

Fall 2004

The Effect of Crosswind and Turbulence in Mental Workload and Pilot Tracking Performance

Bruno E. Vivaldi

Embry-Riddle Aeronautical University - Daytona Beach

Follow this and additional works at: <https://commons.erau.edu/db-theses>



Part of the [Applied Behavior Analysis Commons](#), and the [Aviation Commons](#)

Scholarly Commons Citation

Vivaldi, Bruno E., "The Effect of Crosswind and Turbulence in Mental Workload and Pilot Tracking Performance" (2004). *Theses - Daytona Beach*. 207.

<https://commons.erau.edu/db-theses/207>

This thesis is brought to you for free and open access by Embry-Riddle Aeronautical University – Daytona Beach at ERAU Scholarly Commons. It has been accepted for inclusion in the Theses - Daytona Beach collection by an authorized administrator of ERAU Scholarly Commons. For more information, please contact commons@erau.edu.

THE EFFECT OF CROSSWIND AND TURBULENCE IN MENTAL WORKLOAD
AND PILOT TRACKING PERFORMANCE

by

Bruno E. Vivaldi

A Thesis Submitted to the
Department of Human Factors & Systems
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Human Factors & Systems

Embry-Riddle Aeronautical University
Daytona Beach, Florida
Fall 2004

UMI Number: EP32059

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[®]

UMI Microform EP32059
Copyright 2011 by ProQuest LLC
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

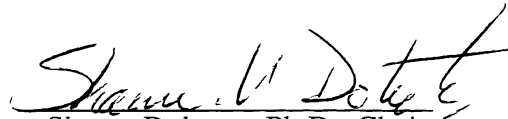
THE EFFECT OF CROSSWIND AND TURBULENCE IN PILOT TRACKING
PERFORMANCE

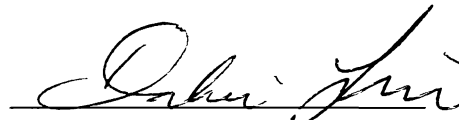
by

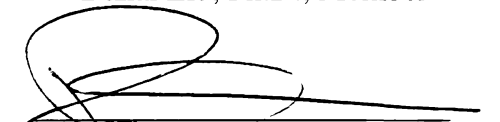
Bruno E. Vivaldi

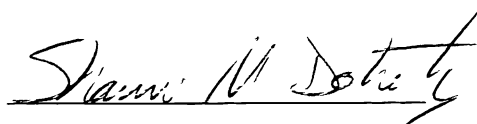
This thesis was prepared under the direction of the candidate's thesis committee chair, Shawn Doherty, Ph.D., Department of Human Factors & Systems, and has been approved by the members of the thesis committee. It was submitted to the Department of Human Factors & Systems and has been accepted in partial fulfillment of the requirements for the degree of Master of Science in Human Factors & Systems.

THESIS COMMITTEE:

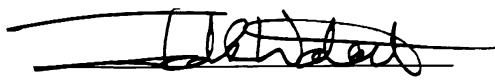

Shawn Doherty, Ph.D., Chair


Dahai Liu, Ph.D., Member


Richard P. Anderson, Ph.D., Member


MS HFS Program Coordinator


Department Chair, Department of Human Factors & Systems


Associate Chancellor of Student Academics

ABSTRACT

The purpose of this study was to evaluate the effect of crosswind and turbulence on mental workload and pilot tracking performance. Based on previous research, it was believed that as the amount of crosswind and turbulence is increased, mental workload would increase and tracking performance would decrease. The objective was to estimate the impact that crosswind and turbulence, of varying degrees, had on performance and workload. Fifteen full time college student volunteers served as experimental participants in a simulated horizontal and vertical tracking task. Each participant flew twelve instrument approaches, experiencing a different crosswind and turbulence combination during each approach. Flight performance and workload were measured using time within standard (TWS) and NASA Task Load Index (TLX) scores, respectively. The most detrimental effect on tracking performance was expected when participants were exposed to both crosswind and turbulence as the pilot had to divert attention between maintaining control of the airplane, establishing and maintaining a crab angle, and correcting for the aircraft being displaced off course in a continuous basis. The results of this study suggest that the impact of crosswind on tracking performance is small and probably not of practical concern. Similarly, the results did not find that crosswind statistically increased mental workload. However, as the turbulence level was increased, observed tracking performance decreased and workload scores increased. The results of the study failed to find a statistically significant interaction between the crosswind and turbulence factors for either the performance or workload data.

TABLE OF CONTENTS

ABSTRACT

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION	1
Human Performance and Workload	3
Flight Performance Measurement	6
Workload Measures	10
Projective Techniques	11
Empirical Measures	12
Task Performance	12
Psychophysical Approaches of Workload	13
Subjective Measures of Workload	14
Data Analysis	17
The Instrument Approach Procedure	17
Wind	21
Wind Terminology and Measurement	21
Causes of Wind	22
Pressure Gradient Force	22
Coriolis Force	23
Geostrophic Balance	24
Friction	24

Wind Review	24
Vertical Motions	25
Turbulence	26
Turbulence Metrics	26
Turbulence Causes and Types	29
Low Level Turbulence	29
Mechanical Turbulence	30
Thermal Turbulence	30
Turbulence in Fronts	31
Turbulence in and Near Thunderstorms	31
Clear Air Turbulence	32
Mountain Wave Turbulence	32
The Present Study	33
Anticipated Outcomes	34
METHOD	36
Participants	36
Materials and Apparatus	37
Other Equipment Used	37
Design	38
Procedure	39
Data Collection	40
RESULTS	41
Results for the Performance Data	42

Results for the Workload Data	44
Regression Equations	47
Discussion	48
Study Limitations and Recommendations	52
REFERENCES	55
APPENDICES	61

LIST OF TABLES

Table 1. Turbulence Reporting Criteria	28
Table 2. Quantitative Measures of Turbulence Intensity	28
Table 3. Descriptive statistics for the performance data.	42
Table 4. Mauchly's test of sphericity for the performance data.	42
Table 5. ANOVA source table for the performance data.	43
Table 6. Test of within subjects contrasts for the performance data.	43
Table 7. Descriptive statistics for the workload data.	45
Table 8. Mauchly's test of sphericity for the workload data.	45
Table 9. ANOVA source table for the workload data.	46
Table 10. Test of within subjects contrasts for the workload data.	46
Table 11. Regression equations for each trend component.	48

LIST OF FIGURES

Figure 1. Instrument Approach Segments	20
Figure 2. Average performance across turbulence level settings.	44
Figure 3. Average workload scores for turbulence.	47

INTRODUCTION

Through the review of human factors, aircraft accidents, and aviation literature, it is clear that pilot performance is critically affected by many variables. Some of these variables are well-defined, easy to study, and manipulate, while others are not and to date have yet to be fully understood.

A study conducted by Boeing in 1996 indicated that 62% of the accidents occur in the arrival phase and 31% in the departure phase of flight. It is in these phases where pilots are bombarded with an overload of information and are required to work as an efficient team to reach a desired goal. What factors influence their performance and challenge their ability to safely and efficiently execute their mission? And if these factors are identified, how and why do they influence performance and what can we do to minimize their risks? These are difficult questions that the scientific community has been working on for years.

The biggest challenge seems to be the enormous amount of information pilots has to process before making a decision and acting during certain segments of flight. Tsang and Wilson (1997) proposed that human operators have limited processing capabilities and once this limit is reached, performance will decrease. This can increase the chances for error and jeopardize flight safety.

Flying an instrument approach is a challenging task and the presence of a crosswind and turbulence does not make the task any easier. Instrument approaches are required when the weather conditions are below three statute miles and/or the ceiling (lowest broken, overcast or obscured layer of clouds) is below 1000 feet. The instrument landing system (ILS) allows pilots to fly aircraft into airports when they are “blind” to the

outside world due to marginal weather conditions. Instrument approaches rely on instrumentation in the cockpit that informs the pilot about the position of the aircraft relative to the ideal approach flight path. The instrumentation will indicate if the aircraft is high or low on the glide slope (vertical component) and if the aircraft is to the left or to the right of the localizer (lateral component). This information allows the pilot to make corrections for any off course deviations. If the course deviations go beyond a certain limit (10 degrees) the aircraft strays away from the desired path and loses positive course guidance. In the presence of a crosswind and turbulence it becomes harder for the pilot to maintain the aircraft on course and/or within the acceptable needle deviations. After the needle displays a full-scale deflection there no longer is guaranteed obstacle clearance and the aircraft can strike terrain, protruding obstacles, or other aircraft. This is obviously a hazardous situation where no pilot likes to be. This is likely to increase the pilots stress, mental demand, and frustration as he attempts to return the aircraft on course or executes an immediate missed approach. In other words, the pilot's mental workload is expected to increase. This study was interested in evaluating the effect of crosswind and turbulence on mental workload and tracking performance.

The problem at hand does not relate to physical work; instead, the major concern is information processing and decision-making. Some experts have agreed to call this "mental workload" and they believe this is the key to understanding human performance. Hart and Staveland (1988, p. 14) saw workload as "a hypothetical construct that represents the cost incurred by a human operator to achieve a particular level of performance". O'Donnell and Eggemeier (1986) viewed workload as that portion of the

operator's limited capacity actually required to perform a particular task. Both these definitions assume that human operators have limited processing capability.

Previous studies have used crosswinds and turbulence as factors to increase mental workload (Hahn, Heintsch, Kaufmann, Schanzer & Swolinsky, 1990; Oman, Rasmussen, Robinson, Huntley, et al., 1995; Oman, Kendra, Hayashi, Stearns & CoheBurki-Cohen, 2001; Ragsdale, Osborne & Seaman, 1974), but to date no published studies have evaluated if these variables really affect mental workload and if so, to what extent.

This study will attempt to answer the question, "how does crosswind and turbulence affect mental workload and pilot tracking performance?" The results of this study will allow for the impact of turbulence and crosswind on performance and workload to be estimated so that future research can alter workload through the systematic alteration of crosswind and turbulence.

Human Performance and Workload

Human performance can be affected by many factors including fatigue (due to sleep debt, time awake, time of day, circadian factors, etc.), alcohol, noise and vibration, temperature and humidity, mood, individual differences and training. Observed levels of performance are also impacted by perceived and real levels of workload. What is mental workload? How and why does mental workload affect performance? And how do we accurately measure workload? These questions have opened up a myriad of ideas and theories which have been scrutinized by many researchers, who over the years have strived to find the answers to these questions. As piloting tasks have become less routine, and especially more recently as the computer has become a cooperative party in decision-

making and control, the appropriate basis for defining mental workload has become more difficult. Sheridan and Simpson (1979) noted this is particularly true when talking about large transient workload demands that occur sometimes unexpectedly in flight.

There is not good agreement in the scientific community as to the cognitive mechanisms associated with mental workload or on which measures can best quantify the level of workload experienced, but most researchers agree that mental workload is an important factor that needs to be evaluated. As an example, Parasuraman and Mouloua (1996) proposed that workload is an important factor in the aviation domain for safety, performance, and efficiency reasons.

When we speak about mental workload, we normally mean something to do with a sense of mental effort or how hard one feels one is working. Mental workload appears to be a subjective state and therefore would vary depending on the individual. Cognitive workload is a transitory, subjective state that has no obvious direct manifestations (Charlton, 1996). This presents a problem when trying to measure mental workload.

It is important to talk about mental workload and its relation to task performance and task demand. Mental workload is not task performance and is not task demand. Sheridan and Simpson (1979) proposed that mental workload seems to be a combination of mental effort, information processing, and emotion in response to task demand. We would therefore expect different persons to have different responses to the same task demands and task performance. In addition, Wickens resource theory proposed that humans have a finite capacity to do certain things. Humans are able to multi task and some tasks can be performed concurrently in harmony, while other tasks will interfere with one another. The main idea is that there is a limited capacity to the human brain.

Hart (1986) showed that task performance and workload are not monotonically related. That is, an increase in workload will not necessarily result in a decrease in performance. For easy tasks (and maybe long, boring, and seldom changing vigilance tasks), increased workload will lead to an increase in performance (Gopher & Donchin, 1986; Lysaght, et al. 1989). Over a range of moderate workload levels, people are able to adjust their level of effort and maintain acceptable levels of performance. For moderately difficult tasks, participants may not be able to increase their effort enough to meet the task demands and thus increased workload is associated with poorer performance. Charlton (1991) observed that for very difficult tasks, participants were not able to continue expending the extra effort in face of what was perceived as unreasonable task demands; instead, they reduced their efforts and allowed task performance to deteriorate in order to return to normal levels of workload.

As mentioned earlier, another factor that can play an important role in understanding and measuring pilot mental workload is individual differences in motivation to perform and responsiveness to task demands. An individual's effort, and thus their cognitive workload, might not be consistent throughout a task. Different people use different strategies to get the job done. Because of these differences across individuals and tasks, many researchers have come to the conclusion that the subjective experience of workload is the collective result of several independent components such as task-related inputs loads, operator effort and motivation, and operator performance (Jahns, 1973). The good news is that extreme levels of workload will degrade performance and is therefore easily and reliably identified. These extreme levels are of

interest to most human factors testers when designing cockpits, cockpit displays, or anything else that would influence pilot performance by varying workload.

With the introduction of autopilot and flight management systems, the role of the pilot has changed from a physical function (flying) to that of a program and monitor function, thus increasing mental workload. Pilots have reported these increases in mental workload since the introduction of glass cockpits (Mouloua, Deaton, & Hitt, 2001). For many years, researchers have been seeking the balance between technology and the amount of mental workload the pilots can safely endure.

Traditionally the measurement of pilot workload is carried out through the use of simulators. Missions are performed in order to determine what amount of workload can be handled without a performance decrement. The key question is whether or not workload encountered in a simulator accurately reflects the workload encountered in actual flying conditions. Because there is no clear consensus to this question and because it is outside the scope of this study to examine this issue, it will be assumed that to a certain degree, workload in a simulated environment will reflect that experienced in actual flying conditions.

Flight Performance Measurement

Measuring flight performance is a difficult task. It is a topic that has been argued for decades and no single well-established method has been adopted to date. There are several methods to measure performance, each having advantages and disadvantages. The method chosen to measure flight performance should be the one that most accurately measures what it is supposed to measure. In other words, having a very precise method to measure performance might not be the best method to use if we have no interest in what

it is measuring. There are two very different ways to measure performance: expert performance ratings and objective performance measures.

Expert performance ratings come from instructors and examiners who over the years have gained invaluable, hands-on experience as to what constitutes satisfactory or unsatisfactory performance for a given task. This is a highly subjective method that places a lot of emphasis on the instructor and/or the examiner. The Federal Aviation Administration (FAA) designed the Practical Test Standards (PTS) for each certificate and rating as a book value, a standard, for all examiners to follow and evaluate pilot performance against. This book describes the minimum performance that should be observed during an evaluation in order for the maneuver/knowledge area to be considered satisfactory or unsatisfactory. However, given the nature of the flying task and the innumerable variables that affect an aircraft in flight (wind, turbulence, visibility, etc.) the PTS allows the examiner the authority to “adjust” the standards for less than favorable flying conditions. It clearly states that the PTS is based on an aircraft flying in a clear, VFR day with no wind/turbulence. Therefore, for less than perfect days (most of the time) the examiners “grey area” is expanded and greater deviations are allowed when performing maneuvers. The PTS also clearly states how the book should be used, and what constitutes satisfactory and unsatisfactory performance. It indicates that deviating from the standards is not necessarily an automatic unsatisfactory for that maneuver. The applicant is expected to establish prompt corrective action from the deviation. This would be considered satisfactory performance. However, constantly exceeding the standards is considered unsatisfactory. The problem is that there is no clear guideline as

to what constantly exceeding the minimums means. This is what makes expert ratings a highly subjective method of measuring flight performance.

A partially subjective metric is time to completion. This metric implies that in order for an individual to reach a certain point in his or her training (first supervised solo, certification check ride) he or she needs to have achieved a minimum level of performance. This metric is objective because it represents the number of flight hours needed to reach this milestone but it is influenced by the subjective evaluation of the instructor. Several things can influence the time it takes an individual to reach a certain milestone (e.g. solo) such as location of training (how busy the airport he or she operates in), weather patterns in the area, student confidence and commitment, instructor commitment, availability of aircraft, etc. This metric is often used to when comparing different training strategies or assessing the impact of training tools like simulators.

Objective methods are usually based on raw deviations from actual flight path to measure flight performance. These are called Flight Technical Error (FTE) measures. One example is the Root Mean Square Error (RMSE), in which deviations from standard are squared in order to eliminate polarity and exaggerate gross deviations, and then averaged across the sample data. The square root of the average is then computed in order to return the metric to its original unit. The advantage with using RMSE is that it is very sensitive to flight path management. On the other hand, it cannot be interpreted as a single average deviation because of square root transformations giving more weight to gross deviations. Therefore, RMSE cannot be compared to performance standards in order to determine if a level of performance was met. The biggest problem associated with using RMSE as a performance measure is that effect sizes expressed in RMSE units

are difficult to interpret in terms of performance differences. This means that a person not familiar with RMSE would have a hard time understanding the data and drawing meaningful conclusions from them.

Other deviation based performance metrics include the number of deviations (ND) outside of tolerance. This metric must be used carefully as an aircraft may stray off course only once but stay there for the duration of the instrument approach. Therefore, ND values must be interpreted taking into consideration the amount of time (TD) spent outside the tolerance.

A similar metric to time spent outside tolerance (TD) is the time within standard metric (TWS). The main difference is that TWS focuses on the amount of time within standard and TD focuses on the amount of time spent outside the standard. The goal of the TWS metric is to quantify performance relative to known standards (PTS). TWS data can be compartmentalized across various flight parameters such as glide slope tracking, localizer tracking, airspeed, etc. This makes this metric easy to use and interpret. Its drawback is it is prone to ceiling effects and it is not as sensitive a measure of performance as RMSE, meaning that small changes in performance are not likely to be detected. This is especially true for a very easy task where a more sensitive measure such as RMSE needs to be used to detect small flight performance changes. The sensitivity of TWS is primarily a function of the standards employed. If tighter tolerances are used (i.e. the ATP PTS over the Instrument PTS), the measure will become more sensitive to a point. The key is to use standards that are appropriate given the population under study so that variability in performance scores is observed. The advantage of TWS over RMSE is that the participant and end users can examine the TWS numbers and determine whether

or not the impact of the treatment is substantial and meaningful. Even though TWS is less sensitive than RMSE, one can argue that differences in performance that cannot be detected by TWS are of little practical consequence. Another advantage of the TWS metric is that multiple standards can be summarized in a single outcome metric. For example, airspeed, altitude, and heading standards can be simultaneously employed and a single TWS number can be used to summarize pilot performance if desired. In contrast, doing something similar with RMSE would require the application a series of mathematical transformations to the data in order to generate some standardized performance score.

Because the flight task in this study is not expected to be easy and due to its simplicity and interpretability, TWS will be the metric used to determine the impact of crosswind and turbulence on pilot performance. Also based on flight and simulator experience, participants are not considered experts and are therefore not expected to perform as such. This should allow for variations in performance and the ability for the experimenter to determine whether or not the impact of outside forces (winds and turbulence) is substantial and meaningful.

Workload Measures

Charlton (1991) described two functionally different types of workload measures: projective techniques and empirical techniques. Projective techniques are used before performing the experiment and attempt to predict the levels of workload that will result under specific conditions. Empirical techniques measure workload during or after the completion of the task. The empirical techniques are the most frequently used in test and evaluation applications.

Projective techniques

These are not a substitute for assessing workload but they can provide valuable indications of potential workload problems. Examples of these techniques include: Time-line analysis, scheduling theory, and pro-SWAT. A brief description of each one of these techniques is described below.

Time-line analysis is used to identify how long tasks and task components will take and if they can be accomplished in the allotted time. It is assumed that if the sum of all of the task times is less than the time available then there will be some operator “slack time” and therefore less potential for operator overload. This procedure is straightforward and fairly simple to perform.

Scheduling theory is based on the belief that time pressure is the major source of cognitive workload. In a manner similar to time line analysis, scheduling theory compares task times to the time available for completion. The primary advantage is the identification of optimal task sequences.

Pro-SWAT (projective application of the subjective workload assessment technique) is essentially a role-playing exercise where the subjects “project” themselves into the system tasks one at a time and complete SWAT ratings. The procedure usually involves a detailed briefing on each task function, some level of equipment mockup, as well as an extensive debriefing in which tasks receiving high workload ratings are discussed in detail. Pro-SWAT offers a relatively low cost method for assessing the workload of developmental systems.

Empirical measures

These measures of workload involve a collection of data from one or more subjects actually performing the task(s). There are a variety of empirical measures available. Examples of these include workload questionnaires, task performance data, and physiological responses.

Task performance. This technique is widely used in workload assessment. Two measures have traditionally been used: primary task performance and secondary task performance. The noted decrement in performance on a primary task is said to be indicative of mental workload. However, Sheridan and Simpson (1979) pointed out several criticisms, mostly directed towards factors such as realism, the test subjects used, the methodology, the equipment variables, simulator fidelity and so on. Therefore use of performance on a primary task as a method of measuring mental workload has many problems. In addition to those already mentioned, one important problem is the lack of generalizability of methods and results. According to Lysaght et al. (1989), because tasks are usually unique to each system, nearly every situation requires its own measure of task performance. Another disadvantage with this method is that when analyzing the performance results, they cannot be generalized to different populations, only the specific group age and experience that was used to generate those results. As an example we might be interested in studying the effect of individuals' mood in short and long-term memory. If we use a sample group of people with ages 18 to 25, we could not draw conclusions based on those results about the effect of mood on memory for people between ages 40-45. Therefore, this method lacks global sensitivity and transferability. Another disadvantage is that workload levels might rise without any degradation in task

performance. This is called dissociation or insensitivity. In addition, workload and performance are not necessarily linearly related. Conversely, high workload levels cannot be inferred from poor task performance. Factors such as training, motivation, communications, user interface problems, and others could be the reason why the subject is not performing successfully and have nothing to do with excessive workload levels. In order to minimize these problems, participants must be trained appropriately and feel completely comfortable with the equipment used for the study. Proper communication between the researcher and the participant should also help minimize any erroneous data. This can be achieved with a thorough briefing about the objective of the study and the methods and procedures to be used throughout the trial runs.

The other performance approach to workload assessment is to measure subject performance on a secondary task as an indication of spare mental capacity. The idea behind using a secondary task is that if the subject is only partially loaded while performing the primary task, performance on a secondary task should remain efficient. As the requirements of the primary task increase, it is expected that the secondary task performance decrease. There are many problems associated with the use of a secondary task. Some researchers have argued that the introduction of a secondary task changes the nature of the primary task and therefore contaminates any measure of workload obtained (O'Donnell & Eggemeier, 1993). An alternative to introducing a secondary task would be to find an embedded task, a concurrent operator task that already exists in the operations procedures (Weirwille & Eggemeier, 1993).

Psychophysical approaches of workload measurement. This technique offers the potential for objective measurement of some physiological correlate of mental workload

and release the researcher from reliance on self-reports and questionnaires. Some of the psychophysical measures used are: heart rate variability, respiration rate, galvanic skin response, pupillary diameter, biochemical changes in blood and urine, electroencephalogram changes, changes in frequency spectrum of voice, and eye movement recording. The problem with these measures is that they can all be affected by stress, diet and other factors. Sheridan and Simpson (1979) proposed that these physiological indices measure something in a very scientific way, the question is whether what they measure is correlated with what we think as mental workload.

Subjective Measures of Workload. By far the most frequently used measures of workload are subjective methods. There are a number of different subjective measures available, ranging from simple, unidimensional scales that provide a single measure of overall workload to multidimensional scales that measure various components of workload. Four of the most commonly used workload scales include: The Cooper-Harper and its derivatives, the Subjective Workload Assessment Technique (SWAT), the NASA Task-Load Index (NASA TLX), and the Crew Status Survey. Charlton (1996) pointed out that each of these measures have shown to be sensitive to changes in workload levels, minimally intrusive, diagnostic, convenient, relevant to a wide variety of tasks, and possess a high degree of operator acceptance. Below is a brief description of the above mentioned subjective measures of workload.

The Cooper-Harper Scale is one of the earliest standardized scales used for measuring workload. The scale is a decision tree that leads the pilot to one of ten ordinal ratings. The primary advantages of the scale is that it is well known in the testing community, easy to use, and the resulting ratings correlate highly with other, more

sophisticated workload scales. Originally developed to evaluate aircraft handling characteristics, the Cooper Harper Scale has been modified to situations outside the aircraft piloting domain. The resultant Modified Cooper Harper Scale provides a sensitive measure of overall mental workload for a wide variety of tasks. This test is typically administered to subjects at the end of the test. The disadvantage with this scale is that the quality of the workload measure will depend on the subject's recollection of the event of interest.

The Subjective Workload Assessment Technique (SWAT) was developed by the U.S Air Force Armstrong Aeromedical Research Laboratory. SWAT is a multidimensional view of workload comprised of mental effort, time load, and psychological stress. Before this technique can be used, the scale must be normalized for each subject. During this first phase called scale development, subjects rank 27 combinations of three different levels of time load, mental effort, and stress by means of a card sorting technique. Then a rule is established for combining the three dimensions for each subject. Once the rule has been established, conjoint scaling is applied to develop an appropriate unidimensional workload scale that ranges from 0 (no workload) to 100 (highest workload). During the data collection phase, subjects provide time load, mental effort, and stress levels at predetermined times during the activity. SWAT is considered to be a reliable, well-developed, and valid measure of workload. However, SWAT requires significant amounts of preparation of materials and pretraining of subject's prior to use.

Crew Status Surveys (CSS) were designed at the U.S Air Force School of Aerospace Medicine to be easily understood by the subjects, easy to administer, and

readily understood by the tester. It has shown to be a sensitive measure across a variety of tasks and to correlate well with other workload measures (Charlton, 1991). The survey involves three ratings: subjective fatigue, maximum workload, and average workload. It is administered throughout the participant's flight mission. The survey's main advantage is that it can be completed very quickly, it is simple, and it has shown agreement with the other workload measures.

The NASA Task Load Index (NASA TLX) is also based on a multidimensional approach to workload and uses an adjustment to normalize ratings for each subject. TLX divides the workload experience into six components: mental demand, physical demand, temporal demand, performance, effort, and frustration. TLX also divides each component subscale into twenty levels. TLX uses a simpler weighing procedure for combining information from the six subscales. Subjects are asked to make pair wise comparisons of each subscale as to which is more relevant to workload for a particular task. The number of times a subscale is chosen over another is used as the weighting for that subscale. Workload scores are computed by multiplying the rating obtained for each subscale by its task weighting, then adding up all of the subscale scores and dividing by the total number of paired comparisons used to obtain the weights. The use of weighted scores over the unweighted averages serves to reduce between subject variability (Hart & Staveland, 1988). The factor structure of the NASA-TLX has recently been questioned by Bailey and Thompson (2001) and Hall, Landa, Hart, and Karkman (2003); specifically, these authors have proposed that the TLX is really only assessing one or two separate factors and not the six as claimed by the authors of the TLX. On the other hand, Hall et al. (2003) and Hall, Doherty, French, and Landa (2004) noted that scores on the TLX did

vary as expected with changes in flight task difficulty providing some evidence for construct validity. Accettullo (2004) used the Total TLX scores as produced by the TLX, arguing that high intercorrelation among the sub-scale scores make using the Total TLX scores the best choice since they represent the combination of all sub-scale scores. Because of its simplicity, high degree of operator acceptance, and its sensitivity in detecting changes in workload levels in a variety of tasks, the NASA TLX was selected as the measure of workload for this study.

Data Analysis

Traditionally, hypothesis testing has been used to determine whether or not a treatment produces significant results beyond chance deviation. While this might sound useful, it actually provides very little information about the impact of the treatment in question. The focus of this study is to estimate the extent to which crosswind and turbulence impact performance and workload. Therefore, several layers of data will be presented for each analysis. First descriptive statistics such as sample size, mean, and standard deviation will be provided. Second, inferential analysis will be presented. Third, when group scores are discussed, confidence intervals will be created around those mean scores. Fourth, confidence intervals for observed mean differences will be presented. And fifth, standardized effect size estimates will be provided to help the reader and interpret the magnitude of group mean differences.

The Instrument Approach Procedure

As previously stated, most accidents occur during the approach and landing phase of flight where pilots become bombarded with an overload of information. Adverse weather conditions, low ceilings (lowest layer of broken or overcast clouds), and/or low

visibilities force pilots to have to “shoot” an instrument approach procedure (IAP) into the airport. There have been many accidents that have occurred as a result of poor execution of these IAPs by a flight crew. Understanding task demand during the IAP and analyzing the different elements that pilots complete during the IAP will help us predict where workload levels might be too high especially if variables such as wind and turbulence are present. A review of the causes, types, and effects of crosswinds and turbulence is also provided. Understanding the nature of crosswinds and turbulence will help develop a better understanding of the impact these variables have in a tracking task and the demands imposed on pilots.

The standard instrument approach procedure allows the pilot to descend safely by reference to instruments from the enroute altitude to a point near the runway at the destination from which a landing can be made visually. A precision approach procedure provides vertical guidance through means of an electronic glide slope, as well as horizontal course guidance. A non-precision approach provides horizontal course guidance with no glide slope information. Although there are many different types of approaches in use, most incorporate common procedures and chart symbology. An instrument approach may be divided into as many as four approach segments: initial, intermediate, final, and missed approach. The purpose of the initial approach segment is to provide a method for aligning the aircraft with the approach course. This is accomplished by using an arc procedure, a course reversal, or by following a route which intersects the final approach course. The initial approach segment begins at the initial approach fix (IAF) and usually ends where it joins the intermediate approach segment. The pilots have spent time preparing for the IAP by setting the aircraft navigation and

communication radios, as well as loading up the approach into the GPS (where applicable), and performing oral briefings of the altitudes, procedures, and airspeeds to be flown. In turbulent conditions this is often difficult as the aircraft experiences changes in flight attitude and forces the crew to divert their attention between flying the airplane and preparing for the approach procedure. The intermediate segment is designed to position the aircraft for the final descent to the airport. On this segment, the crew typically reduces the aircraft's airspeed to the approach speed, set the aircraft configuration for the approach (flaps), complete the before landing checklist, and make a final review of the approach procedure and applicable minimums. The intermediate segment begins at the intermediate fix (IF) or at a point where the pilots are proceeding inbound to the final approach fix. Winds and turbulence often make this task difficult as the pilots are now attempting to track the approach course and prevent the aircraft from getting off course. The division of attention between maintaining aircraft control in turbulence, tracking the final approach course inbound with a crosswind, and performing the pre-landing checklists, make this segment a very busy time for the pilots. It is here where it is easy to "fall behind the airplane" and become overloaded by all the different variables affecting the aircraft in flight. The final approach segment begins at the final approach fix (FAF) or at a point where the aircraft is established on the final approach course. The purpose of the final approach segment is to allow the pilot to navigate safely to a point at which if the required visual references are available, the pilot can continue the approach to a landing. As the aircraft approaches the runway the navigation signals becomes more sensitive and therefore it is easier for the flight crew to get off course and overcorrect for these deviations. Being close to the ground, talking to the control tower for a landing

clearance, making any final configuration changes (landing gear) and dealing with a very sensitive navigation signal makes this segment of the approach a critical one. Winds and turbulence during this segment can be expected to add to the approach difficulty and in some cases push the pilot's capabilities to safely cope with the situation. If the required cues are not seen at the missed approach point the pilots are forced to execute a missed approach procedure (MAP). The purpose of the missed approach segment is to allow the pilot to safely navigate from the missed approach point to a point where another approach can be attempted, a holding pattern can be entered, or a diversion to another airport can be commenced. Figure 1 shows a visual representation of the approach segments.

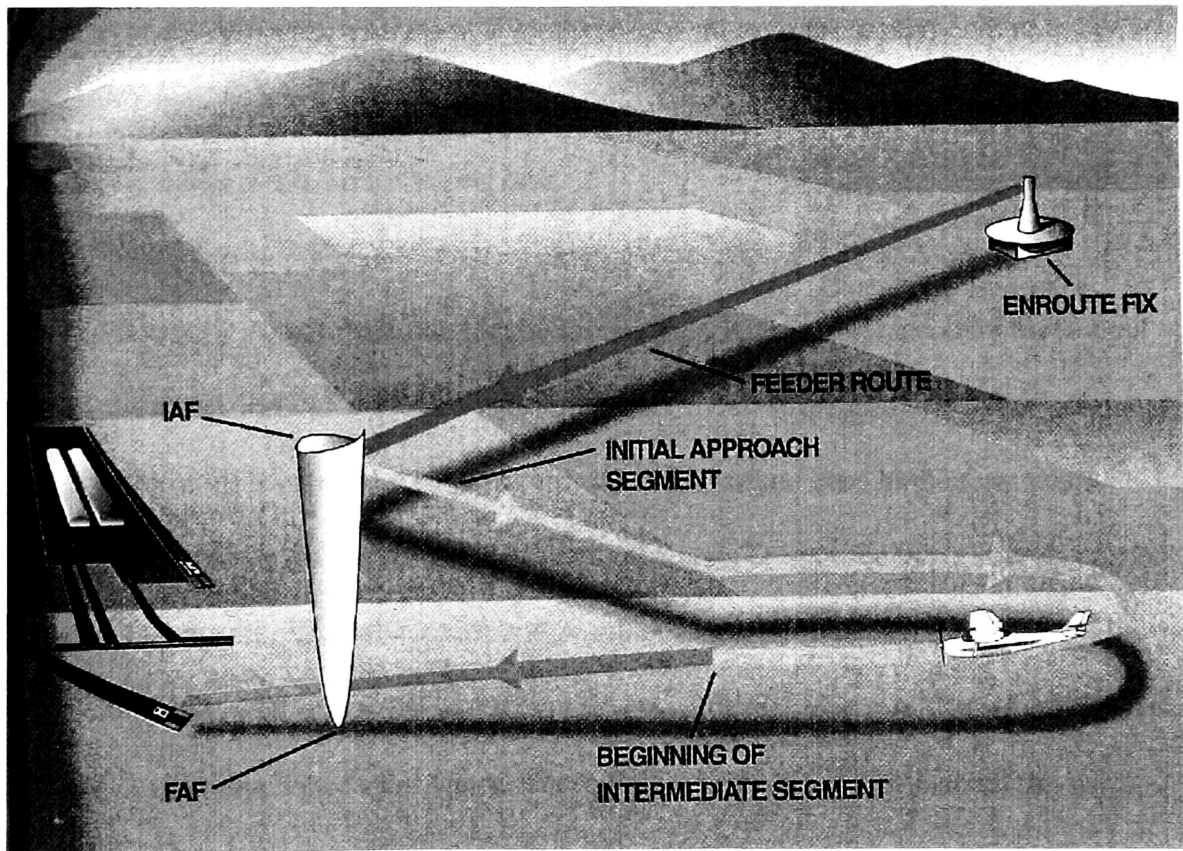


Figure 1. Instrument approach segments. (Source: Willits, 1998).

Wind

The motion of air is important in many weather-producing processes. Moving air carries heat and moisture from one place to another. Air movements can create favorable conditions for the formation and dissipation of clouds and precipitation; in some cases those motions can cause the visibility to decrease to zero; in others they sweep the skies clear.

In flight, winds can have a significant effect on navigation. Erratic air motions cause turbulence which at worst can be catastrophic. Without a question the pilot must understand air motions for efficient and safe flight.

The following is a description of the causes and characteristics of horizontal motions of the atmosphere. Having a deep understanding of the subject will help in the predicting the effect that these horizontal winds will have on the pilot's mental workload and performance while on a tracking task.

Wind Terminology and Measurements

When air moves from one location to another, it can simultaneously move both horizontally and vertically. Horizontal motions are much stronger than vertical motions with the exceptions of a few turbulent phenomena described later in the turbulence section. Also, horizontal motions are easier to measure. During the experimental portion of this study we will be dealing with both vertical and horizontal motions.

Wind is measured at the surface by several different methods. The most common include anemometers and wind vanes. These will provide information about the wind speed (usually expressed in knots) and wind direction (relative to true north). For winds

aloft measurement techniques include free balloons, Doppler radar, aircraft navigation systems, and satellite.

Causes of Wind

In order to understand what makes the wind blow we must ask ourselves two questions: What are the forces that affect air parcels (and therefore make them move), and what are the causes of these forces? The most important forces that affect air motions are: Pressure gradient force, Coriolis force, and Frictional force. A brief discussion of these forces and their causes is provided below.

Pressure Gradient Force

The concept of pressure gradient is better understood when you deal with gases under pressure. For example, if you inflate a tire you establish a pressure gradient across the thickness of the tire. If you puncture it, the air accelerates from the inside to the outside, that is, toward lower pressure. The larger the pressure difference, the greater the acceleration. The force involved here is known as the pressure gradient force. When a horizontal pressure gradient force exists, the atmosphere causes air parcels to be accelerated across the surface towards low pressure. This is the root cause of what we most commonly know as “wind”. But what causes the pressure gradient? In simple terms: uneven heating of the earth’s surface. For example, when the temperature of the land and sea are equal there is no horizontal pressure gradient and therefore no movement across the coastline. As the sun continues to heat the earth’s surface, the land temperature will exceed the water temperature (differential heating). The warmer land heats the overlying air causing it to expand and commence to rise. Because pressure decreases more rapidly with height in cold air than in warm air, the warm air aloft will have a higher pressure

than that one over the water. This causes a horizontal pressure gradient aloft and a movement of warm air toward the lower pressure over the sea. Interestingly, as soon as the mass leaves the upper part of the heated column, the weight of that column decreases and the surface pressure goes down over the land. This creates a second horizontal pressure gradient and air will start to move across the coastline from the sea toward the land. By simply creating a temperature difference between the two locations the air has been caused to move in one direction aloft and in the opposite direction at the surface.

Coriolis Force

Since we observe all motions from a rotating frame of reference (due to the earth's rotation) the effect of that rotation must be taken into account when explaining these observed motions. Coriolis force is a force created as a result of the earth's rotation. It affects all objects moving across the face of the earth. It influences such things as currents, airplanes, and even moving airmasses. Even though a deep analysis of Coriolis force is beyond the scope of this study, it is worth mentioning a few interesting points. This force affects only wind direction, not wind speed. It requires air to be moving and as wind speed increases so does the Coriolis force. It also depends on the latitude, Coriolis varies from zero at the equator to a maximum at the poles. Although this force affects air motion in all scales, in comparison to other forces its effect is minimal for small-scale circulations and very important for large-scale wind systems. Therefore, Coriolis force should not play an important role in this study and is not worth simulating given its small effect on local winds.

Geostrophic Balance

A useful characteristic of the atmosphere is that the pressure gradient force and the Coriolis force tend to balance each other. Coriolis and pressure gradient forces tend to be equal in magnitude but opposite in direction. This is known as the geostrophic balance. It is helpful understanding the characteristics of wind and it provides a good approximation to the actual wind. However, geostrophic balance does not occur in small-scale circulations such as sea breezes (described earlier) and therefore should not be considered as an important factor affecting this study.

Friction

Friction is the force that resists the motion of two bodies in contact. Surface friction is the term used to describe the resistive force that arises from a combination of skin friction and turbulence near the earth's surface. The primary effects of surface friction are experienced through the lowest 2000 feet of the atmosphere. This is called the boundary layer. Surface friction will slow down the wind speed and change its direction anywhere from 10 degrees to 45 degrees.

Wind Review

There are a few important points about wind worth reviewing. In the Northern Hemisphere, wind speed increases with altitude and changes direction clockwise due to Coriolis force. When the winds near the surface are strong the boundary layer is turbulent and the winds are gusty (changing direction and/or speed rapidly). The boundary layer is deeper during the day and in the warmer months of the year. Wind is caused by pressure differences and modified by the earth's rotation and surface friction. This study will model wind so as to simulate a constant wind speed and direction. Even though in real

life there are factors such as surface friction, Coriolis force, and even low-level turbulence in the boundary layer, these factors are not considered significant for this study and will be disregarded. Other factors that will be disregarded are winds produced by vertical motions (thunderstorms), winds around mountainous terrain, and extreme weather phenomenon's such as tornados.

Vertical Motions

As discussed earlier, when an air parcel moves from one location to another, it typically has a horizontal component (wind) and a vertical component, which is called vertical motion. Vertical motions are usually much smaller than horizontal motions, except on some extreme cases. Air may move upward due to a number of causes. The most frequent ones are convergence, orography, fronts, and convection. In the Northern Hemisphere, around large low pressure areas at the surface, the winds will spiral into the center (convergence) and therefore tend to rise. This will cause the upward motion of air. Another simple way to make air rise is by putting an obstacle on its way, such as a mountain. This is called orographic lifting. When the atmosphere itself creates an obstacle to the wind, a similar barrier effect can be produced. For example, when a cold air mass is next to a warm air mass, a sloping boundary is created between the two. This is called a front. If either air mass moves towards the other, the warm air moves upward in a process called frontal lifting. Also, if at a particular level in the atmosphere the air in the atmosphere is warmer than its surroundings, it will rise. This process is a form of convection and can cause vertical motions that can disturb an aircraft in flight. Usually we also find turbulence associated with the convection process and this can add another variable for the pilot to control.

Turbulence

Aviation turbulence is best defined as “bumpiness in flight”. It is important to note that this definition is based on the response of the aircraft rather than the state of the atmosphere. The magnitude on the bumpiness not only has to do with the outside factors but also depends on aircraft design and pilot reactions. Lester (1997) observed that in general, vertical gusts are more likely to have a larger impact on flight than horizontal gusts because they change the angle of attack and lift. However, in some situations (takeoff and landing), horizontal gusts may be as important as vertical gusts. With strong horizontal gusts the airplane has a tendency to weathervane into the wind. With the airplane not lined up with the runway, the pilot must correct by applying the appropriate rudder and aileron inputs to straighten the nose of aircraft while tracking the runway centerline to prevent a sideload on landing. With variable horizontal gusts, the rudder pressures required to maintain proper runway alignment would also vary, adding an additional challenge to that of landing the airplane without a crosswind.

An important issue to consider when talking about aviation turbulence is pilot fatigue. A pilot exposed to turbulent conditions for long periods to time will experience greater fatigue. Also, when the frequency of shaking is very large (4-5 cycles per second), the pilot cannot read the instruments. If the frequency is near one cycle per four seconds, airsickness may result. All of these effects, together with experience and ability, affect the pilot’s response to the turbulence.

Turbulence Metrics

By far the most important property about turbulence is intensity. The most commonly used turbulence criteria are shown in table 1. This turbulence scale has been

used for many years and is the basis for most pilot weather reports. The criteria are highly subjective and are dependent on aircraft type, airspeed, and pilot experience. Quantitative indications of turbulence can be determined from the on-board measurements of g-load (force that arises due to gravity), airspeed fluctuations, and rate-of-climb (see table 2). Normal gravity corresponds to a g-load of 1.0g. A change in g-load above or below the normal value is a rough measure of the intensity of the turbulence. Airspeed fluctuations refer to the largest positive and negative airspeed deviations from the average during a turbulent event. For example, if your average airspeed is 140 knots with variations between 130 and 150 knots, you are experiencing fluctuations of +/- 10 knots. Rate of climb simply refers to the largest positive or negative values during a turbulent event. This can only be used as a rough estimate of the vertical gust speed because it includes both the effect of the vertical gust and the motion of the aircraft.

Table 1. *Turbulence Reporting Criteria Table* (Source: Lester, 1995)

Intensity	Aircraft Reaction	Reaction Inside Aircraft
Light	Turbulence that momentarily causes slight, erratic changes in altitude and/or attitude (pitch, roll, yaw).	Occupants may feel a slight strain against seat belts. Unsecured objects may be displaced slightly. Food service may be conducted and no difficulty is encountered when walking.
Moderate	Turbulence that is similar to light turbulence but of greater intensity. Changes in altitude and/or attitude occur but the aircraft remains in positive control at all times. It usually causes variations in indicated airspeed.	Occupants feel definite strains against seat belts. Unsecured objects are dislodged. Food service and walking are difficult.
Severe	Turbulence that causes large abrupt changes in altitude and/or attitude. It usually causes large variations in indicated airspeed. Aircraft may be momentarily out of control.	Occupants are forced violently against seat belts. Unsecured objects are tossed about. Food service and walking are impossible.
Extreme	Turbulence in which the aircraft is violently tossed about and is practically impossible to control. It may cause structural damage.	

Table 2. *Quantitative measures of turbulence intensity. Values may be positive or negative.* (Source: Lester, 1995).

Turbulence	Airspeed Fluctuation (kts)	G-Load (g)	Derived Gust (fpm)
Light	5-14.9	0.20-0.49	300-1199
Moderate	15-24.9	0.50-0.99	1200-2099
Severe	>25	1.0-1.99	2100-2999
Extreme		>2.00	>3000

Turbulence Causes and Types

Aviation turbulence can be divided into four categories, depending on where the turbulence occurs, what large-scale circulations are present, and what is producing the turbulence. They are:

1. Low-level turbulence (LLT)
2. Turbulence in and near thunderstorms (TNT)
3. Clear-air turbulence (CAT)
4. Mountain wave turbulence (MWT)

To better understand turbulence and its effect in flight a description of each of the above is provided below. However, the main emphasis of this study concentrates on an aircraft shooting an instrument approach at low altitude with varying levels of crosswind and turbulence. Therefore low-level turbulence is the factor that will have the greatest impact in this study and is described in greater detail.

Low-Level Turbulence (LLT)

Low-level turbulence is defined simply as turbulence below 15,000 feet MSL. It can also be defined as that turbulence which occurs primarily within the atmospheric boundary layer (lowest few thousand feet of the atmosphere, that is, where surface heating and friction influences are significant). LLT includes mechanical turbulence, thermal turbulence, and turbulence in fronts. Although wake turbulence may be encountered at any altitude, it is particularly hazardous near the ground, so it is also considered with LLT. However, this study is not interested in turbulence created by other aircraft, so wake turbulence will be disregarded.

Mechanical Turbulence

Over flat ground and with strong winds, surface friction will slow the wind in the lowest layers causing the air above to turn over in turbulent eddies. The turbulent eddies cause fluctuations (gusts) in winds and vertical velocities. The turbulent eddies are then swept along by the sustained wind and cause wind shear near the ground. As the wind becomes stronger the mechanical turbulence extends to greater heights. The presence of obstructions such as buildings and trees increase the effect of surface roughness and strengthen LLT. During strong wind conditions, a trail of turbulent eddies is produced downwind of an obstacle, so hangars and large building near airports can cause control problems during takeoff and landing. Hills can produce very strong turbulent wakes with strong winds. Turbulent eddies downwind of hills are larger because the obstructions that cause them are larger. Steep hillsides encourage the flow to separate from the surface, producing eddies, LLT, and sheared regions. Even though mechanical turbulence caused by high terrain is extremely important to understand, this study will be conducted under a simulated flat ground where this will not be a factor. Similar effects are produced near canyons and valleys and these will not be taken into consideration for this study.

Thermal Turbulence

Thermal turbulence is LLT produced by convection in the boundary layer. It is usually a daytime phenomenon that occurs over land over fair weather conditions. Solar radiation heats the ground generating convection at the bottom of the boundary layer. During the afternoon the convection intensifies and gradually dies out as the earth's surface cools. When cool air moves over land or water, thermal turbulence can occur during any time of day or night. This convection will create thermals, which are simply

warm rising “bubbles” of air. As they move away from the ground they gain speed, grow in size, and become more organized. Glider pilots have taken advantage of the upward motions in thermals to gain altitude and fly long cross-country distances. However, thermal sources of lift for glider pilots are often sources of LLT for powered aircraft. Thermals create upward gusts that can range from 200 to 400 foot per minute (f.p.m). Flight through the boundary layer at midday in the summer will expose you to frequent LLT due to thermals.

Turbulence in Fronts

Fronts are not only sources of wind shear but they can also produce moderate or greater turbulence. Mesoscale fronts such as sea breezes and the thunderstorm gust front will create LLT by creating a source of rising air that will disrupt the motion of the aircraft. Macroscale frontal zones found in the middle and upper troposphere are also sources of turbulence. These are in connection with jet streams and clear-air turbulence, none of which will be considered in this study.

Turbulence in and near Thunderstorms (TNT)

Turbulence within the thunderstorm is caused by strong updrafts and downdrafts. The most frequent and the most intense TNT is found within the cloud (although turbulence below the cloud can have more disastrous consequences, with powerful downdrafts, downbursts, and microbursts). Furthermore it is made worse because it occurs in instrument meteorological conditions with heavy rain, lightning, and possible hail and icing. The combination of these hazards increases the chances for disorientation and loss of control. The weather phenomenon around thunderstorms is complex and beyond the scope of this study, however it is important to mention them as a severe to

extreme source of turbulence. The turbulence simulated in this study will be LLT but not that associated with a thunderstorm in order to eliminate all the other variables (icing, rain, etc.) that would be present with and around the thunderstorm.

Clear Air Turbulence (CAT)

Clear air turbulence is that turbulence that occurs in the free atmosphere away from any convective activity. It occurs in sudden bursts at high altitudes and is the result of high-level frontal passages and the jet-stream (a band of high speed wind). More specifically, CAT is found near high level stable layers that have vertical wind shear. When the air parcel in the stable layer is displaced vertically, atmospheric gravity waves develop. If the vertical wind shear is strong this can create wave crests to overrun the wave troughs, creating a very unstable situation. The reason we are concerned about CAT is that severe and extreme incidents have occurred, causing injuries and occasionally damage to the aircraft. CAT occurs more frequently within a few thousand feet of the tropopause, over mountains than elsewhere, and in winter than in summer.

Mountain Wave Turbulence (MWT)

Mountain wave turbulence is turbulence produced in connection with mountain lee waves. It is responsible for some of the most violent turbulence that is encountered away from thunderstorms. It occurs mainly in two well-defined regions of the lee wave system: near the tropopause and near the ground in the lower turbulent zone (the lower downwind side of the mountain). The intensity of MWT depends on the wind speed near the mountain peaks. The details as to how MWT is formed, how to avoid it, and its hazards are not relevant to the scope of this study and will not be described.

The reader can find more weather information related to crosswinds and turbulence in the following sources: *Aviation weather* (Lester, 1995), *Aviation weather services* (Gleim, 2004), *Severe weather flying* (Newton, 1983).

The Present Study

The purpose of this study is to evaluate the effect of crosswind and turbulence in mental workload and pilot tracking performance. The objective is to estimate the impact of crosswind and turbulence, of varying degrees, on performance and workload. This information should help researchers design studies in which workload needs to be systematically altered. Tsang and Wilson (1997) argued that human operators have limited processing capabilities, and once that limit is reached we find a decrease in performance. This can be very dangerous when applied to certain aviation scenarios, such as tracking an instrument approach in marginal weather conditions, low to the ground, with heavy crosswinds and fatiguing turbulence.

Eggemeir (1980) and Chiles (1979) observed that workload is multidimensional and hard to define. They pointed out that we must tailor the definition to the research situation. Although numerous studies have used crosswinds and turbulence as factors to increase workload and affect tracking performance, the literature does not address if these factors are good variables to be used in order to manipulate mental workload and tracking performance. In addition, they do not provide insights as to the extent that these variables have an impact on mental workload and tracking performance. In other words, the differences in tracking performance could be the result of an increase in mental workload, the factors themselves, or error.

The three questions being examined in this study are as follows:

- 1- What effect does crosswind have on mental workload and tracking performance?
- 2- What effect does turbulence have on mental workload and tracking performance?
- 3- Do crosswind and turbulence interact to produce a larger impact on performance than either one alone?

Based on previous research, it is believed that as the level of crosswind and turbulence is increased, mental workload will increase and tracking performance will decrease. The study will use crosswind and turbulence as variables to evaluate their effect on mental workload and tracking performance.

Anticipated Outcomes

The presence of a crosswind makes an instrument approach more difficult because the aircraft will drift to one side of the localizer when the aircraft is pointed directly at the runway. This presents a hazard as the aircraft can drift off course into terrain or other aircraft. However, it is anticipated that participants exposed to a crosswind will be able to establish a wind correction angle (crab angle) and track the course without large deviations from the centerline. Establishing the proper wind correction angle requires a trial and error process called bracketing, in which the pilot tries several wind correction angles (from large to small) until he or she figures out the wind correction angle needed. Therefore, deviations from centerline are expected until the wind correction angle is found. Given the nature of the instrument approach task and the fact that data will not be collected until the aircraft reaches the outer marker, it is

likely that the correct crab angle will be employed by the time that data collection begins. If crosswind does have a significant impact on performance, such an impact should be clear by comparing the maximum (20 knots) to the minimum crosswind condition (0 knots). Participants are trained during their instrument rating to compensate for crosswinds. With practice this becomes second nature and as a result an individual can perform this task without exerting much mental effort. Therefore, the mental workload is expected to stay low when correcting for a crosswind.

It is anticipated that moderate and severe levels of turbulence will have produce large deviations from the centerline during the approaches. This is due to the fact that as an aircraft is affected by turbulence it gets displaced from its original position, making a perfect tracking task impossible. In addition to this, there is no correction the pilot can establish to anticipate for the effect of turbulence. Continuously scanning the attitude indicator and simultaneously attempting to track the course is expected to increase mental workload to high levels (Oman, Rasmussen, Robinson, & Huntley, 1995). This expected increase in mental workload is a concern as Smith (1979) noted problems in relation to communication, decision-making, planning, leadership, and stress in conditions with increased workload.

The most detrimental effect on tracking performance is expected when participants are exposed to both crosswind and turbulence as the pilot will have to divert attention between maintaining control of the airplane, establishing and maintaining a crab angle, and correcting for the aircraft being displaced off course in a continuous basis. We also expect to see an interaction between crosswind and turbulence on performance and mental workload.

The data will be analyzed using trend analysis computations and confidence intervals as opposed to null-hypothesis significance testing. The trend analysis results will provide insight into the functional relationships between the crosswind and turbulence levels and the performance and workload outcomes. Confidence intervals allow one to determine whether or not significant differences exist while simultaneously estimating the size of those differences.

METHOD

Participants

Fifteen full-time college students from Embry-Riddle Aeronautical University in Daytona Beach, Florida served as experimental participants. Students were told that they were volunteering to assist in an experiment using crosswind and turbulence to manipulate mental workload and measure flight tracking performance. All participants held at least an instrument rating and either a first, second, or third class airman's medical certificate. All participants indicated that they were instrument current (six instrument approaches within the last six months, holding procedures and intercepting and tracking radials and courses using navigational equipment) and had a total flight time between 150-300 hours. All of the participants were male and their average age was 21.2 years (SD= 1.7). A screening procedure was used to gather information regarding the participant's total flight time, simulator experience, and instrument currency. The reported average total flight time was 227.2 hours (SD= 64.3) and the average reported flight time in a C172 was 193.8 hours (SD= 57.5). All of the participants had experience flying a more advanced Frasca flight simulator as part of their flight training with an

average of 25.8 hours (SD= 8.2) being reported. The experiment took approximately one and a half hours per participant. Participants were paid \$10 for participating in the study.

Materials and Apparatus

An Elite iGATE Personal Computer Aviation Training Device (PCATD) was used for the experiment and the standard Cessna 172 flight model and instrument panel configuration was used. The experimenter station had dual monitors and an electronic data switch that allowed the researcher to view and control any of the computers without disturbing the experiment. It also allowed the researcher to set up the flight scenario as well as to monitor in real-time the participant's flight and scenario progress. An Elite iGATE flight control console was used to provide all of the physical flight controls necessary for the experiment (yoke, rudder pedals & power quadrant).

The Elite PCATD was set up to model a standard Cessna 172 (including a localizer and glide slope). All aircraft systems were preset for the participant and wing flaps were preset at 10 degrees for the approach. During the experiment the researcher was able to view and control the Elite PCATD as well as monitor the flight instruments. The flight parameter data was recorded using an add-on Elite software module.

The standard out-the-window view imbedded into the instrument display was used during the practice sessions in order to facilitate accommodation to the simulator. The display of out-the-window information was disabled during the data collection trials.

Other Equipment Used

A FAA U.S. Terminal Procedures Chart (approach plate) published by the FAA National Aeronautical Charting Office for the ILS runway 7L at Daytona Beach International Airport (KDAB) was provided to the participants to conduct the approaches.

A computerized version of the NASA TLX was used to collect subjective workload data from the participants. Directions on how to use the NASA TLX was explained during the pre-briefing and was administered after each approach condition.

Design

The study utilized a completely crossed repeated measures design with two independent variables: crosswind (0, 10, and 20 kts) and turbulence (no turbulence, light, moderate, and severe). The dependent variables were flight performance and workload. This resulted in each participant completing approaches under twelve unique conditions. Treatment presentation order was counterbalanced using a computer-generated random order sequence of conditions. Crosswinds were presented at a 90 degree right angle to the approach path. Turbulence and crosswind were set to the desired setting using the Elite software before every trial run. Elite offered turbulence settings that ranged from 0 to 12. These represented the turbulence intensities, 0 being no turbulence and 12 being extreme turbulence. In order to map Elite's turbulence setting (scale) to the present study, an evaluation was conducted. The aircraft was set to fly in a trimmed condition, straight and level at 100 kts. The turbulence level was then increased to a setting of 3 and several data parameters were collected for 45 seconds. Similar data were collected for the 6, 9, and 12 settings. Airspeed and g-load data were examined and the minimum and maximum points were compared with the FAA guidelines (see Table 2). The data showed that the Elite turbulence levels of 0, 3, 6, and 9 would generally represent the settings of no turbulence, light, moderate, and severe used by the present study.

The FAA ATP practical test standards (PTS) were applied ($\frac{1}{4}$ scale deflection, +/- 5 kts.) on both localizer and glide slope from the outer marker to decision height. The

ATP practical test standards were selected instead of the Instrument practical test standards in order to compensate for any ceiling effects associated with the easier and broader Instrument practical test standard (3/4 scale deflection, +/- 10 kts.). Participant scores could range between 0 and 1, indicating the proportion of the approach within standard on all three parameters, simultaneously.

Procedure

Upon arrival, the researcher welcomed the participant, provided a quick overview the study, briefly explain how this study fitted into other workload programs and went over the informed consent sheet. The participant then signed the informed consent sheet and was instructed to fill out a demographics form that included basic participant information as well as flight ratings held and flight/simulator experience. Once this was completed, the participant was given more detailed information about how the session was to be conducted. The participant was instructed on the controls of the PCATD and was instructed to make climbs, turns, and descents to get a feel for how the PCATD flew. The participant was then given five minutes of free flight practice. After the practice session, the participant was given an approach plate for the Daytona Beach International Airport (ILS 7L). He was told that when the simulation began, the aircraft would be located outside the outer marker at 1600 feet, on course to intercept the localizer. The participant was instructed to intercept the localizer, maintain altitude until intercepting the glide slope, and to maintain 100 knots for the entire approach. The participant was then given a chance to ask questions.

The twelve data collection trials then began. The out-the-window graphics were disabled and wind speed and turbulence were set to the required level. The data collection

was started at the outer marker (OM) and stopped at decision height. The PCATD was then reset, wind turbulence and speed were changed as required, and the next approach was started. This sequence was repeated until each participant flew an approach at each wind speed and turbulence level.

Data Collection

Flight parameter data were collected at 10 Hz and started when the aircraft crossed the OM inbound and ended when the aircraft arrived at the published decision height (DH) for the approach (which the participant was asked to verbally “call out”). The Elite data collection module collected data on a total of 65 different simulator parameters, of which, airspeed, localizer, and glide slope needle deflection data was used to compute TWS. The data collection system coded the needle deflection data in terms of full-scale deflection. In other words, a ½ scale deflection on one of the needles was represented with a value of .500. Each localizer, glide slope, and airspeed datum was compared to the ATP standards. If the datum was within the relevant standard, a corresponding variable was assigned a value of 1 for that time sample. TWS was computed by averaging each binary variable and this average represented the proportion of time spent within that standard. This process occurred for each flight parameter (i.e. airspeed, localizer, and glide slope) at each time stamp and if all three parameters were within standard, a total TWS score of 1 was allocated for that time stamp.

After each instrument approach was accomplished, the participant completed the NASA TLX workload survey. Once all twelve approaches were finished, the researcher debriefed the participant by asking if they have any questions about anything they did. The participants were provided with a contact sheet that had the email address and phone

numbers of the principal investigator for the project. They were encouraged to contact the researcher should they have any questions or concerns after the study. Participants were then paid, thanked and dismissed. In summary, each participant flew twelve approaches and experienced a different crosswind and turbulence combination during each approach. Counterbalancing was accomplished via randomization of presentation order. Flight performance and workload was measured using TWS and TLX scores, respectively.

RESULTS

Fifteen participants completed the study and flew 12 separate instrument approaches. Each approach presented a unique level of crosswind and turbulence. A number of participants did not perform well during the trials resulting in several ATP-TWS scores with a value of zero. Overall, performance as measured by ATP-TWS was fairly low, especially when compared with the performance levels recorded by Hall, Doherty, and Mion (2004) and Accettullo (2003). Thus, the ATP-TWS metric was abandoned because analysis on the data would have been suspect given the observed floor effect. As a result, the RMSE for aircraft position data were used to evaluate performance instead of TWS scores. RMSE values were transformed using the natural log (ln) function in order to normalize the RMSE distributions so that the mean and standard deviation could be used for descriptive purposes and to satisfy the assumption of normality for ANOVA analysis.

The results of the data analysis are presented in separate sub-sections for the performance and workload data. All analyses, unless otherwise noted, were performed using an α level of .05 and confidence intervals of 95%.

Results for the Performance Data

Descriptive statistics for the performance data are presented in Table 3. A (4x3) repeated measures ANOVA was conducted on the performance data. A number of assumptions underlie the use of a repeated measures ANOVA including homogeneity of variance, normality of data, and sphericity. Of these, violations of sphericity are the most serious with regard to the accuracy of reported p values (Keppel, 1991). Analysis of Mauchly's test of sphericity for the performance data (Table 4) shows a violation of sphericity for the turbulence and interaction factors (using $\alpha = .25$). The implication of non-sphericity in the data is that Type I error rates will inflate unless controlled using epsilon correction coefficients. Keppel (1991) suggested that the Greenhouse-Geisser epsilon correction coefficient tends to overcorrect for non-sphericity; thus, the Huynh-Feldt correction factors will be used during the ANOVA process.

Table 3. *Descriptive statistics for the performance data.*

Turbulence	Crosswind ($n = 15$)					
	<u>0</u>		<u>10</u>		<u>20</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
No turbulence	4.79	.61	4.88	.54	5.03	.64
Light	5.18	.74	5.33	.57	5.31	.73
Moderate	5.76	.79	5.45	.53	5.62	.75
Severe	5.60	.60	5.76	1.03	5.77	.74

Table 4. *Mauchly's test of sphericity for the performance data.*

Factor	Mauchly's <i>W</i>	Approx Chi-square	df	<i>p</i>	Epsilon		
					Greenhouse- Geisser	Huynh- Feldt	Lower- bound
Turbulence	.42	10.95	5	.05	.65	.75	.33
Crosswind	.83	2.27	2	.32	.86	.97	.50
Turbulence*Crosswind	.11	25.95	20	.18	.53	.71	.16

The results of the ANOVA analysis indicate that the turbulence main effect was the only statistically significant effect, $F(2.26, 31.69) = 14.28, p < .001$ (see Table 5 for complete results). Even without epsilon correction, the remaining factors were not statistically significant. The statistically significant results for the Turbulence factor were further examined via a trend analysis on the data. Trend analysis was performed in order to establish the form of the relationship between turbulence level settings and pilot performance. As described by Keppel (1991), any factor (given enough manipulated levels) may produce a linear, quadratic, and/or cubic trend in the output variable of interest. These trends are separate from one another and can simultaneously be present in the data. In this case, a positive linear trend is present (see Table 6) in the data, indicating that as turbulence settings in the simulator are increased, performance tends to decrease (ln rmse values tend to increase) in a linear fashion (see Figure 2).

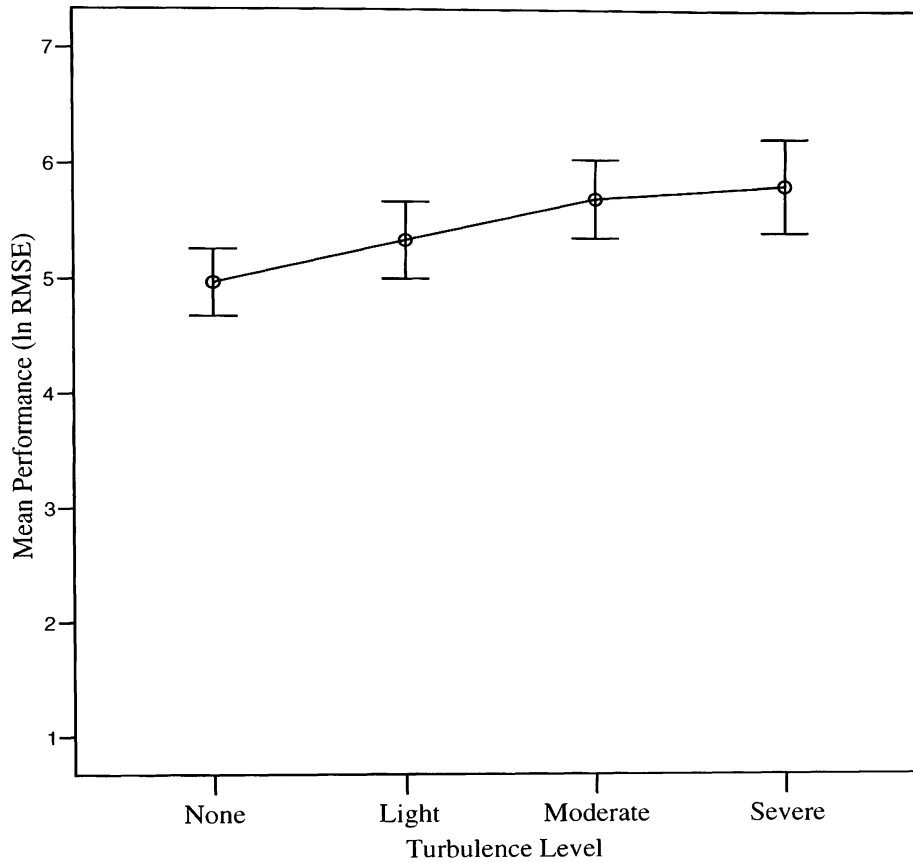
Table 5. ANOVA source table for the performance data.

Source	SS	df	MS	F	<i>p</i>	Eta squared	Power
Turbulence	18.13	2.26	8.00	14.28	<.01	.50	.99
Error (turbulence)	17.77	31.69	(.56)				
Crosswind	.33	1.94	.17	.50	.60	.03	.12
Error (crosswind)	9.34	27.19	(.34)				
Turbulence*crosswind	1.29	4.25	.30	.99	.42	.06	.30
Error (turbulence*crosswind)	18.32	59.62	(.30)				

Note. All reported *p* values are based on the Huynh-Feldt epsilon correction. Values enclosed in parentheses represent mean square errors.

Table 6. Test of within subjects contrasts for the performance data.

Source	Turbulence	SS	df	MS	F	<i>P</i>	Eta squared	Power
Turbulence	Linear	17.19	1	17.19	20.78	<.01	.59	.98
	Quadratic	.84	1	.84	4.315	.057	.236	.490
	Cubic	.10	1	.10	.395	.540	.027	.090



NOTES: Error bars are based on the mean ln RMSE position data +/- 2.145 standard error of the mean units.

Figure 2. Average performance across turbulence level settings.

Results for the Workload Data

After completing each instrument approach, participants completed a computerized version of the NASA TLX survey. Descriptive statistics for the TLX data are presented in Table 7. A (4x3) repeated measures ANOVA was conducted on the data. Analysis of Mauchly's test of sphericity for the workload data (Table 8) shows a violation of sphericity for the turbulence and crosswind factors (using $\alpha = .25$). As with the performance data, Huynh-Feldt correction factors were used during the ANOVA process to control Type I error inflation.

Table 7. *Descriptive statistics for the workload data.*

Turbulence	Crosswind (n = 15)					
	0		10		20	
	M	SD	M	SD	M	SD
No turbulence	29.53	15.08	26.80	15.46	33.53	17.28
Light	38.26	19.13	40.13	16.24	42.86	14.17
Moderate	65.86	17.19	64.53	14.40	61.26	16.26
Severe	63.46	16.05	63.53	17.26	68.86	13.68

Table 8. *Mauchly's test of sphericity for the workload data.*

Factor	Mauchly's W	Approx Chi-square	df	p	Epsilon		
					Greenhouse- Geisser	Huynh- Feldt	Lower- bound
Turbulence	.47	9.40	5	.09	.67	.78	.33
Crosswind	.79	2.91	2	.23	.83	.93	.50
Turbulence*Crosswind	.23	17.12	20	.66	.68	.99	.16

The results of the ANOVA analysis indicate that the turbulence main effect was the only statistically significant effect, $F(2.35, 33) = 48.69, p < .001$ (see Table 9 for complete results). Even without epsilon correction, the remaining factors were not statistically significant. The statistically significant results for the Turbulence factor were further examined via a trend analysis on the data. As with the performance data, trend analysis was performed in order to establish the form of the relationship between turbulence level settings and workload. In this case, linear, quadratic, and cubic trends are simultaneously present (see Table 10) in the data (see Figure 3).

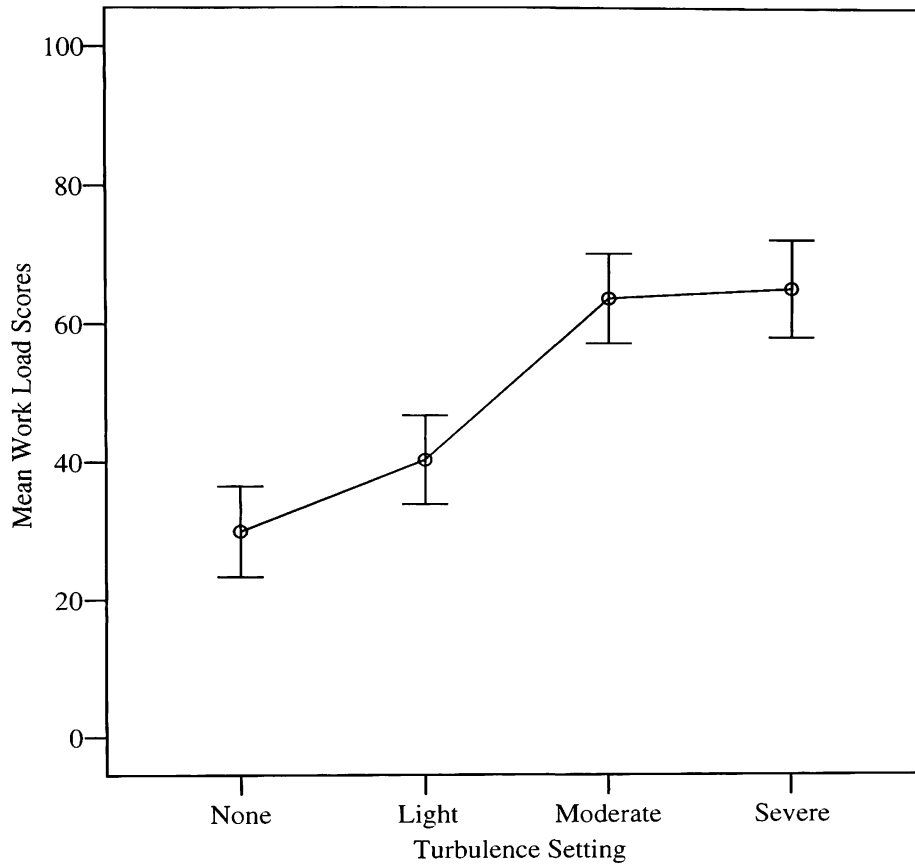
Table 9. ANOVA source table for the workload data.

Source	SS	df	MS	F	<i>p</i>	Eta squared	Power
Turbulence	41405.20	2.35	17565.70	48.69	<.001	.77	1.00
Error (turbulence)	11905.13	33.00	(360.75)				
Crosswind	282.41	1.86	151.60	.56	.56	.03	.13
Error (crosswind)	7050.25	26.08	(270.33)				
Turbulence*crosswind	678.30	5.98	113.42	.77	.59	.05	.29
Error(turbulence*crosswind)	12270.36	83.72	(146.56)				

Note. All reported *p* values are based on the Huynh-Feldt epsilon correction. Values enclosed in parentheses represent mean square errors.

Table 10. Test of within subjects contrasts for the workload data.

Source	Turbulence	SS	df	MS	F	<i>p</i>	Eta squared	Power
Turbulence	Linear	37713.64	1	37713.64	67.86	.00	.82	1.00
	Quadratic	924.80	1	924.80	6.55	.02	.31	.66
	Cubic	2766.76	1	2766.76	18.01	.00	.56	.97



NOTES: Error bars are based on the mean workload score data +/- 2.145 standard error of the mean units.
 Figure 3. Average workload scores for turbulence.

Regression Equations

The data for each of the trend analyses and for each trend component were used to generate lines of best fit. These equations represent the trends mathematically such that the sum of the squared residuals between the line and the four turbulence level means are minimized. All of the equations are presented in Table 11. These equations can be used to roughly estimate the expected performance or workload value given a specific value for the Elite turbulence setting. These equations are not equivalent to regression equations, which allow for the computation of a confidence interval around each estimate.

Table 11. Regression equations for each trend component.

Variable	Trend Component	Equation
Performance	Linear	$\hat{Y} = .10T + 5.02$
	Linear	$\hat{Y} = 4.32 T + 30.47$
Work Load	Quadratic	$\hat{Y} = 6.58T - .25T^2 + 28.20$
	Cubic	$\hat{Y} = -2.57T - 2.67T^2 - .223T^3 + 29.96$

DISCUSSION

Three issues were addressed in this study. The first was to evaluate the effect of crosswind on mental workload and pilot tracking performance. Tracking performance was measured by computing RMSE values which is an indication of deviations of the actual flight path to the ideal approach path. The RMSE data suggest that the impact of crosswind on tracking performance is small and probably not of practical concern. This is not to say that performance measures across the crosswind conditions were statistically equivalent, only that there is no conclusive evidence of a relationship between crosswind and performance. It is possible that an impact on performance could be found if higher crosswind components were simulated.

Similarly, the results did not find that crosswind statistically increased mental workload. These results are consistent with previous research conducted by Hall et al. (2003). The lack of statistical differences in workload across crosswind conditions may be partly attributable to less than perfect reliability of the TLX scale or other sources of error variance that may be obfuscating the true impact of crosswind on workload. As with the tracking performance results, it may also be the case that crosswind does not

practically increase perceived workload, at least at the levels of crosswind that were simulated. Another practical concern is that on the manner in which crosswind is simulated. Simulator programs tend to provide a rather static source of “crosswind”, meaning the aircraft track is shifted relative to the heading of the aircraft. The amount of shift is a function of wind speed and direction relative to the heading of the aircraft, but it tends to be consistent as opposed to variable. Real-world crosswinds tend to vary in relatively short cycles; simulated crosswinds tend to be steady.

The second issue involves the effect of turbulence on mental workload and pilot tracking performance. The study results show that as the turbulence level was increased, observed tracking performance decreased (e.g. RMSE values increased; see Table 3). To some extent, part of the decrease in tracking performance is an artifact of how simulators model turbulence. That is, the simulator is displacing the aircraft’s position by some amount and at some frequency that is commensurate with the turbulence setting. This constant shifting makes it impossible for the aircraft to stay perfectly on course and produce an RMSE score of zero. Part of the decrease in performance is also likely due to the fact that it is more difficult to keep the aircraft on a prescribed pathway when turbulence is present. The fact that performance scores varied across pilots demonstrates this point perfectly; if the decrease in performance were due solely to measurement artifact, there would be no variance in performance across pilots. The fact that turbulence was random and could not be predicted (as in real life) added to the difficulty of the approach. An extremely difficult task can become less challenging when one is able to predict what’s coming next, anticipate its effect, and set a corrective action to prevent unwanted results (i.e. as with crosswind). With the randomness and unpredictability of

turbulence, this was not possible. Once the aircraft was off course and the pilot recognized it, he could then set a correction to get back on course.

Another challenge that was presented to the participants while flying in turbulent conditions was the fact that in many occasions the aircraft yawed and rolled in opposite directions. This required great amount of effort and concentration by the pilot to maintain that aircraft coordinated and in its desired flight path. The scanning of the instruments needed to be changed as a result of the turn coordinator becoming unreliable, showing erratic left and right bank attitudes in moderate and severe levels of turbulence. Most participants placed more emphasis on the attitude indicator in an attempt to maintain the wings level and prevent the aircraft from rolling into a steep bank and an unusual attitude developing into an unsafe situation. This “non-standard” way of scanning the instruments left the participants with less time to concentrate on the tracking task and as a result achieved higher RMSE values. There was also a noticeable increase in the frustration and overall fatigue as the participants attempted to constantly “fight” the turbulence while flying the instrument approach. This observation is supported by an increase in the NASA-TLX workload scores.

The workload scores reflected the difficulty imposed by the turbulence, with scores increasing as turbulence settings increased and markedly increasing between the turbulence settings of 3 and 6. The “step” function exhibited in the workload scores suggests that pilots are well equipped to handle turbulence up to some point without experiencing a large increase in workload, but beyond some threshold, most pilots suddenly found themselves overburdened by the turbulence settings. It is interesting to note that this same pattern was not seen in the performance data, where a precipitous fall

in performance is associated with a small range of turbulence settings. In any event, the sudden change in experienced workload might be explained by non-linearity in the turbulence settings. As discussed earlier in the paper, efforts were made to map the Elite's turbulence settings with accepted rating scales for turbulence. These efforts were somewhat successful showing that g-load measures did increase with increases in the turbulence settings. It may be the case, though, that these increases in turbulence are not linear relative to the associated changes in the psychological experience of workload. For example, demarcations in the turbulence scale may not correspond in a one-to-one fashion with the amount of experienced workload. Thus, moving the turbulence setting from 3 to 4 may produce a five-unit change in workload scores (for example), but moving from a 6 to 7 on the turbulence scale may produce a ten-unit change in workload scores.

The third issue was to evaluate if crosswind and turbulence interacted to produce a larger effect than either one alone. The results of the study failed to find a statistically significant interaction between these two factors for either the performance or workload data. This is probably due the non-impact of the crosswind factor under any circumstances. The presence of a crosswind is countered by using a constant crab angle during the approach, which essentially renders the crosswind irrelevant during the remainder of the approach.

The use of simulators in aviation research extends back many years. The need for tight controls on environmental conditions, safety concerns, and the cost associated with performing research with real aircraft have made simulators a standard fixture in aviation research. To be sure, the generalizability of aviation research findings using simulators is

a concern, but the volume of research in the area of pilot performance would be drastically reduced if all research had to be conducted using real aircraft.

Much of the pilot performance research conducted to date has been performed in simulated environments. The results of this study should help future researchers attempting to manipulate workload in a simulated environment to do so effectively. Specifically, the results of this study provide insight on how to best manipulate workload and performance (e.g. via manipulation of turbulence settings), but also on the degree to which specific environmental manipulations will impact performance and workload.

The fact that only one out of the fifteen pilots was able to perform on two trials to the ATP standard was not surprising. The FAA ATP practical test standards are very strict and require lots of practice, dedication, and experience to be achieved. The low level of performance might also be explained by a lack of experience flying the particular simulator setup. Some of the participants expressed that under the moderate and severe levels of turbulence and as a result of the over-controlling required to maintain the aircraft flying the required course, the yoke would hit their knees, creating a distraction during the instrument approach procedure.

Study Limitations and Recommendations

As with any study, there are several limitations inherent in this study that tempers the results. The generalizability of the results to various groups of pilots is restricted by the fact that the participants in the study were fairly homogenous in terms of age and were all trained to perform instrument approaches using the same methodology. Therefore, estimates of performance at the population level are technically limited to other flight students in the program. Replicating the study with pilots from different

populations (i.e. commercial pilots, more experienced pilots, etc.) would address this issue.

Another limiting factor is that a constant crosswind component was used in this study. A consistent crosswind is relatively easy to compensate for, especially when the goal is to keep the aircraft on a specific ground track. Also the crosswind was always presented from the same direction. Some participants appeared to be aware of this fact and automatically corrected when they detected that the crosswind speed was significant. Part of the challenge of dealing with a crosswind is to estimate its direction. This was not accurately simulated during this study and should be addressed in future crosswind studies.

Some of the examples of poor performance in this study may have been due to a lack of instrument flight proficiency. The participants varied markedly in terms of their instrument flight skills, even though the participants were required to be instrument current. The FAA defines instrument currency as performing 6 instrument approaches within the last 6 months, but this standard does not address the concept of proficiency per se. A pilot's instrument scan, situational awareness, and comfort have a lot to do with how proficient (not necessarily current) he is at flying only with reference to the flight instruments. Using a group of participants who has flown a set number of instrument approaches within a smaller time period might reduce the variability in their performance when performing a simulated instrument-tracking task.

The primary goal of this study was to evaluate the extent to which crosswind and turbulence would influence pilot performance and workload during a simulated instrument-tracking task. Turbulence affected both mental workload and tracking

performance while crosswind did not affect either. Both crosswind and turbulence combined were expected to have a greater impact on the pilot's ability to track the final approach course inbound than either one alone, but this result was not found. Aviation researchers should perhaps reconsider the use of crosswind in performance and workload studies as a way to increase the difficulty of approach tasks as such manipulations are not likely to impact performance or workload. Instead, turbulence manipulations should be considered and are very likely to increase the difficulty of the task.

REFERENCES

- Bailey, L. & Thompson, R. (2001). The TLX: One or more constructs. Paper presented at the Eleventh International Symposium for Aviation Psychology, Columbus, OH: The Ohio State University.
- Boeing Commercial Aircraft Group (1996). Statistical Summary of Commercial Jet Accidents: Worldwide Operations- 1959-1996. Seattle, Washington: Boeing Commercial Airplane Group.
- Charlton, S.G. (1991). Aeromedical Human Factors OT&E Handbook (AFOTEC Tech. Paper 4.1). Kirtland AFB, NM: Air Force Operational Test and Evaluation Center.
- Charlton, S.G. (1996). *Mental Workload Test and Evaluation*. In T.G. O'Brien & S.G. Charlton (Eds.), *Handbook of Human Factors Testing and Evaluation* (pp. 181-199). Mahwah, NJ: Erlbaum
- Croft, John. *Researchers Perfect New Ways To Monitor Pilot Performance*. Aviation Week & Space Technology. New York: July 16, 2001. Vol 155, Iss. 3, Pg. 76.
- Denning, T.V. (2003). *An Experimental Investigation of the Differences in the Subjective Pilot Workload Across Simulated and Real Flight Conditions*. (Master's thesis, Embry-Riddle Aeronautical University, 2003).
- Endsley, M.R. (1996). *Situation Awareness Measurement in Test and Evaluation*. In T.G. O'Brien & S.G. Charlton (Eds.), *Handbook of Human Factors Testing and Evaluation* (pp. 159-180). Mahwah, NJ: Erlbaum.
- Galster, S.M., Duley, J.A., Masalonis, A.J., & Parasuraman, R. (2001). Air traffic controller performance and workload under mature free flight: conflict detection and

resolution of aircraft self-separation. *The International Journal of Aviation Psychology*, 11(1), 71-93.

Gerathwohl, S.J. (1976). Definition and measurement of perceptual and mental workload in aircrews and operators of air force weapon systems. In B. Hartman (Ed.), *AGARD Conference Proceedings: No.181. Higher mental functioning in operational environments* (pp C1-1- C1-7). Texas: Brooks Air Force Base.

Hahn, K.U., Heintsch, T., Kaufmann, B., Schanzer, G., & Swolinsky, M. (1990). *Effect of Wind and Wind Variation on Aircraft Flight Paths*. France: Technical University of Braunschweig. (NTIS AGARD-AG-301 Vol.1).

Hall, S.M., Landa, L., Hart, C., & Karkman, S. (2003). Single pilot performance study report (Tech. Rep. No. ERAU-HF-03-01/SESATS-03-01). Daytona Beach, Florida. Embry-Riddle Aeronautical University, Human Factors Department.

Hancock, P.A. (1996). Effects of control order, augmented feedback, input device and practice on tracking performance and perceived workload. *Ergonomics*, 39(9), 1146-1162.

Hancock, P.A., Williams, G., Manning, C.M., & Miyake, S. (1995). Influence of task demand characteristics on workload and performance. *The International Journal of Aviation Psychology*, 5(1), 63-86.

Harris, W.C., Hancock, P.A., Arthur, E.J., & Caird, J.K. (1995). Performance, workload, and fatigue changes associated with automation. *The International Journal of Aviation Psychology*, 5(2), 169-185.

- Hart, S.G. (1986). Theory and measurement of human workload. In J. Zeider (Ed.), Human productivity enhancement, Vol 1: Training and human factors in system design (pp. 396-456). New York: Praeger.
- Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, Peter A., and Meshkati, Najmedin (Eds.). Human Mental Workload. Amsterdam: North Holland.
- Jahns, D.W. Operator workload: What is it and how should it be measured? Paper presented at the Interagency Conference on Management and Technology in the Crew System Designed Process. Los Angeles, California, September 12-14, 1972.
- Jahns, D.W. (1973). Operator workload: What is it and how should it be measured? In K.D. Gross & J.J. McGrath (Eds.), Crew system design (pp. 281-288). Santa Barbara, CA: Anacapa Sciences Inc.
- Kantowitz, B.H., & Casper, P.A. (1988). Human workload in aviation. In E. Wiener & D. Nagel (Eds.), Human factors in aviation (pp. 157-185). New York: Academic Press.
- Katz, J.A., *Pilot workload in the Air Transport Environment: Measurement, Theory, and the Influence of Air Traffic Control*. Flight Transportation Laboratory, MIT, Cambridge, Mass., May 1980. (FTL Report R80-3)
- Leggatt, A.P., & Noyes, J.M. (2000). Navigation aids: Effects on crew workload and performance. *The International Journal of Aviation Psychology*, 12(2), 89-104.
- Lester, P.F. (1995). *Aviation Weather*. Englewood, CO: Jeppesen Sanderson Training Products.
- Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C., & Wherry, R.J. (1989). *Operator Workload:*

- Comprehensive Review and Evaluation of Operator Workload Methodologies (Tech. Rep. 851). Alexandria, VA: U.S. Army Research Institute.
- Mjos, K. (2001). Communication and operational failures in the cockpit. *Human Factors and Aerospace Safety*: Vol 1(4): 323-340.
- National Aeronautics & Space Administration. *Pilot Performance Affected by Sleep Debt, Time Awake, Time of Day. Air Safety Week*. New York: Jan 29, 2001. Vol. 15, Iss. 5, pg. 1.
- O'Donnell, R.D., & Eggemeier, F.T. (1986). Workload assessment methodology. In Boff, K.R., Kaufman, L., & Thomas, J. (Eds.). *Handbook of perception and human performance, Volume II: Cognitive processes and performance*. New York: John Wiley.
- Oman, C.M., Rasmussen, S.A., Robinson, S.K., Huntley, M.S.Jr., et. al. (1995). *The Use of Analog Track Angle Error Display for Improving GPS Approach Performance*. Cambridge, MA: Institute of Technology. Department of Transportation.
- Oman, A.J., Kendra, A.J., Hayashi, M., Stearns, M.J., & CoheBurki-Cohen, J. (2001). Vertical navigation displays: pilot performance and workload during simulated constant angle of descent GPS approaches. *International Journal of Aviation Psychology*: Vol 11(1): 15-31.
- Parasuraman, R., & Mouloua, M. (Eds.). (1996). *Automation and Human Performance: Theory and Application*. New Jersey: Lawrence Erlbaum Associates.
- Ragsdale, W.A., Osborne, A.R., & Seaman, C.K. (1974). Atmospheric effects on the space shuttle approach and landing. *Conference on Aerospace and Aeronautical Meteorology*, Texas, 21-26.

- Rantanen, E.M. & Talleur, D.A. (2001). Measurement of Pilot Performance During Instrument Flight Using Flight Data Recorders. *International Journal of Aviation Research and Development*, 1(2): 89-102.
- Razmjou, S. (1996). Mental workload in heat: Toward a framework for analyses of stress states. *Aviation, Space, and Environmental Medicine*, 67(6), 530-538.
- Roscoe, A.H., Introduction. IN AGARD Monograph. *Assessing Pilot Workload*. Hartford House, London, 1978.
- Rosenberg, B., Rehmann, J., and Stein, E.S., *The Relationship Between Effort Rating and Performance in a Critical Tracking Task*. FAA Technical Center Technical Report (DOT/FAA/EM 81/13), Atlantic City, N.J., October 1982.
- Rosenberg, B. and Stein, E.S., *The Measurement of Pilot Workload*. FAA Technical Center Technical Report (DOT/FAA/CT-82/23), Atlantic City, N.J., January 1983.
- Rueb, J.D., Vidulich, M.A., & Hassoun, J.A. (1994). Use of workload redlines: A KC-135 crew reduction application. *The International Journal of Aviation Psychology*, 4(1), 47-64.
- Shelnutt, J.B., Childs, J.M., Prophet, W.W., & Spears, W.D. (1980). *Human Factors Problems in General Aviation (FAA-CT-80-194)*. Pensacola, FL: Seville Research Corporation.
- Sheridan, T.B., & Simpson, R.W. (1979). *Toward the definition and measurement of the mental workload of transport pilots*. (FTL Report R79-4). Cambridge, MA: Institute of Technology. Department of Aeronautics and Astronautics.
- Sturdy, J.L., & Hansman, J. Jr. (1988). *Analysis of the Altitude Tracking Performance of Aircraft-Autopilot Systems in the Presence of Atmospheric Disturbances* (FTL Report

R88-1). Cambridge, MA: Institute of Technology. Department of Aeronautics and Astronautics.

Svensson, E.A., & Wilson, G.F. (2002). Psychological and psychophysiological models of pilot performance for systems development and mission evaluation. *International Journal of Aviation Psychology*: Vol 12(1): 95-110.

Tsang, P.S., & Wilson, G.F. (1997). *Mental Workload*. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics*. New York: Wiley.

Wierwille, W.W., & Eggemeier, F.T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors*, 35, 263-281.

Willits, P. (1998). *Instrument Commercial Manual*. Englewood, CO: Jeppesen Sanderson Training Products.

APPENDICES

APPENDIX A

Department of Human Factors and Systems
Embry-Riddle Aeronautical University

CONSENT FORM

Prospective Research Participants: *Read this consent form carefully; ask as many questions as you like before deciding whether or not you wish to participate in the research study. Feel free to ask questions at any time before, during, or after your participation in the research.*

I consent to participating in the research project entitled:
The Effect of Crosswind and Turbulence on Mental Workload and Pilot Tracking Performance. The principle investigator of the study is: Bruno Vivaldi

The purpose of this study is to investigate the effect of crosswind and turbulence on mental workload and pilot tracking performance during an instrument approach (ILS). You will begin by taking the first five minutes of the session to become familiar with the Elite flight simulator and its various controls. You will also practice some basic attitude instrument flying, including climbs, descents, turns, and straight and level.

After completing the familiarization session, you will be asked to fly a total of 12 instrument approaches (each lasting approximately 3 minutes), experiencing in each a different level of crosswind and turbulence. When the aircraft reaches the decision height, the Elite simulator will be paused and you will be asked to complete a form that will address your mental workload during the simulation. At the completion of this task the researcher will reset the simulator and the next trial run will be flown. This process will continue until all 12 trial runs are completed.

The results of this experiment may be used for other ongoing research, but your name will not be used in the reporting of the results. Only group data will be used, as all personal information will be kept completely confidential. If you wish to withdraw from the experiment, you may do so at any time without penalty.

I acknowledge that I have had the opportunity to obtain additional information regarding the study and that any questions I have raised have been answered to my full satisfaction.

Finally, I acknowledge that I have read and fully understand the consent form.

Date: _____ Name (please print): _____

Signed (participant): _____ Researcher: _____

APPENDIX B

NASA-TLX Rating Scale Definitions

title	Endpoints	Descriptions
Mental Demand	Low/High	How much mental and perceptual activity was required (e.g. thinking, calculating, remembering, and searching)? Was the task easy or demanding, simple or complex?
Physical Demand	Low/High	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating)? Was the task easy or demanding, slow or brisk?
Temporal Demand	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Effort	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Performance	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Frustration Level	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content and relaxed did you feel during the task?

NASA TLX Rating Sheet

Instructions: On each scale, place a mark that represents the magnitude of that factor in the task(s) you just performed.

|-----|
LOW MENTAL DEMAND HIGH

|-----|
LOW PHYSICAL DEMAND HIGH

|-----|
LOW TEMPORAL DEMAND HIGH

|-----|
EXCELLENT PERFORMANCE POOR

|-----|
LOW EFFORT HIGH

|-----|
LOW FRUSTRATION HIGH

NASA-TLX

Pairwise Comparison of Factors

Instructions:

Circle the member of each pair that provided the most significant source of variation in the task(s) that you just performed.

PHYSICAL DEMAND / MENTAL DEMAND

TEMPORAL DEMAND / MENTAL DEMAND

PERFORMANCE / MENTAL DEMAND

FRUSTRATION / MENTAL DEMAND

EFFORT / MENTAL DEMAND

TEMPORAL DEMAND / PHYSICAL DEMAND

PERFORMANCE / PHYSICAL DEMAND

FRUSTRATION / PHYSICAL DEMAND

EFFORT / PHYSICAL DEMAND

TEMPORAL DEMAND / PERFORMANCE

TEMPORAL DEMAND / FRUSTRATION

TEMPORAL DEMAND / EFFORT

PERFORMANCE / FRUSTRATION

PERFORMANCE / EFFORT

EFFORT / FRUSTRATION