this type of bias. That pilots on average had a conservative bias or tendency to respond that conditions were VFR is a potential flight safety problem.

A second analysis was based upon fuzzy SDT (Parasuraman et al., 2000). The overall weather was again based upon the worst condition, however it took on a range of values from 0 to 1. The fuzzy responses were generated from the question asking pilots to place the weather condition along a continuum containing LIFR, IFR, MVFR and VFR. On average pilots had an A' of .922. The fuzzy analysis also suggested the pilots had a small conservative response bias. However, there was again variability in individual pilot sensitivities and biases. Four of the pilots had a liberal response criterion (i.e., they tended to respond that conditions were IMC).

The inclusion of the ceiling and visibility estimation questions offered an opportunity to apply separate SDT metrics based solely upon ceiling and visibility. These analyses provided a better understanding of how pilots were making weather estimates. Separate crisp and fuzzy analyses were done on both a ceiling and visibility dimension. The results of the crisp analysis on ceiling revealed that on average a pilot's ability to respond that IFR conditions were IFR (i.e., HR) was below 40%. That is, more than 60% of the IFR conditions were categorized as VMC. The corresponding error, categorizing VMC conditions as IMC (FAR) occurred less than 5%. The result was a low sensitivity (i.e., poor ability to distinguish between VMC and IMC) and a liberal response criterion of $c = 1.146$. Taken alone the poor sensitivity and tendency to overestimate ceilings could be a serious problem. Even more troubling was that four of the pilots demonstrated no sensitivity (i.e., A' = .5) for the crisp ceiling analysis. That is, the pilots never estimated the ceiling as being below 1000 feet when the ceiling was in fact at 400 and 900 feet.

Results from the fuzzy ceiling analysis provide a different picture of the pilots' sensitivities and biases. The fuzzy analysis shows that pilots had an A' of .897 and a c of .329. This represents a rather large increase in sensitivity and decrease in response bias compared with the crisp c results. The fuzzy data yielded hit and false alarm rates of 75% and 10% respectively. This represents an increase of over 35% in hit rate from the fuzzy analysis with only a corresponding 5% increase in false alarm rate.

Considering either the fuzzy or crisp analysis for ceiling alone would provide a different and incomplete understanding of ceiling judgment. To interpret these results, it is helpful to understand how pilots were making their judgments and how fuzzy and crisp logic treated MVFR responses during IFR conditions. Understanding they were flying at 2400 ft, pilots were fairly good at correctly identifying ceilings that were VMC. This is likely due to the fact that two VMC ceilings (i.e., 2900 and 4500 ft) were always above the aircraft. There was little difference between the crisp and fuzzy analyses in their false alarm rate, which suggests the differences between the two did not stem from the estimation of the VFR (and MVFR) conditions. The difference in the analysis most likely resulted from how each analysis considered LIFR and IFR conditions with MVFR responses. If a pilot responded that IFR conditions were VFR, the crisp analysis would mark the trial as a (complete) miss whereas the fuzzy analysis would give the pilot a partial hit depending on how low their estimate was. For example if a pilot estimated a 900 ft ceiling as being 1500 ft, the crisp analysis would consider the response as a 0 for hit and 1 for a miss. The same estimates would be scored a hit of .64 in the fuzzy analysis and a miss of .36. If pilots consistently overestimate ceiling height the crisp analysis would show decreased sensitivity and shift the criterion to be more conservative. The same estimates modeled with the fuzzy analysis would still shift the criterion, however the shift would be smaller and the sensitivity higher.

The differences in the fuzzy and crisp ceiling analysis represent one of the most stark differences between the two techniques and provides support for previous suggestions that researchers use both (Masalonis & Parasuraman, 2003; Parasuraman et al., 2000). Neither analysis provides a complete picture of ceiling judgment.

The results of the crisp and fuzzy analyses of the visibility were more consistent than that of ceiling. However, they were still different. The crisp visibility analysis indicated that pilots had a sensitivity of .853 and a response bias of .239. The fuzzy analysis revealed a lower sensitivity and bias, .761 and .177 respectively. It is likely that the difference comes from the marginal conditions with overestimated responses. A pilot responding that visibility was 4 miles when it was in fact only 3 miles would not be penalized for sensitivity in the crisp analysis, but since the conditions border on being IFR, they would receive a hit of .425 and a miss of .425 in the fuzzy analysis.

As with the ceiling analysis it seems that both a fuzzy and crisp analysis are necessary to understand pilot judgment of weather conditions. Overall, the data seems to suggest that pilots do have problems estimating weather conditions and that their responses are biased such that they tend to overestimate conditions. The second hypothesis, that overall pilots would have a conservative response bias, was supported. In the overall, ceiling, and visibility analyses the pilots exhibited conservative biases regardless of fuzzy or crisp technique. This potential for a conservative bias is a cause for concern particularly since it occurred in the absence of any external or background factors such as time, money or passengers. According to Jensen's model (1995) these background factors would have an additional influence on pilot's response bias. However these judgments are based upon averaged data and there were in fact pilots in the present study who tended to underestimate conditions as well.

Pilot Rating

The third hypothesis that instrument pilots would be better than non-instrument pilots was not supported. It was expected that the extra training required to obtain an instrument rating accounted for the differences between the incidence in which the two groups are involved in VFR into IMC accidents (AOPA Air Safety Foundation, 1996). There were no main effects of rating on any of the performance metrics or SDT sensitivity metrics. There was a significant interaction of rating and visibility on visibility RMSE where instrument pilots had a lower RMSE than non-instrument pilots within the 10-mile visibility conditions. Although this difference does favor the instrument pilots it is not fully consistent with the hypothesis. The hypothesis suggested instrument pilots would have more experience with IFR conditions and therefore a better ability to recognize them. The 10-mile visibility conditions are (based only upon visibility) VFR conditions.

It is most likely that the lack of differences between the instrument and non-instrument pilots are a result of the matching of the two groups on cross-country hours. That is, it is more likely that any difference between the groups that appears in accident data are the result of flight experience and not IFR training. Previous research has found that the number of cross-country hours is a predictor of pilot ability (Wiggins & O'Hare, 1995). However, an inspection of correlations with data from the present study found non-significant negative relationships between both cross-country hours and overall hours with pilots' ability to distinguish between conditions (i.e., sensitivity). The lack of correlation was likely due to the both the number of participants and reduced range of flight hours and cross-country hours. The only data to suggest that instrument pilots would be better came from indirect accident data indicating that 75% of pilots involved in VFR into IMC accidents were not instrument rated (AOPA Air Safety Foundation, 1996).

If there were no differences between the two groups in their ability to distinguish between the two conditions (as found in the present study), an alternative explanation of the AOPA findings might be that the instrument pilots had a more liberal response bias than the non-instrument pilots. The current study did find a significant difference between the two groups in their bias in estimating visibility. The direction of the fuzzy analysis would support such a hypothesis (the instrument pilots had a significantly less conservative response criterion). However the results of the crisp analysis for visibility were also significant with the direction of the difference being reversed. It is not clear why the SDT analyses of visibility would return one set of results for the crisp data and a different set for the fuzzy set. Such results provide further caution in using a fuzzy analysis as a replacement for a crisp analysis.

The AOPA findings that non-instrument rated pilots are more likely to be involved in a VFR into IMC accident is more likely the result of experience. Obtaining an instrument rating requires additional training. The pilots that were selected for the current study represented the lower end (in terms of cross country hours) of the available

IFR pilots, whereas the selected VFR pilots were among the higher end of the available pilots. The AOPA accident data may be a result of IFR pilots having the tendency to have more cross-country experience. In addition to the IFR rating requiring additional training it would also allow those pilots to fly in a wider range of conditions, and therefore have greater opportunity to fly. Unfortunately the previous literature (AOPA Air Safety Foundation, 1996) does not discuss cross-country hours within the different ratings. Separately addressing the issue of ratings and experience would require a large number of participants, and would require a considerable database to match pilots within both groups on experience. Future research should continue to examine the issue of pilot ratings, experience and judgment.

Window Based Decision Making Strategies

The present experiment conducted an analysis similar to that used in previous studies of pilot weather decision making (Driskill et al., 1997; Hunter et al., 2003). This previous literature concluded that pilots' comfort in continuing flight was most correlated with compensatory decision strategies. That is, pilots appeared to take a mathematical average of all of the different weather conditions when making an overall rating of comfort. This would allow pilots to compensate low ceilings for high visibility when deciding comfort. The present study suggested that the results might be different if pilots were presented with a visual depiction of weather conditions as opposed to textual weather information. It was expected that pilots would actually use a decision model based upon the worst condition when given the overall visual. Such a model would be consistent with Federal Aviation Regulations.

There were several issues in calculating the decision model. The largest obstacle was pilots' inaccuracy in estimating weather conditions. Due to the estimation inaccuracy two series of models were created, one based upon what the conditions actually were and the second based upon the pilots' estimates of those conditions. Overall the use of the pilots' estimates of the weather conditions increased the correlations with comfort. Both the estimated and actual conditions show the two compensatory models having higher correlations with comfort than the worst condition model. However, the magnitude of difference between the worst condition model and the better of the compensatory models is only .04 for the actual conditions and .02 for the estimated conditions.

It is difficult to conclude that pilots use a compensatory decision model as opposed to the worst factor model. Unfortunately based upon the results it is not clear what model is most representative of pilots' decision making.

The nature of the weather presentation also creates another problem in comparing the results with those obtained through textual conditions and previous literature (Driskill et al., 1997; Hunter et al., 2003). The FAA studies, actually presented pilots with all of the weather condition combinations cards at once and then asked them to sort them based upon their comfort. Then after sorting the cards, pilots were asked to provide a rating of comfort. The present study simply gave pilots a weather depiction and asked them to provide a rating for comfort. Weather combinations were presented one at a time because presenting 16 videos of weather conditions simultaneously was beyond the resources available for the current experiment. A more direct replication of the FAA studies would have had the pilots have all conditions to be sorted simultaneously.

Despite limitations in a direct comparison between the current methodology and previous literature, it appears that the previous conclusion that pilots follow a compensatory decision making model should be further examined. In addition, the method for validating the pilots' decision model should also be given further consideration in future research using both textual and visual presentation of weather.

Card Sort Task

Previous research (Driskill et al., 1997; Hunter et al., 2003) had found that pilots followed a compensatory decision making model as just discussed. Although the previous hypothesis questioned these results because of the methodology used to obtain them it was expected that using a similar text based card sort task the results could be replicated. However, the results in the FAA study were not replicated within the present study. Results of the present study indicated that within the textual ratings the worst factor model was most highly correlated with a pilot's judgment. However, this correlation was only different by .001 and .028 compared to the multiplicative and additive compensatory models respectively. Although this difference does not necessarily suggest pilots did follow a worst factor model it does not provide strong support for the FAA's conclusions from previous research that suggest either compensatory model would be the most accurate. There are several potential differences between the current experiment and previous literature that may account for this lack of strong support.

One major difference between the present study and the previous FAA research was the inclusion of precipitation information. Both FAA studies used varying levels of precipitation in addition to varying ceiling and visibility. However, because the conditions also had to be rendered graphically, a decision was made not to include a precipitation manipulation in the current investigation. This makes it difficult to directly compare the results obtained in the current study with those obtained in previous work. It is not clear if the inclusion or exclusion of precipitation would change the pilots' decision making strategy.

An additional consideration is that the pilots in the present study were specifically trained to use a worst factor model. In their training for the experimental paradigm pilots were reminded of FAR definitions for categories which is consistent with the worst factor model. It is possible that training the pilots to make this judgment in the experiment impacted their strategies in the card sort task.

It is not clear if either of the two methodological differences accounted for the inability to replicate the data from previous research. However, combined with the data and models generated from the out the window images it does suggest that the issue needs further examination.

Influence of the Graphical METARs

The sixth hypothesis that the graphical METARs would have an impact on pilots' weather judgments was supported. In combination there were 12 different signal detection metrics calculated ((fuzzy/crisp) x (overall/ceiling/visibility) x (sensitivity/bias)). Of these 12 metrics, the GWIS had a significant effect on all 6 of the sensitivity measures and 4 of the bias measures.

With respect to the overall crisp sensitivity only one of the inaccurate display conditions (i.e., the visibility worse condition) had a significantly lower sensitivity than the no-display conditions. The accurate display, two better display conditions, and the ceiling worse condition were not significantly different from the no-display condition. The overall fuzzy sensitivity data revealed similar trends. The visibility-worse condition resulted in significantly poorer sensitivity than the no-display condition. None of the other display conditions resulted in different crisp sensitivity values relative to the no-display conditions. The crisp and fuzzy data that reveals a lower sensitivity for the visibility worse condition was likely obtained because pilots with the display were more likely to incorrectly report that conditions with 3-mile visibility (which is at the boundary between VMC and IMC) were IMC. This type of an increase in FAR would be reflected in the response bias.

The overall crisp and fuzzy response criterion also had significant main effects of GWIS. The crisp bias data indicated that the visibility worse condition significantly shifted pilots' response criterion compared to every other condition with the exception of the ceiling worse such that they lowered their estimates of visibility. The fuzzy bias confirmed these results finding both the ceiling worse and visibility worse conditions to be significantly different from all of the other conditions.

Taken together the overall sensitivity and response bias paints a positive picture of pilots' use of the GWIS. Although the inaccurate display of information did serve to lower the pilots' sensitivity it was in the direction of improved safety. The pilots did not allow the displayed information that was better than the out-the-window conditions to bias their answers. Pilots did however demonstrate a greater tendency to respond that conditions were IFR (i.e., a liberal response bias) when the display portrayed conditions worse than those out the window.

The separate ceiling and visibility SDT analyses had more mixed results as to the impact of the GWIS. With respect to ceiling, the fuzzy sensitivity analysis did not reveal any differences between the GWIS conditions. The crisp analysis indicated that none of the display conditions had a significantly lower sensitivity than the no-display conditions. However, the ceiling worse conditions yielded a significantly higher sensitivity than the no-display condition. Both the crisp and fuzzy analysis of ceiling response bias indicated that the ceiling better conditions significantly increased response bias (i.e., a conservative shift) compared to the no-display conditions. Conversely, the ceiling-worse display condition did have a significantly lower response bias in the crisp, but not the fuzzy analysis.

The visibility analyses also provided some results that suggest that the GWIS could lead to poorer decisions by the pilots. The crisp visibility sensitivity analysis found that the visibility better condition had a significantly lower sensitivity than the no-display condition, however this difference was not present in the fuzzy analysis. Additionally within the crisp analyses the visibility better condition had a significantly more liberal response criterion than the no-display condition. The no-display condition was not different from any other display condition in the fuzzy response bias analysis.

Unlike the overall SDT analysis the separate ceiling and visibility analyses revealed that the GWIS could have a negative impact on pilot decision making (i.e., creating a tendency to estimate conditions as better than they are.) When looking only at pilots' ability to judge ceiling, the ceiling-better display further shifted the pilots' already conservative response criterion in both the fuzzy and crisp analysis. Without any GWIS information pilots had a conservative bias and were on average likely to overestimate ceiling conditions. However, the potential for the graphical METAR information to add to the problem of overestimating ceiling could be dangerous. On the positive side, the crisp ceiling bias was less liberal when the display depicted the ceilings as being worse. The same trend was present when looking at the visibility SDT analysis. When the display suggested the visibility was better than it was there was a reduction in the pilot's crisp sensitivity and an increase in crisp bias when compared to the no-display condition.

Regardless of the positive or negative impact of the GWIS display it did factor into the pilot's weather judgment. If the GWIS is to have a positive impact on pilots' weather decision making, they should be trained in how to use the information.

Fuzzy Signal Detection Theory

The process of analyzing weather data using fuzzy and crisp signal detection theory provided a unique comparison of the two paradigms. Several papers have been published highlighting the potential benefits of fuzzy signal detection theory over crisp signal detection theory (Hancock et al., 2000; Masalonis & Parasuraman, 2003; Parasuraman et al., 2000). Ultimately the current analysis arrives at the same conclusion as previous research that both techniques should be used. However, the current study had perhaps the most significant differences in the values generated by a fuzzy and crisp analysis. This study found some benefits and drawbacks to fuzzy SDT not discussed in prior literature.

A major difference between the two types of analysis was in the incidence of extreme scores. Extreme scores are those in which the observer has hit and false alarm rates that are equal to either 1 or 0. These extreme scores can be due to either perfect discrimination or sampling variability (Hautus, 1995). Typically perfect discrimination can be ruled out, particularly in the present experiment where there were several instances of hit rates that were at 0.

There were fewer extreme scores when using the fuzzy analysis. The crisp data from the baseline trials contained 11 extreme scores, whereas the same fuzzy data did not contain any extreme scores. Particularly in the case of a strong bias it is easy to understand how extreme scores can be obtained in a crisp analysis. However, with the fuzzy analysis and an imperfect observer there is a greater chance for having some error on every trial. The fuzzy analysis was created as a more precise measure of sensitivity (Parasuraman et al., 2000). This added precision helped to reduce the possibility of obtaining extreme scores in the current study.

Unfortunately this added precision in the sensitivity metric is not without its drawbacks. There were substantial differences in the results obtained from the crisp and fuzzy analysis of ceiling sensitivity. The current study originally suggested that although the fuzzy analysis might make gains in being a more precise metric of sensitivity it does so by reducing variability in response bias. The crisp analysis of ceiling shows that, pilots were very poor at distinguishing between IMC and VMC ceilings. The participants were very good at determining VMC ceilings; however, they determined less that 40% of the IMC ceilings to actually be IMC. The participants had a strong tendency to overestimate the ceilings. Four observers actually missed every IFR ceiling that was presented. As a result the overall sensitivity within the ceiling analysis was low, an A' of only .746. The low hit rate was coupled with a low false alarm rate resulting in a fairly conservative response bias of 1.146.

The crisp analysis suggests pilots are poor at differentiating conditions and have a tendency to overestimate conditions. The fuzzy analysis indicates that pilots were inaccurate at estimating the conditions however their estimates did correspond to the actual conditions. Estimating a ceiling to be 1500 ft when it is actually 900 would result in a miss within the crisp analysis, but a partial hit within the fuzzy analysis. The fuzzy analysis suggests that pilots are better able to discriminate conditions (i.e., an A' of .897) and have a less conservative response bias (i.e., a c of .329) than the crisp analysis.

The two different explanations of sensitivity and response bias determined by the fuzzy and crisp analyses ultimately leads to the question of which technique is better. Given the disparity in results obtained it is a sound question. Both analyses provide valuable insight into how pilots estimate ceiling. The arguments for the use of fuzzy SDT suggest that it provides a more precise metric of sensitivity. Based upon the ceiling results this is a reasonable argument. The poor sensitivity suggested by the crisp analysis may indeed hide the true picture that pilots were still relatively close in their estimates of cloud height. However, with this more precise measure of accuracy there was a substantial reduction in response bias. The fuzzy analysis does not show how, when forced to categorize ceilings as VMC or IMC, pilots were more likely to decide that ceilings were VMC. The differences in the obtained results only reaffirm prior recommendations that both fuzzy and crisp analysis should be performed. Both analyses added unique explanations of events, and as such the present study found some cause for questioning the meaning of each analysis alone.

Ceiling and Visibility Interaction

Although it was not anticipated there was a significant interaction of ceiling and visibility on all of the dependent variables in the baseline analyses. As a result of the prevalence of the interactions several additional analyses were performed (See Appendix G).

The nature of the interaction was such that there was a significant drop in accuracy for the two IFR ceilings at 10 miles visibility. This combination of low IFR ceilings and high visibility should have been interpreted as IFR. However, accuracy at 400 ft with 10 mile visibility was 60% whereas accuracy with 900 ft ceilings and 10 mile visibility was 40%. This means that despite the IFR ceilings pilots erred and interpreted the conditions to be VMC. What is also unusual is that the problem was not present at the 5-mile visibility conditions as pilots' estimates for the 400 and 900 ft ceilings were 90% and 80%.

The visibility estimation data (in Appendix G) shows that as the ceiling increased pilots' estimation of visibility also increased. Pilot's estimation of visibility was always below the actual visibility within the two IFR ceiling conditions. However, as the ceiling increased (i.e., 2900 and 4500 ft) pilots' estimate of visibility increased. The two IFR ceilings revealed significantly lower estimates of visibility compared to the two VMC ceilings within each level of visibility. The average estimates for ceiling were above the actual ceilings in the baseline data. The data within Appendix G reveal that as the visibility increased pilot's estimates of cloud height also increased at the 400, 900 and 4500 ft ceilings.

It is not clear as to why pilots' estimations of ceiling would be influenced by visibility, or why ceiling would influence visibility estimates. One potential explanation is that pilots do use a compensatory decision making model as suggested by previous authors (Driskill et al., 1997; Hunter et al., 2003). The pilots may be using the high visibility to compensate for the low ceilings. The compensatory model suggested by the previous literature suggested that when a pilot knew both the ceiling and visibility he or she would use a mathematical combination of the two values to determine their overall comfort.

However in the current study, the exact ceilings and visibility are not known and the pilot must make an estimate. It may be that the pilot evaluates one dimension, and this in turn may impact his or her response bias for the second dimension. Based upon the examination of the results of the SDT analysis in Appendix G this appears to be the case. For example, if pilots estimate that the ceiling is VMC then their bias for visibility may in turn experience a conservative shift. This is in fact the case in the visibility response bias data for both crisp and fuzzy SDT. Pilots reveal a liberal bias for visibility when ceilings are IFR, and a conservative response bias for visibility when ceilings are VMC. The fuzzy ceiling bias data also suggests that as visibility improves there is a conservative shift in response bias.

From a safety perspective a conservative shift in response criterion for one dimension due to high values in the other dimension is a potentially serious problem. Previous research suggested pilots use a compensatory strategy to average the different dimensions when making an overall judgment. The compensatory model was based upon the actual values of ceiling and visibility. The data generated from this study suggest a problem earlier in the decision process. Pilots may have a bias in estimating one dimension based upon levels of the other dimension. In forming their estimate of visibility pilots demonstrate a conservative or liberal bias based upon the actual ceiling.

That is, before pilots even have the two pieces of data necessary to make an overall decision they may have already been biased in their estimation. The bias suggested by this experiment does not exclude the compensatory decision model suggested by previous research (Driskill et al., 1997; Hunter et al., 2003) but, may actually occur in addition to it. Future research should further investigate the ability for estimates in one dimension of weather to influence response bias in the other dimensions of weather.

2-D Simulation of 3-D

The validity of generalizing results obtained in a 2-D simulation of weather conditions to the real 3-D world must be examined. Several arguments can be brought up in the defense of the 2-D methodology used in the present study. The first argument is that the present study also included the dimension of time or apparent motion. Pilots were not asked to make judgments based solely upon a static 2-D depiction. Motion (and apparent motion) provides additional cues to depth that would have aided pilots' judgments. The most notable cue is motion parallax. Motion parallax is a monocular depth cue in which the speed and direction of objects are based upon their distance from the observer. The clouds rendered in the experiment were broken. The holes in the cloud layer would have given the pilot a cue as to how far he or she was from the clouds. The faster the hole appears to move the closer the clouds are.

A second argument comes from the estimation of height in 2-D versus 3-D displays (Dixon & Proffitt, 2002). Research has investigated the use of 2-D, 3-D and the real world estimates of height as a function of display size. The authors conducted a study of the vertical horizontal illusion where vertical objects are overestimated when asked the height of the vertical object using a horizontal scale. Previous research (Yang, Dixon, & Proffitt, 1999) had found that the magnitude of the overestimation was larger (and closer to perception in the real world) for 3-D versus 2-D displays. However more recent research (Dixon & Proffitt, 2002) has found that the magnitude of estimation errors is related to the size of the display and not whether it is 2-D or 3-D. The larger the 2-D or 3-D image the closer the visual system will be to perception of height in the real world. In the current investigation the weather conditions were presented with a projector on a screen at 34.75 x 26 inches and were therefore relatively large when compared to other visual displays.

Although the use of a 2-D display is a limiting factor in the present experiment the inclusion of motion and the use of a projected image would have aided in the realism. Motion was included at the recommendation of several pilots during a pretest phase. Additionally the inclusion of terrain features based upon satellite imagery also eliminated problems based upon indistinct terrain features.

Limitations

The current study contained a few limitations that may have had an impact on the results obtained and their generalization to the larger GA population. The pilot population selected represents a relatively homogenous group of GA pilots. This homogeneity resulted from matching of instrument and non-instrument pilots on cross-country hours. Pilots were intentionally matched to isolate the impact of instrument training. This typically resulted in the selection of lower hour instrument pilots and higher hour non-instrument pilots. None of the pilots selected for the study had over 1000 flight hours. Based upon cross-country hour cut offs used in previous research

(Wiggins & O'Hare, 1995) all of the pilots were novices. Future research should incorporate a wider range of pilot experience.

On an actual flight, pilots typically acquire weather information from a variety of sources. This helps them to build a mental model of current weather conditions and trends. Having this model of the weather conditions helps pilots set expectations, frame and interpret what they see out the window. Although in the instances of inaccurate forecasts this can lead to problems, it usually aids pilots in their judgments. This study intentionally focused on in flight point estimation of weather, and therefore no prior weather information was provided. In a typical flight pilots would have this information and it would likely affect their bias in interpreting the weather conditions.

Although the use of a simulator is a limitation in the present study, it would have been impossible to use an aircraft. The precise measurement of the actual ceiling and visibility at the aircrafts location would require the aircraft flight path be next to weather equipment. Replications would require the cooperation of weather and the ability to take each pilot on multiple trips to see the same conditions. Several pretest pilots were brought in to ensure the depictions had a high level of realism.

These limitations were the result of the specific questions being addressed within the present study. Eliminating variability such as flight experience and available weather information was necessary. However, future research may be needed to more fully examine the impact these variables have on pilot judgment.

Summary and Conclusions

What does the current study reveal about accidents due to VFR flight into IMC? Previous research cites the difficulty that pilots have estimating weather conditions as a possible cause for their decision to continue into IMC (Goh & Wiegmann, 2001b; Wiegmann et al., 2002). The evidence from this experiment demonstrates that pilots do indeed have problems accurately estimating the weather conditions.

It was originally hypothesized that pilot's instrument ratings and additional training might aid in the earlier stages of Jensen's judgment model which compose sensitivity. However, this was not supported within the present study. The lack of substantial differences between instrument and non-instrument rated pilots can likely be attributed to the approximate equivalence of the two groups on cross-country hours. This suggests that the additional IFR training undertaken by instrument pilots does not aid in their judgment of weather.

Based upon other researchers categorizations (Wiggins & O'Hare, 2003) all of the pilots in the present study were novices (i.e., they had less than 1000 cross-country hours). The use of SDT with a wider range of cross-country hours may help to reveal potential differences attributed to experience. A clear relationship between pilot experience and sensitivity in weather judgment has not yet been established.

The ability to distinguish between conditions (i.e., pilot's sensitivity) is only a part of the judgment process (Jensen, 1995). Pilots also have motivational influences that take the form of response bias in SDT. The greater the difficulty in distinguishing between weather conditions, the stronger the potential for response criterion to influence pilots decisions. The GWIS provided additional information that had both conservative and liberal shifts on response criterion depending upon the consistency of information in the display.

With respect to their use of the GWIS, pilots revealed a potential to allow the display to influence their judgments. What is troubling is that pilots might use inaccurate METAR information that is better than the actual conditions to further their overestimation of weather conditions. When questioned after debriefing the majority of pilots realized that the conditions on the display and projected window did not always correspond. However, the majority of the pilots stated that when this happened they would trust their own judgment over that of the display. Despite this claim the separate ceiling and visibility SDT analyses suggests that the inaccurate displays did impact their decisions both by causing a shift in the response criterion and changing their sensitivity. Both conservative and liberal shifts were seen in the GWIS data, as well as both increases and decreases in sensitivity based upon the accuracy of GWIS information.

Although the present study points out possible problems with the use of GWIS METAR information it did not allow pilots to interact with the display. Under normal circumstances, the GWIS display would allow the pilot to access information on time of the report, more precise textual data from the report and also information from other local stations. However, the potential for pilots to use the GWIS to overestimate weather conditions suggests the need for training on the proper use of the display. Pilots should be taught to make use of the METAR information, but should also be trained on where to look for additional information when something does not agree.

A particularly alarming result was pilots' inability to accurately estimate ceilings. Although the data were generated from a simulation it revealed a potential estimation problem. Previous research has suggested pilots use a compensatory decision model in estimating weather conditions (Driskill et al., 1997; Hunter et al., 2003). However the current experiment found evidence of another potential problem. Data suggests that high VMC levels of visibility may have resulted in a conservative shift in pilot's response criterion in estimating ceiling. The same was true of the high visibility conditions causing a conservative shift in ceiling response criterion. Although this was not the compensatory decision model suggested by previous authors it may in turn be an equally serious problem. There may be a potential for both a criterion shift caused from one dimension of weather and then a compensatory judgment model after a biased estimation has been made.

Future research needs to further investigate the potential of the different weather dimensions to influence response bias in other dimensions. Ideally a larger number of ceilings and visibilities would be used to determine the extent of the biasing effects. It is not clear if one dimension has a stronger impact on the other dimension. That is, can ceiling influence visibility judgments more than visibility can influence ceiling judgments? The present study always asked pilots to estimate ceiling first. This may have caused pilots to consider ceiling before visibility thus establishing ceiling as an anchor for visibility. Although the data demonstrates criterion shifts in both dimensions the methodology may have contributed to a stronger biasing effect for ceiling because it was asked first. The requirement of pilots to estimate both ceiling and visibility may have also aided in the criterion shift. When a pilot estimates ceiling to be VMC it may serve as a VMC anchor for visibility. If pilots did not have to estimate both would there still be a criterion shift? Future research may also consider manipulating both dimensions but only asking pilots to estimate a single dimension.

Overall the experiment found evidence to suggest both situation assessment and motivation could contribute to a decision to continue into IMC. Pilot weather judgment is vulnerable to a number of different factors. These factors include poor sensitivity, a tendency to overestimate conditions, and inaccurate weather data. The current investigation also identified a new possible variable that influences pilot weather judgment, the interaction of the different elements of weather. Attributing a VFR into IMC accident to simply one of the factors would be a mistake. This experiment demonstrated the success of both crisp and fuzzy SDT in assessing the different components of weather judgment, and a need for future research to incorporate both. The current investigation raises some new concerns for research in pilot weather judgment. It is important for research to continue to examine how the two dimensions of weather interact with pilot judgment, and to determine the potential for this interaction to push pilots to continue a VFR flight into IMC.

REFERENCES

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*, 85-104.

Aircraft Owners and Pilots Association Air Safety Foundation. (1996). *Safety review: General aviation weather accidents*. Frederick, MD: Aircraft Owners and Pilots Association: Air Safety Foundation.

Aircraft Owners and Pilots Association Air Safety Foundation. (2001). *2001 Nall Report: General Aviation Accident Trends and Factors for 2000*. Frederick, MD: AOPA Air Safety Foundation.

Aircraft Owners and Pilots Association Air Safety Foundation. (2002). *2002 Nall Report: Accident Trends and Factors for 2001*. Frederick, MD: AOPA Air Safety Foundation.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*, 124-140.

Beringer, D. B., & Ball, J. D. (2003). Effects of NEXRAD graphical data resolution and direct weather viewing on pilot judgments of weather severity and willingness to continue flight. In *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 26-30). Santa Monica, CA: Human Factors and Ergonomics Society

Boatman, J. K. (2001). *Ounce of prevention, shades of gray, part 10 of 12, VFR-into-IMC's slippery slope*, from http://www.aopa.org/pilot/features/ounce0110.html.

Buck, R. N. (1998). *Weather Flying* (4 ed.). New York: McGraw-Hill.

Chamberlain, J. P., & Latorella, K. A. (2001). Convective weather detection be general aviation pilots with convectional and data-linked graphical weather information sources. *The 20th Digital Avionics Systems Conference*.

Dixon, M. W., & Proffitt, D. R. (2002). Overestimation of heights in virtual reality is influenced more by perceived distal size than by 2-D versus 3-D dimensionality of the display. *Perception, 31*, 103-112.

Driskill, W. E., Weismuller, J. J., Quebe, J., Hand, D. K., Dittmar, M. J., & Hunter, D. R. (1997). *The Use of Weather Information in Aeronautical Decision Making* (No. NTIS DOT/FAA/AM-97/3). Washington, D.C.: Federal Aviation Administration.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32-64.

Goh, J., & Wiegmann, D. A. (2001a). Visual flight rules (VFR) flight into instrument meteorological conditions (IMC): A review of the accident data. In *Proceedings of the 11th International Symposium on Aviation Psychology.* Columbus, OH: Ohio University.

Goh, J., & Wiegmann, D. A. (2001b). Visual flight rules flight into instrument meteorological conditions: An empirical investigation of the possible causes. *International Journal of Aviation Psychology, 11*, 359-379.

Goh, J., & Wiegmann, D. A. (2002). Human factors analysis of accidents involving visual flight rules flight into adverse weather. *Aviation Space & Environmental Medicine, 73*(8), 817-822.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Griffin, W. C., & Rockwell, T. H. (1987). A methodology for research on VFR flight into IMC. In *Proceedings of the 4th International Symposium on Aviation Psychology* (pp. 278-285). Columbus, OH: Ohio University.

Hancock, P. A., Masalonis, A. J., & Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomic Science, 1*, 207-230.

Hautus, M. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d'. *Behavior Research Methods. Instruments & Computers., 27*, 46-51.

Huberty, C. J., & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analyses. *Psychological Bulletin, 105*, 302-308.

Hunter, D. R., Martinussen, M., & Wiggins, M. (2003). Understanding how pilots make weather-related decisions. *International Journal of Aviation Psychology, 13*, 73-87.

Jensen, R. S. (1995). *Pilot Judgment and Crew Resource Management.* Burlington, VT: Ashgate.

Jensen, R. S., Guilkey, J. E., & Hunter, D. R. (1998). *An evaluation of pilot acceptance of the personal minimums training program for risk management* (No. DOT/FAA/AM-98/6). Washington, DC: Federal Aviation Administration.

Kaempf, G. L., & Klein, G. (1994). Aeronautical decision making: The next generation. In N. Johnston, N. McDonald & R. Fuller (Eds.), *Aviation Psychology in Practice* (pp. 223-254). Aldershot: Ashgate.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39*, 341-350.

Lankford, T. T. (2001). *Controlling Pilot Error: Weather*. New York: Mc-Graw Hill.

Latorella, K. A., & Chamberlain, J. (2002). Tactical vs. strategic behavior: General aviation piloting in convective weather scenarios. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 101-105). Santa Monica, CA: Human Factors and Ergonomics Society

Layton, C. F., & McCoy, E. (1989). General aviation pilot perceptions of deteriorating weather conditions. In *Proceedings of the 5th International Symposium on Aviation Psychology* (pp. 377-388). Columbus, OH: Ohio University.

Lind, A. T., Dershowitz, A., & Bussolari, S. R. (1994). *The influence of data link provided graphical weather on pilot decision making.* (No. DOT/FAA/RD-94/9). Washington, DC: Federal Aviation Administration.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of Naturalistic Decision Making. *Journal of Behavioral Decision Making, 14*, 331-352.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin, 98*, 185-199.

Masalonis, A. J., & Parasuraman, R. (2003). Fuzzy signal detection theory: Analysis of human and machine performance in air traffic control, and analytic considerations. *Ergonomics, 46*, 1045-1074.

Novacek, P. F., Burgess, M. A., Heck, M. L., & Stokes, A. F. (2001). *The effect of ownship information and NEXRAD resolution on pilot decision making in the use of a cockpit weather information display* (No. NASA/CR-2001-210845). Hampton, VA: NASA Langley Research Center.

O'Hare, D. (1990). Pilots' perception of risks and hazards in general aviation. *Aviation Space & Environmental Medicine, 61*(7), 599-603.

O'Hare, D. (2003). Aeronautical decision making: Metaphors, models and methods. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and Practice of Aviation Psychology* (pp. 201-237). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Hare, D., Owen, D., & Wiegmann, D. A. (2001). The 'where' and 'why' of cross-country VFR crashes: Database and simulation analysis. In *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 78-81) Santa Monica, CA: Human Factors and Ergonomics Society

O'Hare, D., & Smitheram, T. (1995). "Pressing on" into deteriorating conditions: An application of behavioral decision theory to pilot decision making. *International Journal of Aviation Psychology, 5*, 351-370.

Parasuraman, R., Masalonis, A. J., & Hancock, P. A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42*(4), 636-659.

Potter, S. S., Rockwell, T. H., & McCoy, C. E. (1989). General aviation pilot error in computer simulated adverse weather scenarios. In *Proceedings of the 5th International Symposium on Aviation Psychology* (pp. 570-575). Columbus, OH: Ohio University.

Smith, D. R. (2001). *Controlling Pilot Error: Controlled Flight into Terrain (CFIT/CFTT)* (Vol. 4). New York: McGraw-Hill.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods. Instruments & Computers., 31*, 137-149.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1-26.

Transportation Safety Board. (1990). *Report of a Safety Study on VFR Flight into Adverse Weather* (No. 90-SP002). Ottawa: Transportation Safety Board of Canada.

Vidulich, M. A. (2003). Mental workload and situation awareness: Essential concepts for aviation psychology practice. In M. A. Vidulich (Ed.), *Principles and Practices of Aviation Psychology* (pp. 115-146). Mahwah, New Jersey: Lawrence Earlbaum Associates.

Wichman, H., & Ball, J. (1983). Locus of control, self-serving biases, and attitudes towards safety in general aviaition. *Aviation Space & Environmental Medicine, 54*, 507-510.

Wickens, C. D. (2001). Workload and situation awareness. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, Workload, and Fatigue*. Mawah, NJ: Lawrence Earlbaum Associates.

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human Factors, 44*(2), 189-197.

Wiggins, M. W., & O'Hare, D. (1995). Expertise in aeronautical weather-related decision making: A cross-sectional analysis of general aviation pilots. *Journal of Experimental Psychology: Applied, 1*(4), 305-320.

Wiggins, M. W., & O'Hare, D. (2003). Expert and novice pilot perceptions of static in-flight images of weather. *The International Journal of Aviation Psychology, 13*, 173-187.

Wilson, D. R., & Fallshore, M. (2001). Optimistic and ability biases in pilots' decisions and perceptions of risk regarding VFR flight into IMC. In *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: Ohio University.

Yang, T. L., Dixon, M. W., & Proffitt, D. R. (1999). Seeing big things: Overestimation of heights is greater for real objects than for objects in pictures. *Perception, 28*, 445-467.

Yuchnovicz, D. E., Novacek, P. F., Burgess, M. A., Heck, M. L., & Stokes, A. F. (2001). *Use of a Data-Linked Weather Information Display and Effects on Pilot Navigation Decision Making in a Piloted Simulation Study* (No. NASA/CR-2001-211047). Hampton, VA: NASA Langley Research Center.

APPENDIX A:

WEATHER CLASSIFICATION BASED UPON FARS

TABLE A1: Ceiling and Visibility Categories

Category	Ceiling	Visibility
VFR	> 3000	> 5 sm
MVFR	1000-3000	3 to 5 sm
IFR	500 to 1000	1 to 3 sm
LIFR	< 500 ft	Less than 1 sm

TABLE A2: Basic VFR Weather Minimums

Airspace	Visibility	Clearance
Class A	N/A	N/A
Class B	3 SM	Clear of Clouds
Class C & D	3 SM	500 feet below 1000 feet above 2000 feet horizontal
Class E less than 10,000 MSL	3 SM	500 feet below 1000 feet above 2000 feet horizontal
Class E at or above 10,000 feet MSL	5 SM	1000 feet below 1000 feet above 1 SM horizontal
Class G Day 1,200 feet AGL or less then 10,000 MSL	1 SM	500 feet below 1000 feet above 2000 feet horizontal
Class G Night below 10,000 feet MSL	3 SM	500 feet below 1000 feet above 2000 feet horizontal
Class G at or above 10,000 feet MSL and above 1,200 AGL	5 SM	1000 feet below 1000 feet above 1 SM horizontal

APPENDIX B:

GLOSSARY

AGL	Above Ground Level
AIRMET	Airman's Meteorological Information
ANOVA	Analysis of Variance
AOPA	Aircraft Owners and Pilot Association
ARTCC	Air Route Traffic Control Center
ASOS	Automated Surface Observation System
ATC	Air Traffic Control
ATIS	Automatic Terminal Information Service
AWOS	Automated Weather Observing System
CAMI	Civil Aerospace Medical Institute
CFIT	Controlled Flight into Terrain
DUATS	Direct User Access Terminals
EFAS	Enroute Flight Advisory System
FAA	Federal Aviation Administration
FAR	Federal Aviation Regulations
FISDL	Flight Information Services Data-Link
FSS	Flight Service Station
FW	Flightwatch
GA	General Aviation
GPS	Global Positioning System
GWIS	Graphical Weather Information System
HIWAS	Hazardous In-flight Weather Advisory Service
ICAO	International Civil Aviation Organization
IFR	Instrument Flight Rules
IMC	Instrument Meteorological Conditions
LIFR	Low Instrument Flight Rules
MANOVA	Multivariate Analysis of Variance
METAR	Meteorological Report
MSL	Mean Sea Level
MVFR	Marginal Visual Flight Rules
NAS	National Airspace System
NEXRAD	Next Generation Radar
NOAA	National Oceanic and Atmospheric Administration
NTSB	National Transportation Safety Board
NWS	National Weather Service
PIREP	Pilot Report
SDT	Signal Detection Theory
SIGMET	Significant Meteorological Information
sm	Statue Miles
TAF	Terminal Aerodrome Forecast
VFR	Visual Flight Rules

APPENDIX B: (continued)

VMC	Visual Meteorological Conditions
VOR	Very high frequency Omnidirectional Range

APPENDIX C:

SIGNAL DETECTION FORMULAS

Sensitivity

The most commonly used metric for sensitivity is *d*'. It is the standardized distance between the normal curve that approximates the signal distribution and the noise distribution. It is the distance between the curves in units of standard deviations.

$D' = Z(HR) - Z(FAR)$

Where HR and FAR are equal to the hit rates and false alarm rates respectively. Z(HR) is equal to the position of the observed hit rate along the normal curve. Z(FAR) is equal to the position of the observed false alarm rate along the normal curve. Use of *d*' as a metric of sensitivity requires two assumptions be fulfilled. First the data must be normal. Second the standard deviations of the signal and noise distributions must be equivalent.

When assumptions for *d*' are violated researchers turn to nonparametric measures of sensitivity. The most popular nonparametric measure of sensitivity is A'(Stanislaw & Todorov, 1999). A' ranges from .5 when a signal cannot be differentiated from noise to 1 when there is perfect performance.

$A' = 1-(0.25*[(FAR/HR) + ((1-HR)/(1-FAR))])$

However, problems with both metrics arise when there is extreme performance. If HR = 1 or FAR = 0 both metrics cannot be computed. Two approaches have been advocated for dealing with extreme performance. One is to add 0.5 to the number of hits and false alarms and to add 1 to both the number of signals and the number of noise trials. This method would be applied regardless of extreme scores. The second approach is to

APPENDIX C (continued)

replace rates of 0 with (0.5/*n*) and rates of 1 with (*n*-0.5), where *n* is the number of signal or noise trials. Both corrections can be applied to either *d'* or A'.

Response Bias

Similar to sensitivity there are multiple metrics for response bias the two most popular are β and c. The parametric version β assumes responses are based upon a likelihood ratio. Subjects who favor neither a yes nor no response have β = 1. A β greater than one corresponds to a participant who is more willing to say no and a β less than one corresponds to a participant more willing to say yes.

$$\beta = Y(HR)/Y(FAR)$$

Where Y(HR) represents the ordinate of the normal curve for the hit rate and Y(FAR) represents the ordinate of the normal distribution for the false alarm rate.

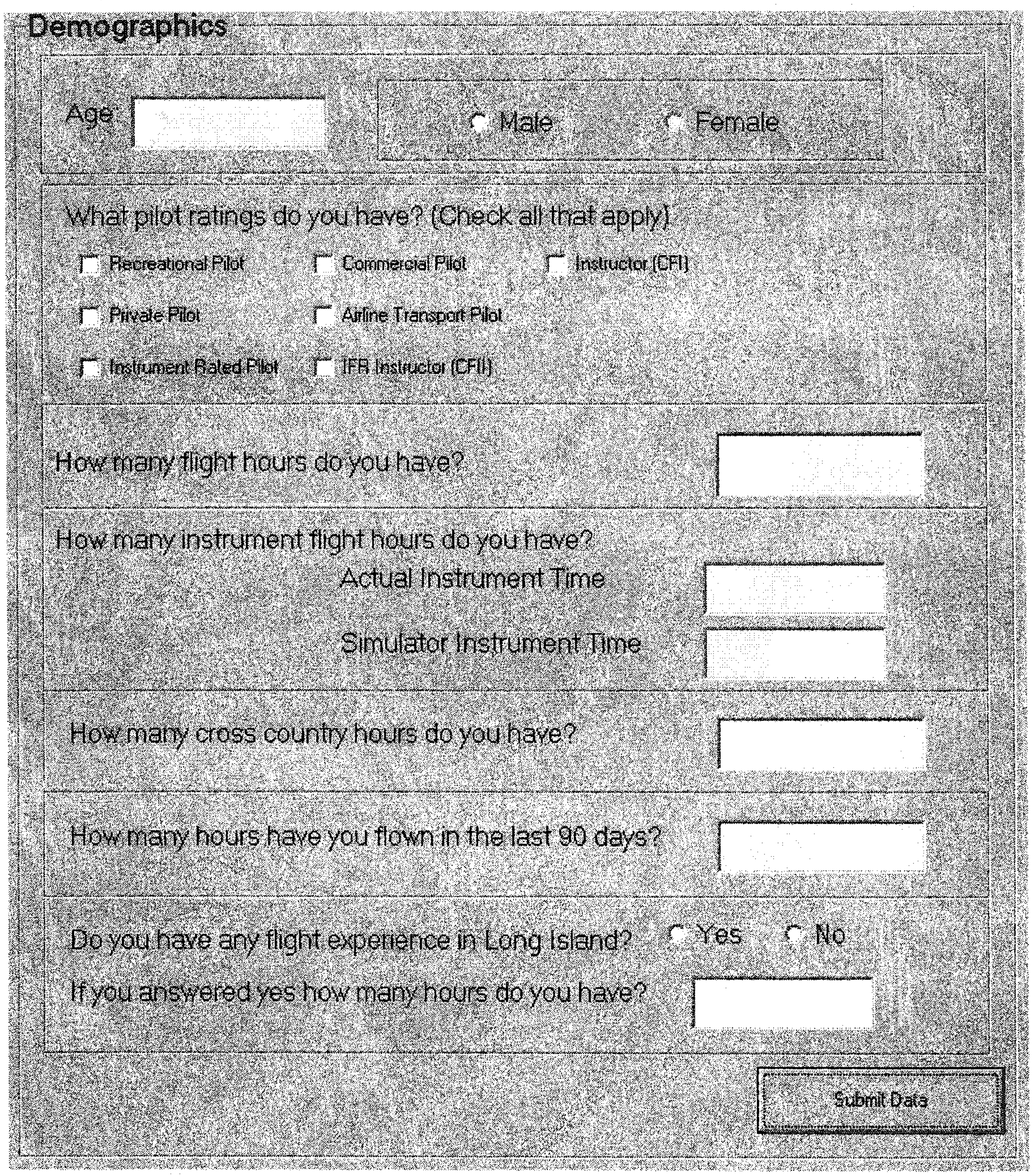
The most commonly used non-parametric version of response criterion is c. c is defined as the distance between the criterion and the neutral point. The neutral point is where β = 1. At this point c has a value of 0. Negative values of c represent a bias towards responding yes, where as positive values of c are representative of a bias to say no.

$$c = -0.5(Z(HR) + Z(FAR))$$

As in the formula for *d*', Z(HR) is equal to the position of the observed hit rate along the normal curve. Z(FAR) is equal to the position of the observed false alarm rate along the normal curve.

APPENDIX D:

QUESTIONNAIRE



Demographics

Age

Male Female

What pilot ratings do you have? (Check all that apply)

- Recreational Pilot
- Commercial Pilot
- Instructor (CFI)
- Private Pilot
- Airline Transport Pilot
- Instrument Rated Pilot
- IFR Instructor (CFII)

How many flight hours do you have?

How many instrument flight hours do you have?

Actual Instrument Time

Simulator Instrument Time

How many cross country hours do you have?

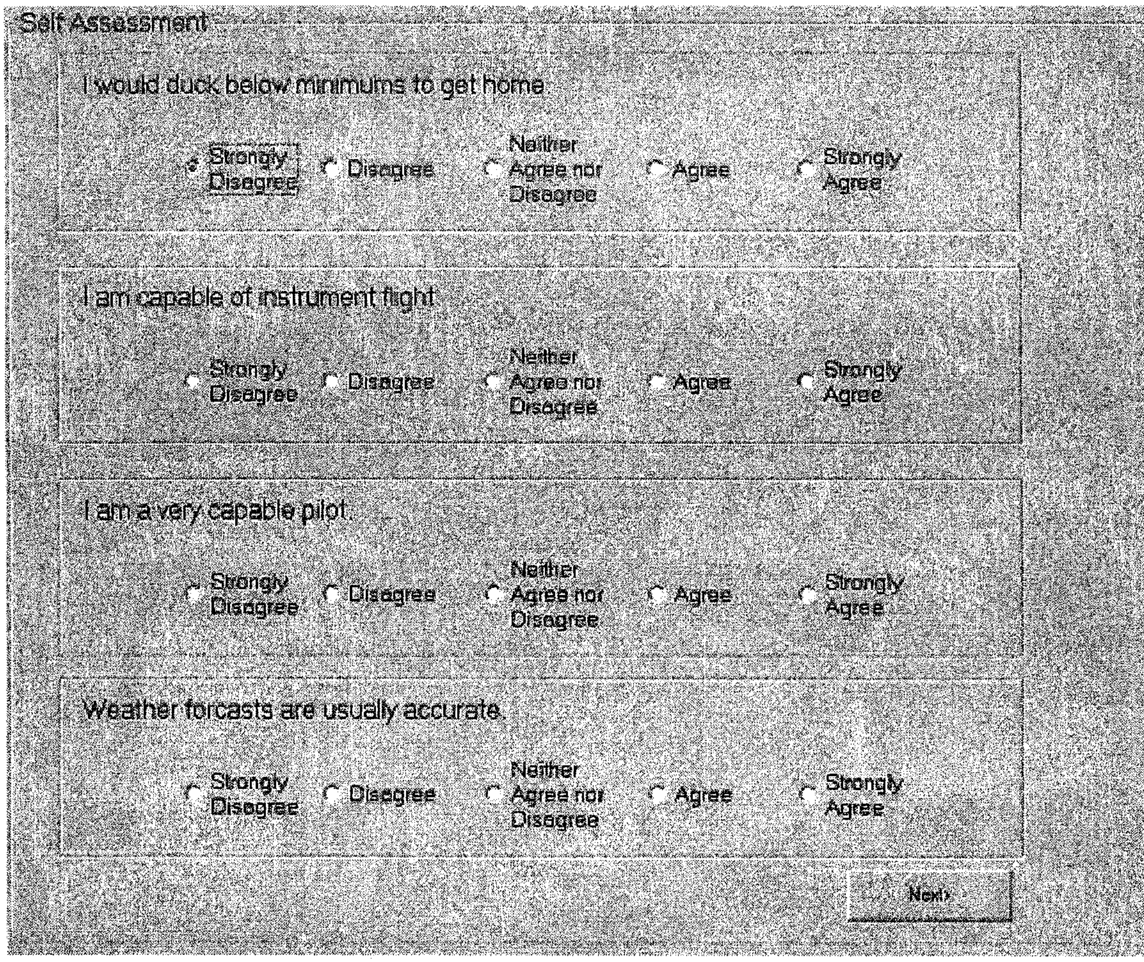
How many hours have you flown in the last 90 days?

Do you have any flight experience in Long Island? Yes No

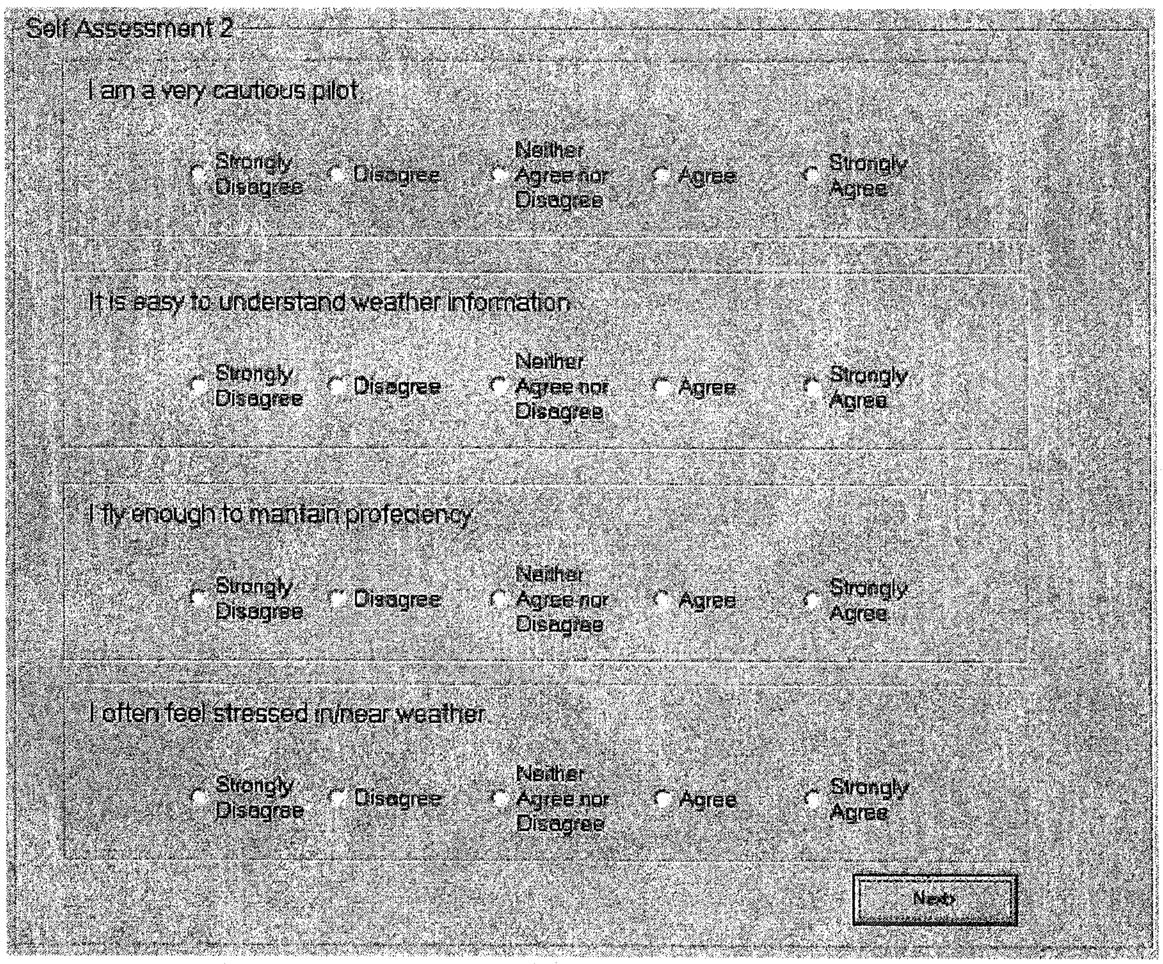
If you answered yes how many hours do you have?

Submit Data

APPENDIX D (continued)



APPENDIX D (continued)



Self Assessment 2

I am a very cautious pilot.

Strongly Disagree | Disagree | Neither Agree nor Disagree | Agree | Strongly Agree

It is easy to understand weather information.

Strongly Disagree | Disagree | Neither Agree nor Disagree | Agree | Strongly Agree

I fly enough to mantain profeciency.

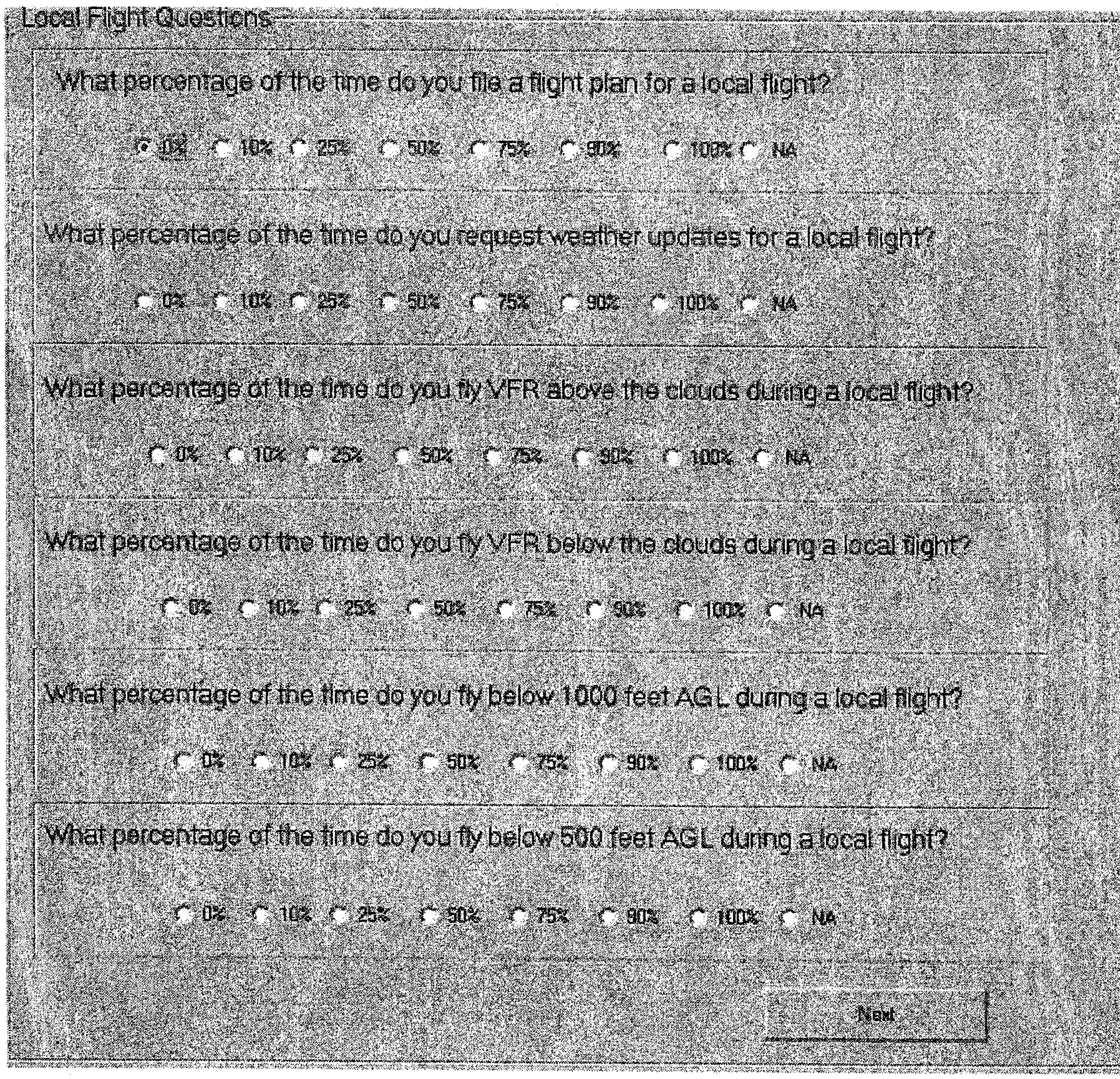
Strongly Disagree | Disagree | Neither Agree nor Disagree | Agree | Strongly Agree

I often feel stressed in/near weather.

Strongly Disagree | Disagree | Neither Agree nor Disagree | Agree | Strongly Agree

Next

APPENDIX D (continued)



Local Flight Questions

What percentage of the time do you file a flight plan for a local flight?

0% 10% 25% 50% 75% 90% 100% NA

What percentage of the time do you request weather updates for a local flight?

0% 10% 25% 50% 75% 90% 100% NA

What percentage of the time do you fly VFR above the clouds during a local flight?

0% 10% 25% 50% 75% 90% 100% NA

What percentage of the time do you fly VFR below the clouds during a local flight?

0% 10% 25% 50% 75% 90% 100% NA

What percentage of the time do you fly below 1000 feet AGL during a local flight?

0% 10% 25% 50% 75% 90% 100% NA

What percentage of the time do you fly below 500 feet AGL during a local flight?

0% 10% 25% 50% 75% 90% 100% NA

Next

APPENDIX D (continued)

Cross Country Flight Questions

What percentage of the time do you file a flight plan for a cross country flight?

0% 10% 25% 50% 75% 90% 100% NA

What percentage of the time do you request weather updates for a cross country flight?

0% 10% 25% 50% 75% 90% 100% NA

What percentage of the time do you fly VFR above the clouds during a cross country flight?

0% 10% 25% 50% 75% 90% 100% NA

What percentage of the time do you fly VFR below the clouds during a cross country flight?

0% 10% 25% 50% 75% 90% 100% NA

What percentage of the time do you fly below 1000 feet AGL during a cross country flight?

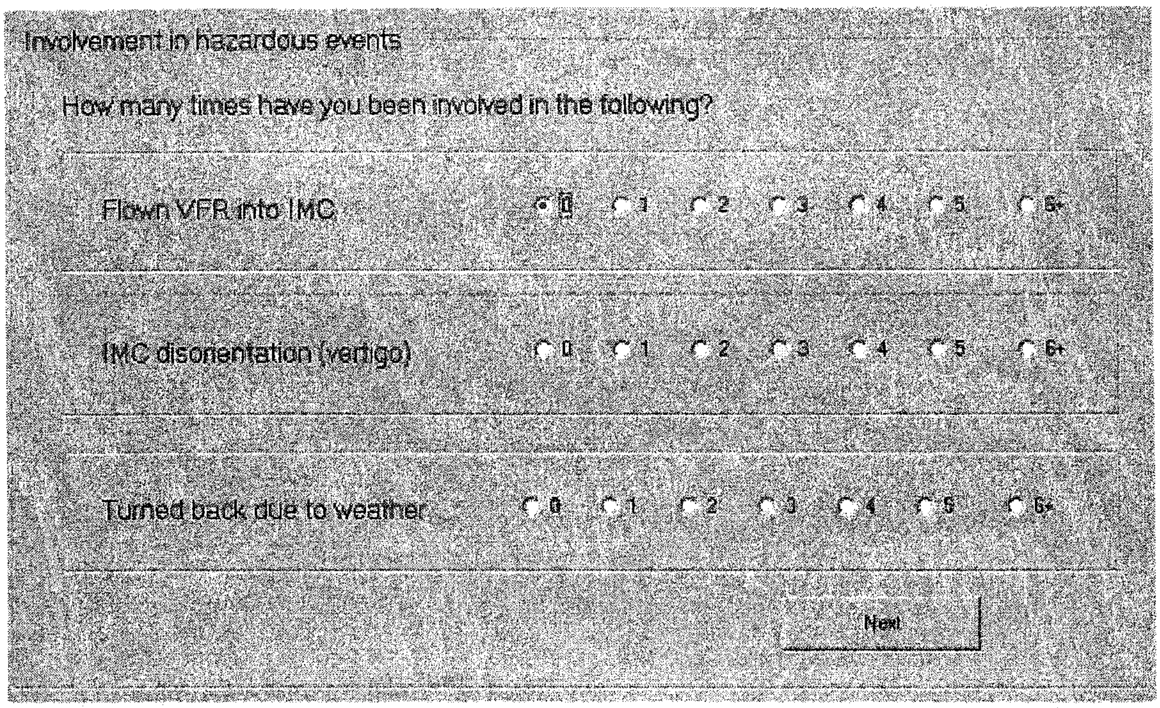
0% 10% 25% 50% 75% 90% 100% NA

What percentage of the time do you fly below 500 feet AGL during a cross country flight?

0% 10% 25% 50% 75% 90% 100% NA

Next

APPENDIX D (continued)


Involvement in hazardous events

How many times have you been involved in the following?

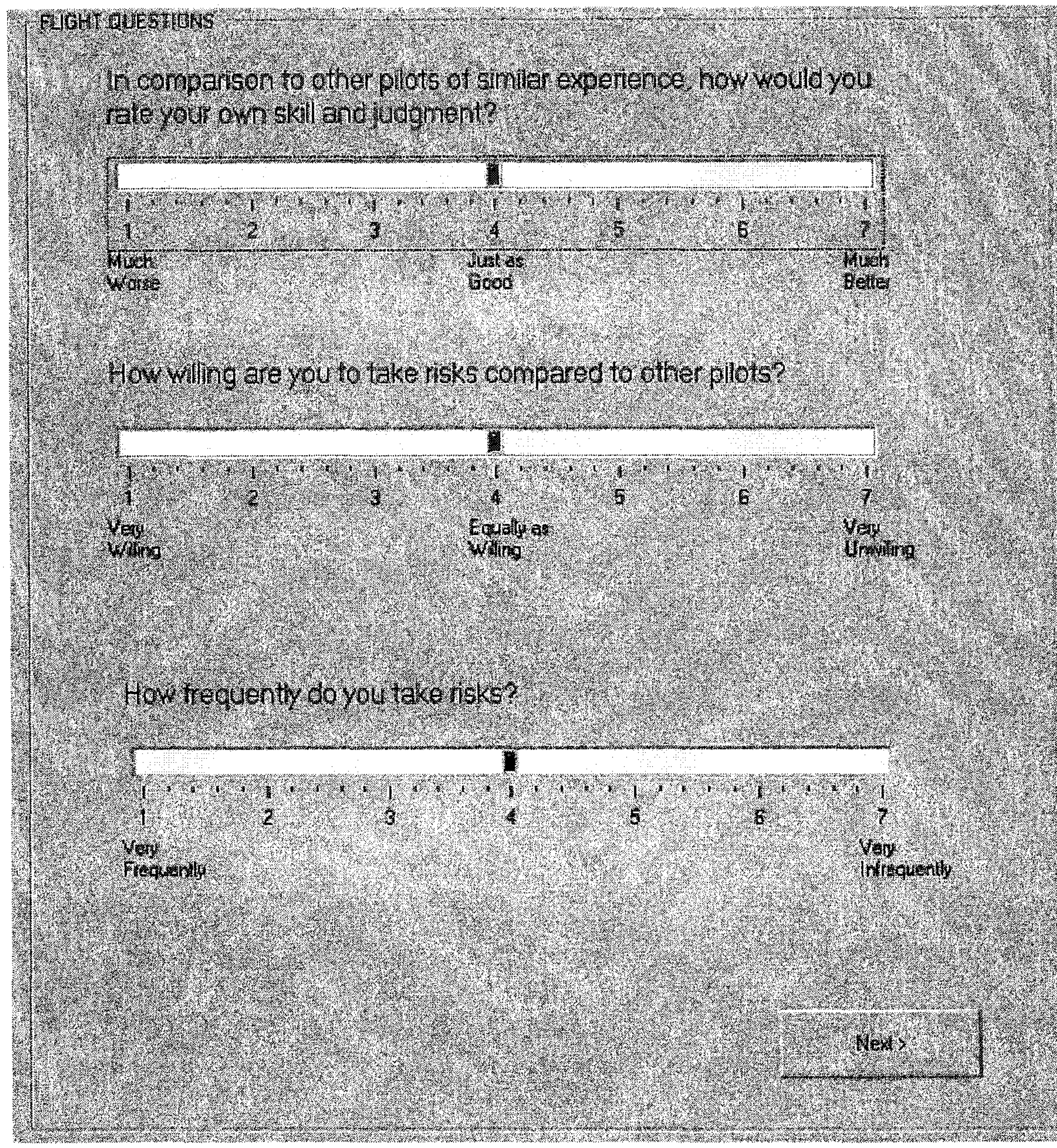
Flown VFR into IMC 0 1 2 3 4 5 6+

IMC disorientation (vertigo) 0 1 2 3 4 5 6+

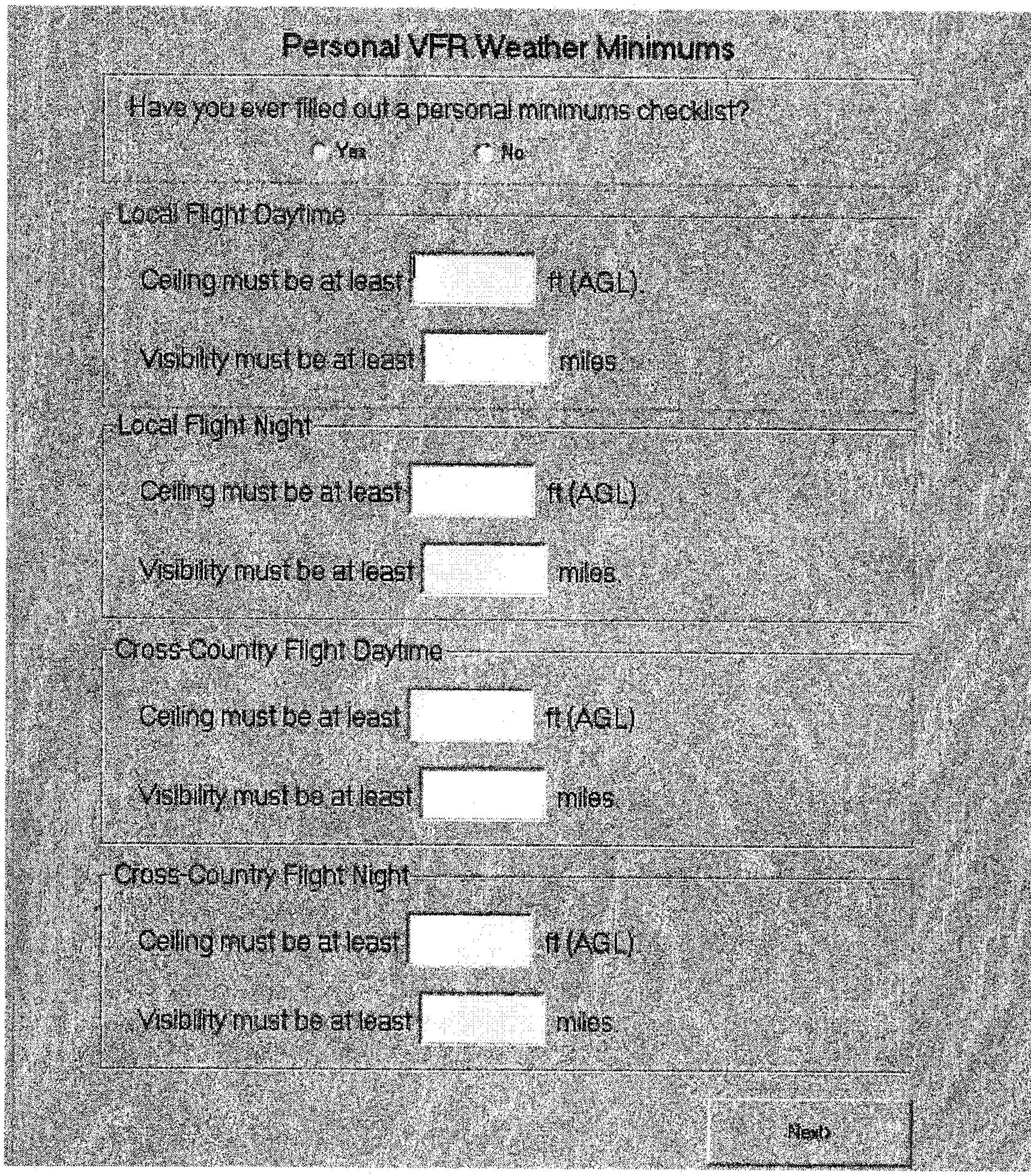
Turned back due to weather 0 1 2 3 4 5 6+

Next

APPENDIX D (continued)



APPENDIX D (continued)



Personal VFR Weather Minimums

Have you ever filled out a personal minimums checklist?

Yes No

Local Flight Daytime

Ceiling must be at least ft (AGL).

Visibility must be at least miles.

Local Flight Night

Ceiling must be at least ft (AGL).

Visibility must be at least miles.

Cross-Country Flight Daytime

Ceiling must be at least ft (AGL).

Visibility must be at least miles.

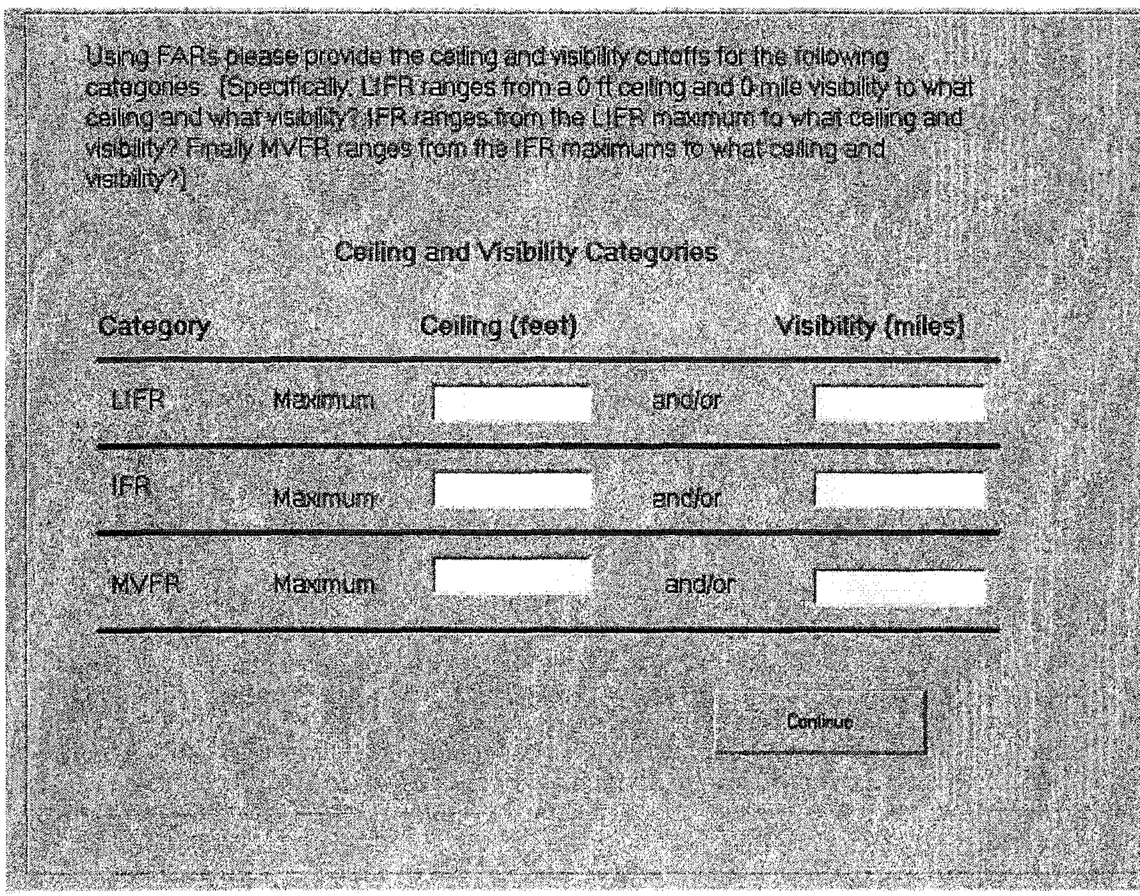
Cross-Country Flight Night

Ceiling must be at least ft (AGL).

Visibility must be at least miles.

Next

APPENDIX D (continued)



Using FARs please provide the ceiling and visibility cutoffs for the following categories. (Specifically, LIFR ranges from a 0 ft ceiling and 0 mile visibility to what ceiling and what visibility? IFR ranges from the LIFR maximum to what ceiling and visibility? Finally MVFR ranges from the IFR maximums to what ceiling and visibility?)

Ceiling and Visibility Categories

Category		Ceiling (feet)		Visibility (miles)
LIFR	Maximum		and/or	
IFR	Maximum		and/or	
MVFR	Maximum		and/or	

Continue

APPENDIX E:

ANOVA AND MEANS TABLES

Baseline Accuracy and Means Tables

TABLE E1: Baseline Accuracy ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
Between Subjects							
Rating (R)	1	.0326	22	.153	.212	.649	.010
Within Subjects							
Ceiling (C)	3	.296	66	.136	2.181	.099	.090
Visibility (V)	3	1.171	66	.200	5.848	.001	.210
R x C	3	.0569	66	.136	.418	.740	.019
R x V	3	.0152	66	.200	.076	.973	.003
C x V	9	1.946	198	.117	16.647	.000	.431
R x C x V	9	.0210	198	.117	.179	.996	.008

TABLE E2: Accuracy Means Table for Rating

Rating	Mean	N	Std Dev
Instrument	0.839	384	0.368
Non-instrument	0.852	384	0.356

TABLE E3: Accuracy Means Table for Visibility

Visibility	Mean	N	Std Dev
2	0.880	192	0.326
3	0.839	192	0.369
5	0.922	192	0.269
10	0.740	192	0.440

TABLE E4: Accuracy Means Table for Ceiling

Ceiling	Mean	N	Std Dev
400	0.870	192	0.337
900	0.786	192	0.411
2900	0.865	192	0.343
4500	0.859	192	0.349

APPENDIX E: (continued)

TABLE E5: Accuracy Means (and Standard Deviations) for Ceiling at Each Visibility

	Ceiling			
Visibility	400	900	2900	4500
2	.958 (.202)	.979 (.144)	.771 (.425)	.813 (.394)
3	1.00 (0.00)	.958 (.202)	.771 (.425)	.625 (.489)
5	.896 (3.09)	.813 (.394)	.979 (.144)	1.00 (0.00)
10	.625 (.489)	.396 (.494)	.938 (.245)	1.00 (0.00)

Baseline Comfort ANOVA Tables and Means

TABLE E6: Baseline Comfort ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.173	22	4708.193	.000	.995	.000
			Within Subjects				
Ceiling (C)	3	84105.44	66	824.74	101.977	.000	.823
Visibility (V)	3	96889.17	66	582.726	166.269	.000	.883
R x C	3	409.83	66	824.74	.497	.686	.022
R x V	3	721.970	66	582.726	1.239	.303	.053
C x V	9	2549.733	198	300.140	8.495	.000	.279
R x C x V	9	337.104	198	300.140	1.123	.348	.049

TABLE E7: Comfort Means Table for Rating

Rating	Mean	N	Std Dev
Instrument	43.625	384	34.268
Non-instrument	43.655	384	34.136

TABLE E8: Comfort Means Table for Visibility

Visibility	Mean	N	Std Dev
2	17.451	192	20.164
3	34.898	192	28.713
5	52.691	192	32.364
10	69.519	192	29.852

APPENDIX E: (continued)

TABLE E9: Comfort Means Table for Ceiling

Ceiling	Mean	N	Std Dev
400	23.312	192	27.905
900	27.876	192	25.900
2900	62.448	192	29.751
4500	60.923	192	32.183

TABLE E10: Comfort Means (and Standard Deviations) for Ceiling at Each Visibility

	Ceiling			
Visibility	400	900	2900	4500
2	5.518 (11.627)	10.535 (13.598)	30.037 (23.004)	23.715 (20.170)
3	13.025 (17.302)	16.849 (17.770)	57.154 (23.683)	52.564 (22.990)
5	25.240 (24.549)	33.281 (23.067)	74.939 (20.023)	77.302 (2.033)
10	49.467 (31.519)	50.837 (26.340)	87.660 (13.999)	90.111 (14.216)

Baseline Visibility RMSE ANOVA Tables and Means

TABLE E11: Baseline Visibility RMSE ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	13.814	22	5.780	2.390	.136	.098
			Within Subjects				
Ceiling (C)	3	20.647	66	3.233	6.387	.001	.225
Visibility (V)	3	121.581	66	5.222	23.283	.000	.514
R x C	3	8.828	66	3.233	2.731	.051	.110
R x V	3	31.963	66	5.222	6.121	.001	.216
C x V	9	38.960	198	1.577	24.706	.000	.529
R x C x V	9	1.037	198	1.577	0.658	.747	.029

TABLE E12: Visibility RMSE Means Table for Rating

Rating	Mean	N	Std Dev
Instrument	1.430	384	1.526
Non-instrument	1.698	384	1.873

APPENDIX E: (continued)

TABLE E13: Visibility RMSE Means Table for Visibility

Visibility	Mean	N	Std Dev
2	0.651	192	0.861
3	1.328	192	1.064
5	1.719	192	1.308
10	2.557	192	2.506

TABLE E14: Visibility RMSE Means Table for Ceiling

Ceiling	Mean	N	Std Dev
400	1.932	192	1.936
900	1.740	192	2.017
2900	1.354	192	1.373
4500	1.229	192	1.322

TABLE E15: Visibility RMSE Means (and Standard Deviations) for Ceiling at Each Visibility

	Ceiling							
Visibility	400		900		2900		4500	
2	0.833	(0.663)	0.313	(0.589)	0.708	(0.824)	0.750	(1.176)
3	1.000	(0.825)	1.083	(0.794)	1.750	(1.246)	1.479	(1.167)
5	2.021	(1.407)	1.521	(1.148)	1.625	(1.315)	1.708	(1.336)
10	3.875	(2.481)	4.042	(2.501)	1.333	(1.742)	0.979	(1.407)

Baseline Ceiling RMSE ANOVA Tables and Means

TABLE E16: Baseline Ceiling RMSE ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	5192607	22	4643364	1.118	.302	.048
			Within Subjects				
Ceiling (C)	3	62798380	66	3305274	18.99	.000	.463
Visibility (V)	3	3785047	66	1203028	3.14	.031	.125
R x C	3	1144118	66	3305274	.346	.792	.015
R x V	3	731861	66	1203028	.608	.612	.027
C x V	9	7177263	198	1130735	6.34	.000	.224
R x C x V	9	1494251	198	1130735	1.321	.228	.057

APPENDIX E: (continued)

TABLE E17: Ceiling RMSE Means Table for Rating

Rating	Mean	N	Std Dev
Instrument	1113.7	384	1515.7
Non-instrument	949.2	384	998.6

TABLE E18: Ceiling RMSE Means Table for Visibility

Visibility	Mean	N	Std Dev
2	1201.3	192	1677.7
3	892.7	192	1103.5
5	943.2	192	966.6
10	1088.5	192	1267.7

TABLE E19: Ceiling RMSE Means Table for Ceiling

Ceiling	Mean	N	Std Dev
400	778.6	192	984.7
900	647.1	192	819.6
2900	817.7	192	930.8
4500	1882.3	192	1771.7

TABLE E20: Ceiling RMSE Means (and Standard Deviations) for Ceiling at Each Visibility

	Ceiling							
Visibility	400		900		2900		4500	
2	383.3	(414.3)	740.6	(1195.3)	981.3	(991.4)	2700.0	(2368.8)
3	606.3	(426.0)	356.3	(535.1)	816.7	(1052.7)	1791.7	(1471.0)
5	843.8	(619.2)	604.2	(528.7)	760.4	(871.0)	1564.6	(1349.0)
10	1281.3	(1658.8)	887.5	(760.9)	712.5	(793.2)	1472.9	(1482.1)

APPENDIX E: (continued)

Baseline SDT ANOVA Tables

TABLE E21: Baseline Crisp and Fuzzy ANOVA Table

DV	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
Crisp A'	1	.00003	22	.0017	.017	.897	.001
Crisp C	1	.07692	22	.169	.454	.508	.020
Fuzzy A'	1	.00000	22	.00122	.000	.988	.000
Fuzzy C	1	.568	22	.199	2.858	.105	.115
Crisp Ceiling A'	1	.0001	22	.0218	.045	.834	.002
Crisp Ceiling c	1	.0355	22	.215	.165	.688	.007
Fuzzy Ceiling A'	1	.0009	22	.0036	.247	.624	.011
Fuzzy Ceiling c	1	.106	22	.0965	1.103	.305	.048
Crisp Visibility A'	1	.0002	22	.0050	.044	.835	.002
Crisp Visibility c	1	1.540	22	.254	6.054	.022	.216
Fuzzy Visibility A'	1	.0296	22	.0165	1.799	.193	.076
Fuzzy Visibility c	1	.0908	22	.0160	5.673	.026	.205

APPENDIX E: (continued)

GWIS Accuracy ANOVA Tables and Means

TABLE E22: GWIS Accuracy ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	1.021	22	.244	4.184	.053	.160
			Within Subjects				
Ceiling (C)	1	2.370	22	.333	7.122	.014	.245
Visibility (V)	2	.146	44	.359	.406	.669	.018
GWIS (G)	5	.249	110	.0854	2.911	.017	.117
R x C	1	.113	22	.333	.341	.565	.015
R x V	2	.0486	44	.359	.135	.874	.006
R x G	5	.0444	110	.0854	.520	.760	.023
C x V	2	4.267	44	.298	15.507	.000	.413
C x G	5	.491	110	.0982	5.001	.000	.185
V x G	10	.230	220	.0917	2.506	.007	.102
R x C x G	5	.0537	110	.0982	.547	.740	.024
R x C x V	2	.238	44	.298	.799	.456	.035
R x V x G	10	.04931	220	.0917	.538	.862	.024
C x V x G	10	.138	220	.0833	1.653	.093	.070
R x C x V x G	10	.0600	220	0833	.720	.706	.032

TABLE E23: Accuracy Means Table for GWIS

GWIS	Mean	N	Std Dev
None	0.878	288	0.327
Accurate	0.906	288	0.292
Ceiling Better	0.885	288	0.319
Ceiling Worse	0.837	288	0.370
Visibility Better	0.872	288	0.335
Visibility Worse	0.830	288	0.376

APPENDIX E: (continued)

TABLE E24: Accuracy Means (and Standard Deviations) for GWIS at Each Ceiling

	Ceiling			
GWIS	400		900	
None	0.917	(0.277)	0.840	(0.368)
Accurate	0.931	(0.255)	0.882	(0.324)
Ceiling Better	0.854	(0.354)	0.917	(0.277)
Ceiling Worse	0.896	(0.307)	0.778	(0.417)
Visibility Better	0.910	(0.288)	0.833	(0.374)

TABLE E25: Accuracy Means (and Standard Deviations) for GWIS at Each Visibility

	Visibility					
GWIS	2		3		5	
None	0.875	(0.332)	0.865	(0.344)	0.896	(0.307)
Accurate	0.885	(0.320)	0.938	(0.243)	0.896	(0.307)
Ceiling Better	0.927	(0.261)	0.906	(0.293)	0.823	(0.384)
Ceiling Worse	0.885	(0.320)	0.833	(0.375)	0.792	(0.408)
Visibility Better	0.833	(0.375)	0.896	(0.307)	0.885	(0.320)

APPENDIX E: (continued)

GWIS Comfort ANOVA Tables and Means

TABLE E26: GWIS Comfort ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.821	22	15157.10	.000	.994	.000
			Within Subjects				
Ceiling (C)	1	449929.29	22	4587.57	98.706	.000	.817
Visibility (V)	2	149335.27	44	1063.59	140.40	.000	.865
GWIS (G)	5	1055.84	110	227.21	4.647	.001	.175
R x C	1	1116.71	22	4587.57	.243	.623	.011
R x V	2	329.98	44	1063.59	.310	.735	.014
R x G	5	70.42	110	227.21	.310	.906	.014
C x V	2	23737.20	44	966.14	24.569	.000	.528
C x G	5	133.89	110	175.43	.763	.578	.034
V x G	10	84.96	220	154.49	.550	.853	.024
R x C x G	5	262.46	110	175.43	1.496	.197	.064
R x C x V	2	731.15	44	966.14	.757	.475	.033
R x V x G	10	46.48	220	154.49	.301	.980	.013
C x V x G	10	166.85	220	132.53	1.259	.255	.054
R x C x V x G	10	247.08	220	132.53	1.864	.051	.078

TABLE E27: Comfort Means Table for GWIS

GWIS	Mean	N	Std Dev
None	37.133	288	30.289
Accurate	37.096	288	30.819
Ceiling Better	40.086	288	30.412
Ceiling Worse	35.916	288	31.656
Visibility Better	38.186	288	29.996
Visibility Worse	34.483	288	30.406

APPENDIX E: (continued)

GWIS Ceiling RMSE ANOVA Tables and Means

TABLE E28: GWIS Ceiling RMSE ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	8566892	22	7375333	1.162	.293	.050
			Within Subjects				
Ceiling (C)	1	19564917	22	5994796	3.264	.085	.129
Visibility (V)	2	6828903	44	1435893	4.756	.013	.178
GWIS (G)	5	2369227	110	619141	3.82	.003	.148
R x C	1	5470876	22	5994796	.913	.350	.040
R x V	2	92530	44	1435893	.064	.938	.003
R x G	5	85656	110	619141	.138	.983	.006
C x V	2	3083155	44	1628849	1.893	.163	.079
C x G	5	328741	110	731498	.449	.813	.020
V x G	10	533695	220	554497	.962	.477	.042
R x C x G	5	363704	110	731498	.497	.778	.022
R x C x V	2	1179819	44	1628849	.724	.490	.032
R x V x G	10	1551724	220	554497	2.798	.003	.113
C x V x G	10	1128337	220	1128337	1.452	.159	.062
R x C x V x G	10	417433	220	776827	.537	.863	.024

TABLE E29: Ceiling RMSE Means Table for GWIS

GWIS	Mean	N	Std Dev
None	709.896	288	911.549
Accurate	590.625	288	765.273
Ceiling Better	698.872	288	819.695
Ceiling Worse	839.080	288	1336.731
Visibility Better	604.688	288	804.032
Visibility Worse	649.306	288	966.296

APPENDIX E: (continued)

TABLE E30: Ceiling RMSE Means Table for GWIS by Visibility by Rating

Visibility	Rating	Ceiling 2		3		5	
None	Instrument	961	(1345)	706	(1043)	704	(757)
	Non-instrument	760	(783)	467	(622)	660	(691)
Accurate	Instrument	593	(669)	579	(877)	833	(1142)
	Non-instrument	581	(618)	452	(578)	505	(502)
Ceiling Better	Instrument	1258	(1243)	492	(579)	579	(554)
	Non-instrument	593	(662)	606	(656)	665	(805)
Ceiling Worse	Instrument	1000	(1019)	869	(1255)	756	(1508)
	Non-instrument	895	(1114)	615	(595)	900	(2082)
Visibility Better	Instrument	693	(897)	498	(602)	863	(1276)
	Non-instrument	621	(600)	438	(540)	517	(614)
Visibility Worse	Instrument	719	(1391)	633	(923)	808	(1262)
	Non-instrument	756	(753)	471	(496)	508	(644)

APPENDIX E: (continued)

GWIS Visibility RMSE ANOVA Tables and Means

TABLE E31: GWIS Ceiling RMSE ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	22.891	22	11.247	2.035	.168	.085
			Within Subjects				
Ceiling (C)	1	91.103	22	8.257	11.033	.003	.334
Visibility (V)	2	150.175	44	2.706	55.495	.000	.716
GWIS (G)	5	1.146	110	0.811	1.414	.225	.060
R x C	1	33.018	22	8.257	3.999	.058	.154
R x V	2	1.225	44	2.706	.453	.639	.020
R x G	5	1.456	110	.811	1.796	.120	.075
C x V	2	11.389	44	2.701	4.217	.021	.161
C x G	5	1.210	110	.662	1.827	.113	.077
V x G	10	.722	220	.686	1.052	.401	.046
R x C x G	5	.879	110	.662	1.328	.258	.057
R x C x V	2	14.121	44	2.701	5.228	.009	.192
R x V x G	10	.385	220	.686	.561	.844	.025
C x V x G	10	.319	220	.609	.524	.872	.023
R x C x V x G	10	.706	220	.609	1.159	.320	.050

TABLE E32: Visibility RMSE Means Table for GWIS

GWIS	Mean	N	Std Dev
None	1.167	288	1.138
Accurate	1.160	288	1.192
Ceiling Better	1.104	288	1.109
Ceiling Worse	1.142	288	1.254
Visibility Better	1.281	288	1.166
Visibility Worse	1.118	288	1.056

APPENDIX E: (continued)

GWIS SDT ANOVA Tables

TABLE E33: Crisp A' ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.022	22	.012	1.832	.190	.077
			Within Subjects				
GWIS (G)	5	.019	110	.004	4.364	.001	.166
R x G	5	.002	110	.004	.510	.768	.023

TABLE E34: Fuzzy A' ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.003	22	.007	.375	.547	.017
			Within Subjects				
GWIS (G)	5	.003	110	.0007	4.002	.002	.154
R x G	5	.0009	110	.0007	1.357	.246	.058

TABLE E35: Crisp c ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.598	22	.547	1.092	.307	.047
			Within Subjects				
GWIS (G)	5	.470	110	.100	4.705	.001	.176
R x G	5	.010	110	.100	.103	.991	.005

TABLE E36: Fuzzy c ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	1.056	22	.604	1.748	.200	.074
			Within Subjects				
GWIS (G)	5	.283	110	.023	12.227	.000	.357
R x G	5	.005	110	.023	.236	.946	.011

APPENDIX E: (continued)

TABLE E37: Crisp Ceiling A' ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.062	22	.094	.660	.425	.029
			Within Subjects				
GWIS (G)	5	.070	110	.012	5.676	.000	.205
R x G	5	.006	110	.012	.462	.804	.021

TABLE E38: Fuzzy Ceiling A' ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.005	22	.016	.306	.586	.014
			Within Subjects				
GWIS (G)	5	.001	110	.0008	1.745	.130	.073
R x G	5	.0005	110	.0008	.649	.663	.029

TABLE E38: Crisp Ceiling c ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.421	22	.690	.610	.443	.027
			Within Subjects				
GWIS (G)	5	.670	110	.063	10.597	.000	.325
R x G	5	.074	110	.063	1.167	.330	.050

TABLE E39: Fuzzy Ceiling c ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.294	22	.438	.670	.422	.030
			Within Subjects				
GWIS (G)	5	.197	110	.029	6.786	.000	.236
R x G	5	.028	110	.029	.974	.437	.042

APPENDIX E: (continued)

TABLE E40: Crisp Visibility A' ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.007	22	.063	.109	.744	.005
			Within Subjects				
GWIS (G)	5	.030	110	.012	2.432	.039	.100
R x G	5	.015	110	.012	1.180	.324	.051

TABLE E41: Fuzzy Visibility A' ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	.009	22	.007	1.287	.269	.055
			Within Subjects				
GWIS (G)	5	.003	110	.002	1.462	.208	.062
R x G	5	.002	110	.002	.954	.449	.042

TABLE E42: Crisp Visibility c ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	8.456	22	1.061	8.054	.010	.268
			Within Subjects				
GWIS (G)	5	.415	110	.083	4.982	.000	.185
R x G	5	.061	110	.083	.735	.599	.032

TABLE E43: Fuzzy Visibility c ANOVA Table

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
			Between Subjects				
Rating (R)	1	3.781	22	.636	5.941	.023	.213
			Within Subjects				
GWIS (G)	5	.082	110	.023	3.497	.006	.137
R x G	5	.002	110	.023	.094	.993	.004

APPENDIX F:

INDIVIDUAL SDT METRICS

TABLE F1: Baseline SDT scores for individual observers

	Crisp				Fuzzy			
ID	HR*	FAR*	A'	c	HR	FAR	A'	c
				Instrument Pilots				
1	0.786	0.115	0.903	0.203	0.847	0.031	0.951	0.419
2	0.786	0.115	0.903	0.203	0.695	0.015	0.917	0.834
3	0.643	0.038	0.892	0.701	0.741	0.022	0.926	0.683
4	0.69	0.038	0.906	0.636	0.752	0.113	0.893	0.265
5	0.833	0.269	0.862	-0.176	0.645	0.01	0.906	0.969
6	0.738	0.115	0.887	0.28	0.883	0.039	0.959	0.288
7	0.69	0.115	0.871	0.351	0.674	0.028	0.906	0.73
8	0.929	0.346	0.879	-0.535	0.892	0.041	0.96	0.249
9	0.929	0.038	0.971	0.152	0.897	0.047	0.96	0.203
10	0.786	0.038	0.932	0.489	0.876	0.286	0.875	-0.293
11	0.786	0.038	0.932	0.489	0.759	0.033	0.927	0.565
12	0.929	0.423	0.855	-0.636	0.575	0.021	0.882	0.918
				Non-Instrument Pilots				
13	0.69	0.192	0.835	0.186	0.792	0.054	0.928	0.396
14	0.738	0.115	0.887	0.28	0.909	0.463	0.83	-0.622
15	0.929	0.115	0.949	-0.133	0.884	0.066	0.95	0.155
16	0.976	0.038	0.984	-0.106	0.864	0.019	0.96	0.489
17	0.833	0.038	0.945	0.401	0.886	0.035	0.961	0.301
18	0.643	0.115	0.854	0.416	0.666	0.069	0.885	0.529
19	0.881	0.346	0.856	-0.392	0.972	0.26	0.924	-0.634
20	0.643	0.038	0.892	0.701	0.737	0.032	0.921	0.606
21	0.976	0.269	0.923	-0.683	0.95	0.306	0.901	-0.57
22	0.738	0.192	0.854	0.116	0.629	0.021	0.897	0.851
23	0.881	0.038	0.958	0.295	0.817	0.027	0.945	0.515
24	0.881	0.269	0.883	-0.282	0.904	0.06	0.958	0.122

* Values are those obtained after a loglinear correction

APPENDIX F: (continued)

TABLE F2: Baseline Ceiling and Visibility SDT scores for individual observers

	Ceiling				Visibility			
	Crisp		Fuzzy		Crisp		Fuzzy	
ID	A'	c	A'	c	A'	c	A'	c
				Instrument Pilots				
1	0.7828	1.2592	0.9185	0.6652	0.8624	1.0269	0.5283	0.0513
2	0.5	1.8895	0.9054	0.5401	0.8882	0.346	0.7617	0.0816
3	0.8394	1.0187	0.9129	0.6733	0.8624	1.0269	0.67	0.0263
4	0.7529	0.5986	0.9074	0.0069	0.7978	1.3216	0.4083	0.0289
5	0.5	1.8895	0.882	0.8474	0.8926	0.8858	0.7867	0.0487
6	0.6818	1.6206	0.8589	0.7015	0.8505	-0.2215	0.85	0.2816
7	0.7828	1.2592	0.9233	0.4291	0.8348	0.0914	0.705	0.1816
8	0.7828	1.2592	0.8176	0.4369	0.9354	-0.3389	0.8817	0.2013
9	0.9218	0.6303	0.9698	-0.0681	0.853	0.163	0.8183	0.2013
10	0.8396	0.6019	0.8659	-0.1491	0.8806	-0.0976	0.7933	0.1763
11	0.8217	1.0944	0.9381	0.275	0.7847	0.5402	0.7267	0.0395
12	0.8217	1.0944	0.9402	0.3842	0.8312	1.168	0.78	0.0711
				Non-instrument Pilots				
13	0.7463	0.9048	0.8695	0.4111	0.8162	0.0269	0.7333	0.2289
14	0.6818	1.6206	0.7958	0.7749	0.8348	0.0914	0.6767	0.1763
15	0.9058	0.7158	0.9278	0.0676	0.9354	-0.3389	0.8817	0.1697
16	0.5	1.8895	0.9047	0.7381	0.8806	-0.0976	0.8217	0.3329
17	0.9058	0.7158	0.9245	0.2842	0.872	0.6363	0.7017	0.1211
18	0.6161	1.2004	0.8821	-0.1268	0.8882	0.346	0.8467	0.0395
19	0.7597	1.3552	0.9161	-0.0688	0.8399	-0.7966	0.9667	0.5658
20	0.8732	0.8708	0.9623	0.3601	0.6877	0.0116	0.73	0.4342
21	0.5	1.8895	0.7026	0.2862	0.9468	-0.2564	0.9383	0.2882
22	0.8732	0.8708	0.9382	0.297	0.6627	0.5272	0.5533	0.1316
23	0.9377	0.5344	0.9601	0.2478	0.9237	0.1571	0.8217	0.0592
24	0.5741	0.7248	0.9098	-0.127	0.9124	-0.4749	0.8817	0.3184

APPENDIX G:

EXPLORATORY CEILING BY VISIBILITY ANALYSES

Several additional analyses were performed to fully understand the nature of the ceiling by visibility interaction that occurred in the accuracy, comfort (in continuing the flight), and two RMSE dependent variables. The subsequent analyses are exploratory in nature and can provide additional insight into the nature of the interaction.

The two RMSE variables provide data on the magnitude of errors that occur in estimating ceiling and visibility. However, RMSE does not provide data on the direction of this error. An examination of the mean estimate for ceiling and visibility would show if pilots would provide data on the direction of pilot's error. An analysis of simple main effects was performed to investigate the effect of visibility at each level of ceiling on the pilot's estimate of ceiling.

Ceiling Estimation across Visibility

Visibility had a significant simple main effect at the 400, 900, and 4500 ft ceiling conditions. The ANOVA results showing the simple main effects are presented in Table G1. Post hoc analysis at the 400 ft ceiling condition found that pilots' estimations of ceiling was significantly higher at the 5 and 10-mile visibilities than the 2-mile visibility. The 3-mile visibility condition was also significantly lower than the 10-mile condition. Within the 900 ft ceiling condition estimates of ceiling were significantly higher at the 10-mile visibility condition compared to the 3-mile condition. The data and post hoc analysis for the 400 and 900 ft ceiling conditions is presented in Figure G1.

APPENDIX G: (continued)

TABLE G1: Simple Main Effects of Visibility at Each Level of Ceiling on Ceiling Estimation

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
Visibility at Ceiling = 400	3	7994913	66	858044	9.318	.000	.298
Visibility at Ceiling = 900	3	3318381	66	1050085	3.16	.030	.126
Visibility at Ceiling = 2900	3	1219513	66	996407	1.224	.308	.053
Visibility at Ceiling = 4500	3	12251181	66	3459751	3.541	.019	.139

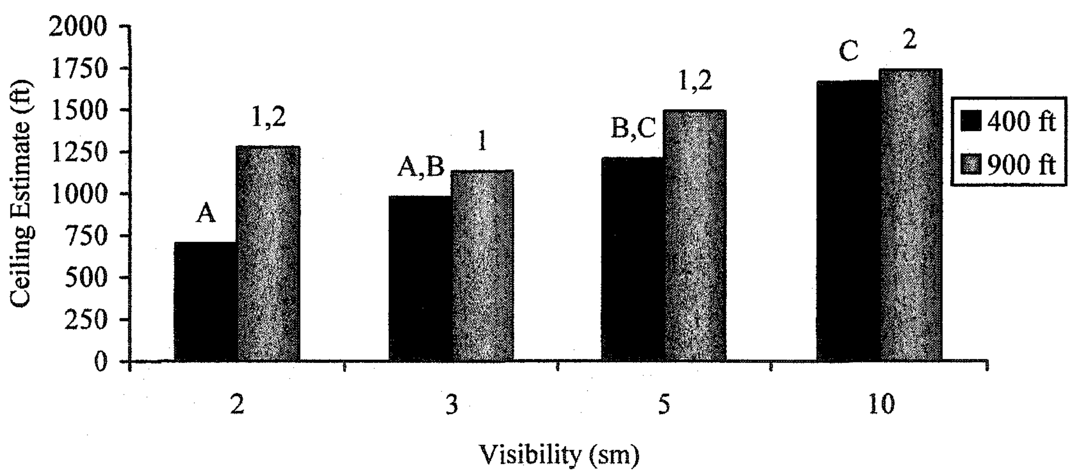


Figure G1. Mean ceiling estimation data for visibility at the 400 and 900 ft ceiling conditions. Comparisons should only be made within each ceiling condition (i.e., 400 or 900). Means with different numbers or letters are significantly different (at $p < .05$).

There was no simple main effect of visibility on ceiling estimation within the 2900 ft ceiling condition. Within the 4500 ft ceiling condition pilots estimate of ceiling at 2 miles was significantly lower than pilot's estimation of ceiling at 10 miles. The

APPENDIX G: (continued)

means and post hoc analysis data for the 2900 ft and 4500 ft ceilings is presented in Figure G2.

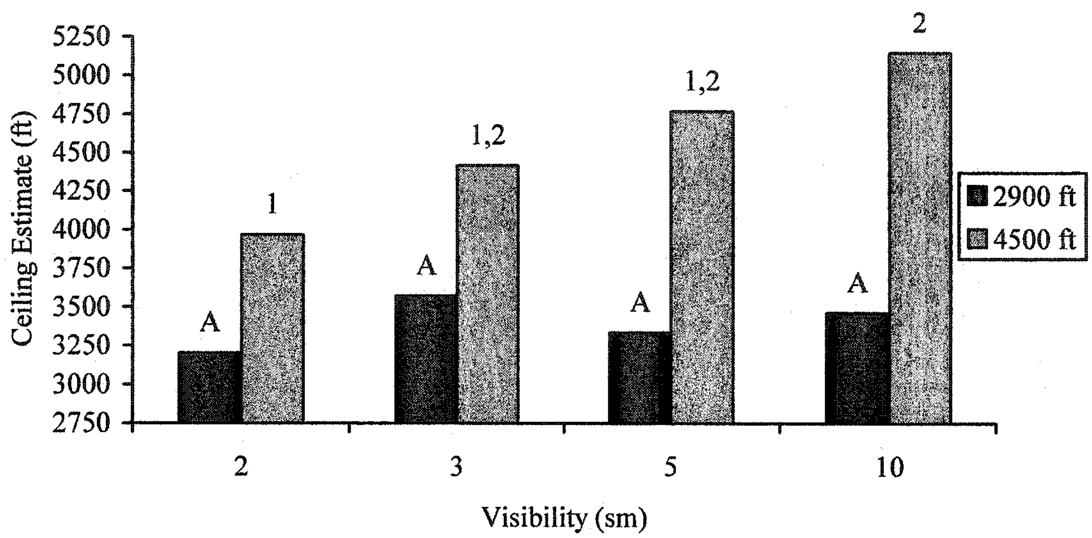


Figure G2. Mean ceiling estimation data for visibility at the 2900 and 4500 ft ceiling conditions. Comparisons should only be made within each ceiling condition (i.e., 2900 or 4500). Means with different numbers or letters are significantly different at $p < .05$.

Visibility Estimation across Ceiling

An analysis of simple main effects of ceiling was performed on the visibility estimation data at each level of visibility. The ANOVA Table for the simple main effects of ceiling is presented in Table G2.

The analysis of simple main effects at 2 miles visibility revealed that visibility estimates within the 400 ft ceiling condition were significantly lower than visibility estimates at any other ceiling. Additionally, the 900 ft ceiling condition was significantly lower than the 2900 ft and 4500 ft ceiling conditions. At 3 miles visibility, visibility

APPENDIX G: (continued)

estimates within the 400 and 900 ft ceilings were significantly lower than estimates made at the 2900 and 4500 ft ceilings. Means and post hoc analysis for ceiling at the 2 and 3-mile visibilities are presented in Figure G3.

TABLE G2: Simple Main Effects of Ceiling at Each Level of Visibility on Visibility Estimation

Source	df effect	MS effect	df error	MS error	*F*	p	Partial η^2
Ceiling at Visibility = 2	3	24.186	66	.585	41.362	.000	.653
Ceiling at Visibility = 3	3	77.116	66	1.280	60.255	.000	.733
Ceiling at Visibility = 5	3	95.424	66	1.974	48.337	.000	.687
Ceiling at Visibility = 10	3	138.630	66	3.597	38.536	.000	.637

The analysis of simple main effects at both 5 and 10 miles visibility revealed that visibility estimates were significantly smaller at the two IMC ceilings (i.e., 400 and 900 ft) than the two VMC ceilings (i.e., 2900 and 4500 ft). The means and post hoc groupings are presented in Figure G4.

Ceiling and Visibility SDT Analysis

It was assumed the bias metric from SDT would be able to provide the necessary insight into the direction of pilot's estimation error. Within the design of the current experiment, the overall SDT metrics could not examine bias at the different levels of ceiling and visibility. The overall analysis could not provide separate SDT metrics at the different levels of ceiling or visibility because they compose the signal and noise distributions necessary to calculate the SDT metrics.

APPENDIX G: (continued)

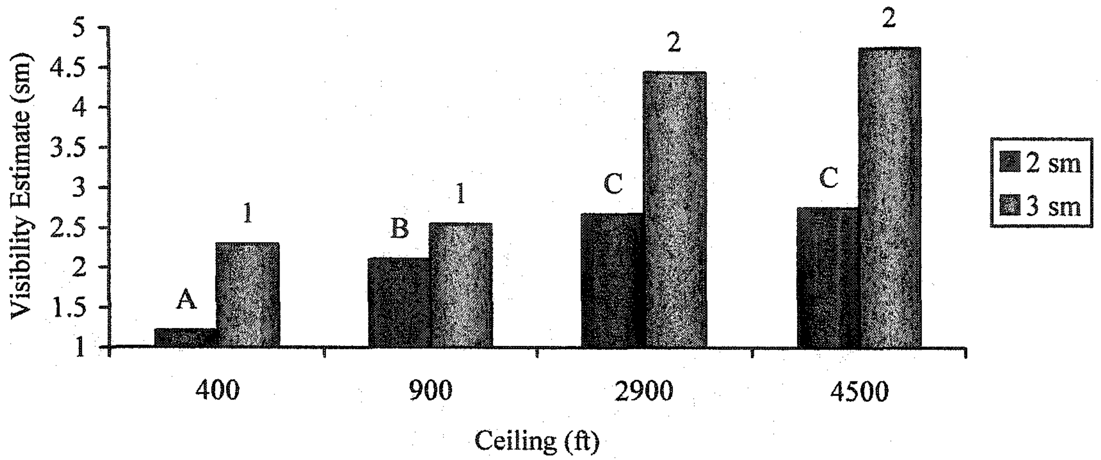


Figure G3. Mean visibility estimation data for ceiling at the 2 and 3-mile visibility conditions. Comparisons should only be made within each visibility condition (i.e., 2 or 3). Means with different numbers or letters are significantly different at $p < .05$.

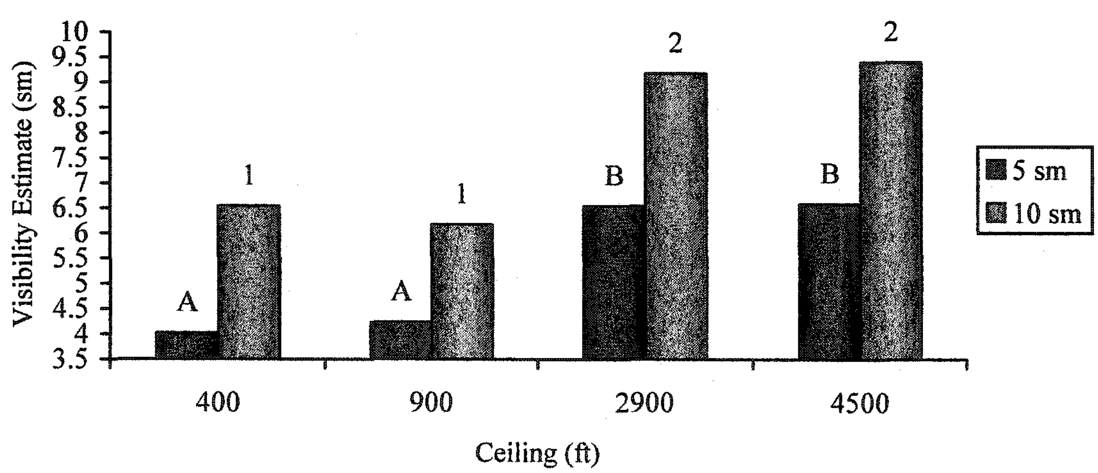


Figure G4. Mean visibility estimation data for ceiling at the 5 and 10-mile visibility conditions. Comparisons should only be made within each visibility condition (i.e., 5 or 10). Means with different numbers or letters are significantly different at $p < .05$.

APPENDIX G: (continued)

The calculation of separate SDT metrics for ceiling and visibility would allow for the examination of visibility SDT metrics at the different ceilings, and the ceiling SDT metrics at the different levels of visibility. However, the number of trials used to calculate SDT for each observer would be decreased by 75%. Due to this reduction in data an overall hit rate and false alarm rate were calculated across participants. SDT metrics were created from the overall data. Since only one data point was computed for each SDT metric no statistical comparison can be made across groups.

The means for the fuzzy and crisp ceiling sensitivity at each visibility is provided in Figure G5. There is a reduction in the crisp ceiling A' from the 2 and 3 mile visibility conditions and the 5 and 10 mile conditions.

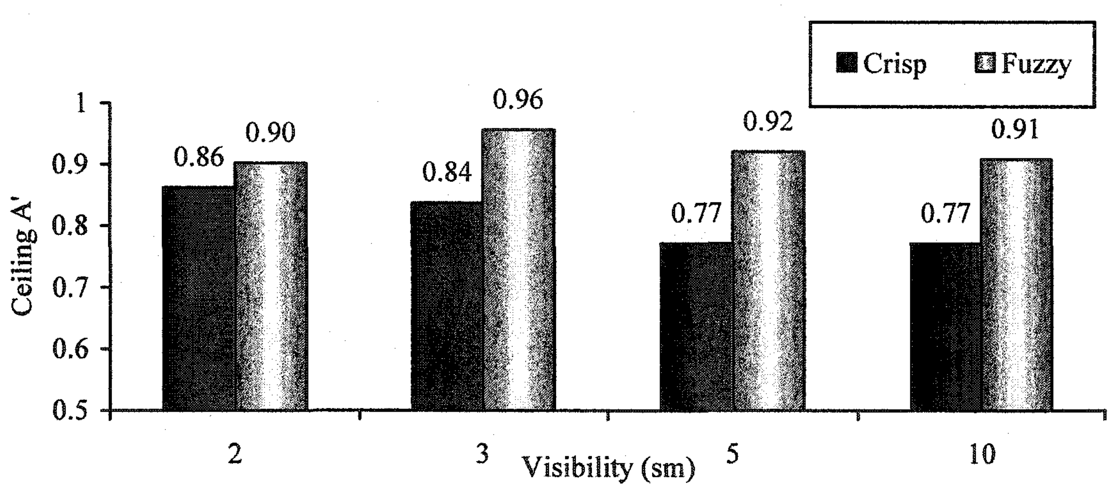


Figure G5. Fuzzy and crisp ceiling A' at each visibility.

Inspection of the crisp ceiling c shows a conservative bias at each level of visibility. The fuzzy ceiling c reveals a small liberal response bias at 2 miles with a

APPENDIX G: (continued)

progressively more conservative response bias as visibility increases. The ceiling response bias data is provided in Figure G6.

The means for the fuzzy and crisp visibility sensitivity at each ceiling is provided in Figure G7. The maximum difference in the crisp visibility A' between ceilings is .04, and the maximum difference in the fuzzy visibility A' is .01.

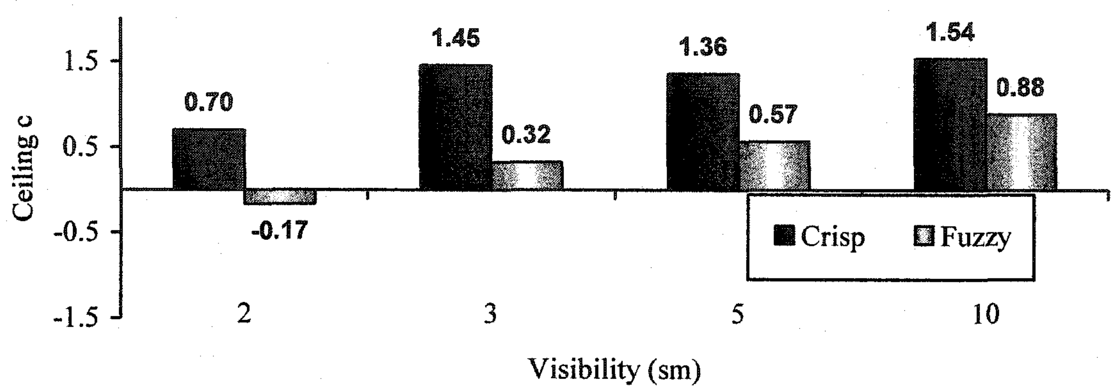


Figure G6. Fuzzy and crisp ceiling bias at each visibility.

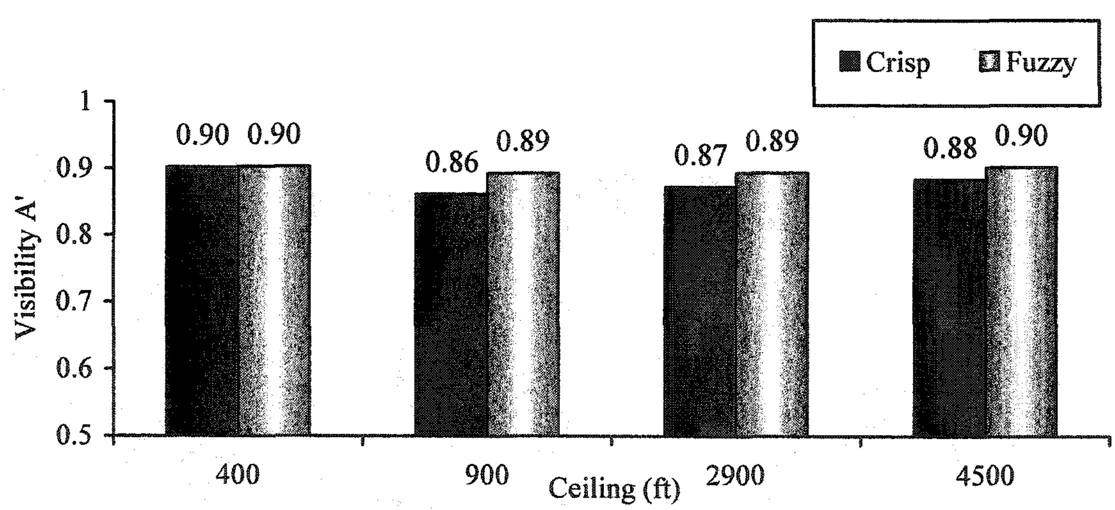


Figure G7. Fuzzy and crisp visibility A' at each ceiling.

APPENDIX G: (continued)

The crisp visibility c shows a liberal response bias for the two IMC ceilings (i.e., 400 ft and 900 ft) and a conservative bias at the two VFR ceilings (i.e., 2900 ft and 4500 ft). The occurrence of a liberal bias at the IMC ceilings and a conservative bias at the VFR ceilings was also present in the fuzzy c data. The visibility response bias data is provided in Figure G8.

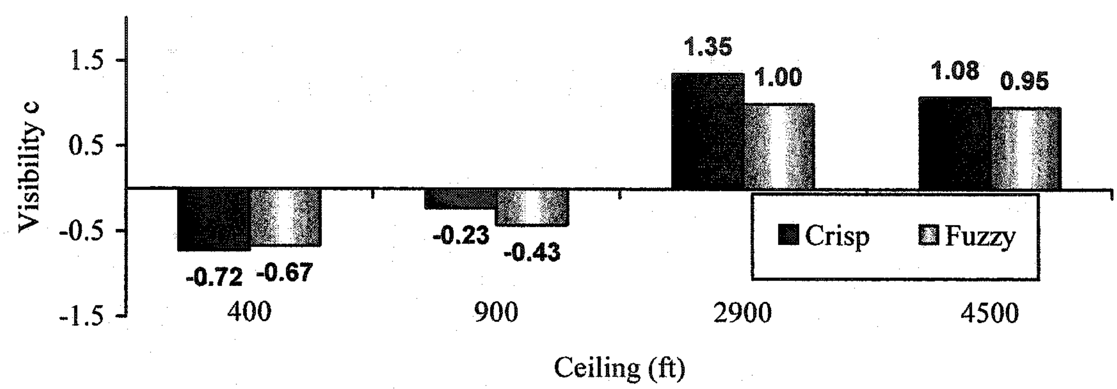


Figure G8. Fuzzy and crisp visibility bias at each ceiling.

A full discussion of the implications of the interaction of ceiling and visibility is provided in the discussion section of this paper. The analyses contained within this appendix are provided to supplement the discussion of the interaction.

VITA

Joseph T. Coyne
7420 Hampton Blvd Apt 3D
Norfolk, VA 23505

EDUCATION

Ph.D. Industrial/Organizational Psychology, 2004 - *Old Dominion University,* Norfolk, Virginia
M.S. Psychology, 2002 – *Old Dominion University,* Norfolk, Virginia
B.A. Psychology, 1999 – *University at Albany,* Albany, New York

AWARDS AND HONORS

- 10/01-10/04 NASA Graduate Student Researchers Program Training Grant NGT-01-01030
- 10/00-10/01 NASA Graduate Student Researchers Program Training Grant NGT-01-001
- Graduate Student Excellence Award, Old Dominion University, 2003-2004

SELECTED CONFERENCE PROCEEDINGS

Coyne, J. T., & Baldwin, C. L. (2003). Comparison of the P300 and other workload assessment techniques in a simulated flight task. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting,* 601-605.

Coyne, J. T., & Baldwin, C. L. (2003). ERP indices of mental workload for traditional and text-based ATC commands during a simulated flight task. *Proceedings of the 12th International Symposium on Aviation Psychology*, 261-266.

Coyne, J. T., & Lee, M. D. (2002). The effects of visual cues and interstimulus interval on accuracy in auditory localization and detection. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting,* 1613-1617.

Coyne, J. T., Lee, M. D., & Cunningham, A. (2001). The effects of speaker gender, frequency and distinctiveness on auditory search performance. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting*, 1348-1351.

Coyne, J. T., & Lee, M. D. (2000). Errors in multi-channel auditory search: Localization versus detection. *Conference Proceedings of the Human Performance, Situation Awareness and Automation Meeting*, 184-189.

PROFESSIONAL AFFILIATIONS

- Human Factors and Ergonomics Society (HFES)
- HFES ODU Chapter (Vice President, 2002)
- American Psychological Association, Division 21