Simulator Fidelity: The Effect of Platform Motion

Judith Bürki-Cohen
U.S. Department of Transportation
Research and Special Programs Administration
John A. Volpe
National Transportation Systems Center
Cambridge, MA 02142

Nancy N. Soja
Brookline, MA 02446

Tiauw H. Go
Massachusetts Institute of Technology
Cambridge, MA 02139

Edward M. Boothe
Stone Mountain, GA 30083

Robert DiSario
Bryant College
Smithfield, RI 02917-1284

Young Jin Jo
Battelle Memorial Institute
Cambridge, MA 02142

Final Report
October 2001

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY	2. REPORT DATE July 2000	3. REPORT TYPE AND DATES COVERED April 2000

4. TITLE AND SUBTITLE
Simulator Fidelity: The Effect of Platform Motion

5. FUNDING NUMBERS

6. AUTHOR(S)
Judith Bürki-Cohen*, Nancy N. Soja**, Tiauw H. Go***, Edward M. Boothe****, Robert DiSario*****, Young Jin Jo******

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
*U.S. DOT-Volpe National Transportation Systems Center, Cambridge, MA 02142
* Consultant, Brookline, MA; ***Massachusetts Institute of Technology, Cambridge, MA; ****Consultant, Stone Mountain, GA; *****Bryant College, Smithfield, RI; ******Battelle, Cambridge, MA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Department of Transportation
Federal Aviation Administration
Office of the Chief Scientific and Technical Advisor for Human Factors
Washington, D.C. 20591
Program Manager: Dr. Eleana Edens

10. SPONSORING/ MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

12b. DISTRIBUTION CODE

13. ABSTRACT (200 words)

As part of the Federal Aviation Administration's (FAA) initiative towards affordable flight simulators for U.S. commuter airlines, this study empirically examined the effect of six-degree-of-freedom simulator platform motion on recurrent pilot training and evaluation in the presence of a wide field-of-view visual system. Sound scientific data on the relationship between the motion requirement and its effect on the transfer of pilot performance/behavior to and from the airplane is all the more important given that the FAA may mandate the use of simulators for airline pilot training and evaluation. The study addressed the question of whether the motion provided by an FAA qualified Level C simulator affects 1) pilot performance/behavior and instructor grading criteria during **First Look evaluation**, 2) the **course of Training** in the simulator, and 3) the **Transfer of skills** acquired during Training in the simulator with or without motion to the simulator with motion as a stand-in for the airplane. Every effort was made to avoid deficiencies in the research design identified in a review of prior studies, by measuring pilot stimulation **and** response, testing both maneuvers and pilots that are **diagnostic** of a need of motion, avoiding pilot and instructor **bias**, and ensuring sufficient statistical **power** to capture operationally relevant effects.

14. SUBJECT TERMS	15. NUMBER OF PAGES
	16. PRICE CODE

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

Preface

This work was performed as part of an ongoing research program at the U.S. Department of Transportation's John A. Volpe National Transportation Systems Center investigating fidelity requirements for simulators used in training and evaluation of airline pilots. The research is supported by the Federal Aviation Administration's Office of the Chief Scientific and Technical Advisor for Human Factors (AAR-100). We thank its Air Carrier Training Program manager Dr. Eleana Edens for her effective guidance and assistance. The need for this work was established by the Advanced Qualification Program (AFS-230). We thank its manager Dr. Thomas Longridge for his insights. Many thanks also to Mr. Paul Ray, manager of the National Simulator Program Office (AFS-210), and Mr. Ed Cook of the same office, for their continued interest and advice.

Many people have contributed to this work. Most of all, we would like to thank the host airline and training center for their participation. Without the expertise and collaboration of their personnel, this work would not have been possible. Ms. Julie Guinn provided excellent support during her senior year in college. Dr. Shuang Wu designed and implemented the laptop computer interface controlling the experiment. Mr. Frank Cardullo of SUNY Binghamton educated three of us at his yearly simulator courses and generously shared his expertise. Mr. John K. Pollard of the Volpe Center's Operator Performance and Safety Division provided invaluable technical advice regarding measurement equipment and programming. Dr. Michael Zuschlag of the same division provided cogent comments on an earlier draft. Dr. E. Donald Sussman, Chief of the Operator Performance and Safety Division, provided comments throughout the report generation. We extend our thanks to all of them.

Blank Page for Metric/English Conversion

Table of Contents

List of Figures

Figure **Page**

List of Tables

Acronyms and Abbreviations

AGL:	Above Ground Level
ALPA:	Air Line Pilots Association
ANOVA:	Analysis of Variance
ATC:	Air Traffic Control
cg:	center of gravity
DAC:	Data Acquisition Computer
DOF:	Degrees of Freedom
FAA:	Federal Aviation Administration
FAR:	Federal Aviation Regulation
FOV:	Field of View
I/E:	Instructor/Evaluator
LOFT:	Line Oriented Flight Training
LP:	Low Pressure
N, *n*:	Sample size
NSPO:	National Simulator Program Office
p:	Probability of null hypothesis (i.e., no effect of motion)
PF:	Pilot Flying
PNF:	Pilot Not Flying
PTS:	FAA Practical Test Standards
QTG:	Qualification Test Guide
r:	Pearson correlation coefficient
RMS:	Root Mean Square
RTO:	Rejected Take-Off
SAS:	Statistical Analysis Software
SFAR:	Special Federal Aviation Regulation
SME:	Subject Matter Expert
STD:	Standard Deviation
STIG:	Simulator Technical Issues Group
V_1:	Take-off decision speed; the minimum speed in the take-off, following a failure of the critical engine, at which the pilot can continue the take-off and achieve the required height above the take-off surface within the take-off distance.
V_2:	Take-off safety speed; a speed that will provide at least the gradient of climb required by the airplane certification rules with the critical engine inoperative.
V_R:	Rotation speed; the speed at which backpressure is applied to the controls to increase the airplane's pitch attitude to the target take-off pitch attitude.

Definitions[1]

First Look:	First Look evaluation; the first experiment (i.e., the first part of the combined experiment) examining the effect of motion when using the simulator as an *evaluation* tool of pilots' aviating skills. The crews flew one V_1 cut followed by on RTO, without having been in the simulator for the previous six months and without having been informed that engine failures would occur. The two First Look evaluation trials also served as the first Training trials of the second experiment (i.e., the second part of the combined experiment) examining the simulator as a *training* tool.
Motion Group:	those crews that experienced simulator motion during all parts of the combined experiment, First Look, Training and Transfer.
motion group(s):	is used when referring to both the Motion and the No-Motion groups.
No-Motion Group:	those crews that did not experience simulator motion during First Look and Training, but were tested with the simulator motion turned on during Transfer.
RTO:	engine failure below V_1 with crews required to reject the take-off.
Training:	the first part of the second experiment (i.e., the second part of the combined experiment) examining the effect of motion when using the simulator as a *training* tool. It included First Look as the first Training trial for each maneuver, and might have been followed by additional Training trials. Any additional RTOs were given first, followed by additional V_1 cuts. Training was concluded for each maneuver when either a grade of 3 or 4 was achieved or after three trials, including First Look. To enhance training during the second and third Training trials, crews were informed about each upcoming engine failure.
Transfer:	Transfer testing; the final part of the combined experiment examining the effect of simulator motion on training effectiveness. It determined whether motion affects how the skills learned in the simulator transfer to the simulator with motion as a stand-in for the airplane. Crews flew one V_1 cut followed by one RTO without having been alerted to the engine failures.
V_1 cut:	engine failure at or after V_1 with crews required to continue take-off.

[1] The objective measures and their definitions are listed in Appendix I (for RTOs) and Appendix J (for V_1 cuts).

Executive Summary

SUMMARY OF BACKGROUND AND METHOD

This research effort is part of the Federal Aviation Administration's (FAA) initiative towards the promotion of availability and affordability of effective flight simulators for all U.S. airlines. Simulators provide a safe and effective means for pilot training and evaluation. This enables presentation of scenarios including emergencies requiring both technical and crew resource management skills. Therefore, the FAA is proposing a rule that would mandate the use of simulators for all air carrier flight crew training and qualification, limiting the use of the aircraft itself as a training option even for small regional airlines. However, there is a lack of sound scientific data on the relationship between certain key training device features, such as platform motion cuing, and their effect on the transfer of performance to and from the airplane. The goal of this project is to provide a scientific basis to assure that FAA requirements are commensurate with safety objectives. Particularly, it addresses the question of the need for simulator motion for commuter airline pilot recurrent training and evaluation in the presence of a state-of-the-art visual system. The data will also help the FAA to evaluate air carrier proposals for the alternative use of full flight simulators, whose availability and affordability may be limited especially for small regional airlines, and other training equipment.

The first stage of this multi-year project was a state-of-the-art review of key aspects of simulation, involving both FAA/Industry subject matter expert workshops and an extensive literature review. Based on this review, an empirical investigation of the requirement of simulator platform motion has been initiated, which seeks to correct deficiencies in the research design of prior studies. The present study examined the effect of FAA qualified Level C six degree-of-freedom synergistic motion in the presence of a wide-angle collimated visual system with cross-cockpit viewing on pilot training and pilot evaluation. Transfer of skills acquired in the simulator to the airplane was measured by comparing the effect of training received in the simulator, with and without motion, on performance and behavior in the simulator with motion. This “quasi-transfer” to the simulator with motion as a stand-in for the airplane ensured the safety of the participants while allowing full experimental control.

The research was conducted using regional airline pilots in recurrent training. Two test maneuvers were chosen as diagnostic for an effect of motion on pilot training and evaluation, namely, engine failures on take-off with either rejected take-off (RTO) or continued take-off (V_1 cut). These maneuvers minimally disrupted the host airline’s training program while satisfying the criteria recommended in the literature as diagnostic for detection of a motion requirement. These criteria included 1) closed loop, to allow for motion to be part of the control-feedback loop to the pilot; 2) unpredictable and asymmetric disturbance, to highlight an early alerting function of motion; 3) high gain and high thrust, to magnify any motion effects; 4) high workload, to increase the need for redundant cues such as provided by motion, out-the-window view, instruments and sound; and 5) short duration, to prevent pilots from adjusting to a lack of cues. To prevent bias, the state of the motion system was kept concealed from all participants.

Both the stimulation of pilots by the simulator and the pilots' responses were measured by recording nearly eighty simulator-state and control-input variables at a high sampling rate, resulting in a vast amount of objective data on simulator performance and pilot performance, behavior, and workload. Two forms of subjective data were collected. First, at the conclusion of each maneuver the instructor/evaluator (I/E) provided a grade for the just-completed maneuver. Second, at the end of the Training period and again at the end of the Transfer period the I/Es, Pilots Flying, and Pilots Not Flying provided ratings for the simulator acceptability and training effectiveness, and pilot performance, behavior, workload and comfort. The study addressed the questions of whether the motion provided by an FAA qualified Level C simulator affects

1) First Look evaluation of pilot performance/behavior prior to any simulator practice,
2) the course of Training in the simulator, and
3) the Transfer of training acquired during Training in the simulator with or without motion to the "airplane."

The analysis also examined, by determining the relationship between subjective grades and objective measures, how the I/Es' grading criteria were affected by the presence or absence of motion.

SUMMARY OF RESULTS

Motion Stimulation Provided by Simulator

For roll and longitudinal accelerations, the actually measured stimulation followed the airplane model fairly well given the limitations inherent to all simulators. Failure induced lateral acceleration, however, was not well represented by the motion system of the test simulator. Not only was it greatly attenuated, but also visual inspection of the measured response did not lead to an easy distinction of failure-induced lateral acceleration, unlike the response derived from the equations of motion (relatively high peak shortly after engine failure). This may have represented an important deficiency in pilot stimulation, because lateral acceleration may act as a useful cue for proper failure recognition and for delivery of appropriate action. To the best of our knowledge, however, the importance of lateral versus other cues in failure recognition has not been systematically examined in the literature. Comparison of the lateral acceleration data of the test simulator with data from some other Level C and D simulators indicates that this deficiency may well be shared with other FAA qualified simulators.

Grades Provided by Instructor/Evaluators Following Each Maneuver

The possible grades were 1 (unsatisfactory), 2 (FAA Practical Test Standards), 3 (company standards), and 4 (excellent). The experimental session appeared to have been effective in simulating a real training session in that the crews' performance improved across the session. Specifically, combining the two motion groups (or looking at them individually), the grades for RTOs and V_1 cuts improved across the Training trials. This was even stronger for the V_1 cuts, which elicited lower grades than the RTOs during First Look, but caught up by Transfer.

Turning to the effect of motion, the presence or absence of motion had no effect on the grades for the RTOs at either First Look or Transfer. There was also no effect of motion for the V_1 cuts at First Look. At V_1 cut Transfer, motion training did not affect grades when comparing the group means or the number of low vs. high grades in each group (i.e., grades of 1 and 2 vs. grades of 3 and 4). However, crews that had previously had motion did receive more grades of 2 than the crews who had not previously had motion, and fewer grades of 1 (none actually). Despite this single effect of motion, there was no effect of motion on the course of Training or on the amount of Training required before reaching the criterion needed to proceed to Transfer.

Objective Measures

The objective measures that are included in this summary are the ones that are either listed in the FAA Practical Test Standards (PTS) (FAA, 1995), were used by the instructors for grading as shown by correlation/regression analyses, or showed an effect of motion. For each measure, the statistical power was determined (i.e., the smallest effect that could be detected given the idiosyncratic variability between crews). The power of the experiment was found to be sufficient to capture any operationally relevant effects.

I/Es' Grading Criteria

Regression analyses on the relationship between the grades and the objective measures were used to infer the I/Es' grading criteria and whether the platform motion had an effect on these criteria. For RTOs, regardless of whether the platform motion was on or off, the measures of lateral and heading deviations played an important role in predicting the I/E grades. For V_1 cuts, regardless of motion status of the simulator, bank angle appeared to affect I/E grades. Longitudinal measures, however, appeared to matter mainly when the platform motion was on. It is conceivable, then, that I/Es could have overlooked a deficit in longitudinal performance during First Look without motion. Findings from the objective data analysis, however, showed that the differences in the longitudinal performance between the two groups were negligible. Moreover, the regression models obtained accounted for only a small portion of the variance in the grades. These findings confirm the lack of an effect of motion on pilot performance/behavior found in the grades (see below).

First Look Evaluation, RTOs

Motion did not affect performance in heading deviation, lateral deviation, Power Lever Reaction Time,[2] nor did it affect any workload measures (operation of controls). It did, however, improve Integrated Yaw Activity, a measure that was not found to be an important I/E grading criterion. This suggests that for First Look evaluation, motion may improve performance, but not to the extent of affecting grades.

First Look Evaluation, V_1 Cuts

Motion did not affect bank angle or heading control variables (and these, especially bank angle, were important for grades) or reaction times. Interestingly, it might have improved the

[2] See Appendices I and J for a description of all objective measures.

Pitch Angle Standard Deviation marginally ($p < .10$). This effect was, however, physically small and not accompanied by any other performance or workload effects. This, together with the fact that Pitch Angle Standard Deviation alone does not affect grading criteria, indicates that platform motion would not affect pilot grades during First Look evaluation.

Transfer, RTOs

Transfer of training was tested for all crews on the simulator with motion activated as a stand-in for the airplane. Despite the fact that the Motion crews were trained and tested on the same simulator configuration, they did not do any better than the No-Motion crews with any RTO variable. Additionally, the overall power of the experiment was higher after Training, and still no effects of prior motion were found. One *caveat* is that for heading control, although there was no difference between the two groups, the No-Motion group improved more than the Motion group between the last Training and the Transfer testing. This may indicate that the addition of motion was beneficial, although during Transfer testing the two groups performed at the same level (as just described).

Transfer, V_1 Cuts

Having been trained with motion did improve speed control (i.e., Integrated Airspeed Exceedance) with V_1 cuts during Transfer. It came at the price of increasing Pitch Angle Standard Deviation, but may still be advantageous because of the critical role speed plays in aircraft control and clearing obstacles. However, there was also an increase in Integrated Yaw Activity shown by the motion-trained group, even though it did not appear to affect heading.

With regard to workload during V_1 cuts, the motion-trained group worked harder on the pedal but less hard on the wheel than the fixed-base trained group (with very minor performance differences, however). The reason for this is not clear. The difference was not there at First Look, and a combined ANOVA of Motion/No-Motion by First Look vs. Transfer did not find a significant interaction. The questionnaire data indicated that the Motion group felt the pedal was less like the airplane than the No-Motion group did.

Note that all the differences found were physically small and may not be relevant operationally.

Training Progress, RTOs and V_1 Cuts

There were no differences between the two motion groups for RTOs on the course of Training. Training progress for V_1 cuts reflected the Transfer results. Motion did improve progress during Training for Integrated Airspeed Exceedance, but hindered Training for STD Pitch (the price for reduced Integrated Airspeed Exceedance) and also for the heading variables. For workload variables, there were no differences between the two motion groups.

Questionnaire Data

Pilots Flying (PF), Pilots Not Flying (PNF), and I/Es were given two questionnaires, one after Training and one after Transfer). Each asked about the PF's control precision/performance, control strategy and technique,[3] workload, and Training progress. Each also asked about the respondents' own comfort, and only the crew members about the acceptability of the simulator. Some of the questions contained many subparts. Responses were examined for motion effects and for timing effects (i.e., whether they changed after Transfer).

Only four statistically significant, but contradictory, differences were found between the Motion and No-Motion crews. The *No-Motion* PFs rated themselves higher with regard to Training progress, but only when the two questionnaires were combined. Their control precision was rated higher by the PNFs (but only after Training) and themselves (but only after Transfer). The latter finding, however, was contradicted by the I/Es, who rated the *Motion* PFs higher after Transfer. This difference, however, was not evident from the grades, at least not when comparing the group means or the number of low grades (1 and 2) vs. high grades (3 and 4).

CONCLUSION

The results of this study indicate that the motion provided by the test simulator does not, in an operationally significant way, affect either First Look evaluation, Training progress, or Transfer of training acquired in the simulator with or without motion to the simulator with motion. Motion also does not consistently affect the PFs', PNFs', and I/Es' subjective perception of the PFs' performance, workload, and Training progress, or of their own comfort in the simulator. Neither does it affect the acceptability of the simulator by the PF and the PNF.

Two *caveats* have to be kept in mind, however. First, even though the test simulator was an FAA qualified Level C simulator, it may not have provided sufficient motion to be effective. The measurements indicate that the simulator may have failed to provide lateral acceleration cuing representative of the aircraft for the test maneuvers (RTO and V_1 cut). Data from some other Level C and D simulators, however, suggest that the motion performance of the test simulator was representative.

A second *caveat* is that the current study used the simulator with motion as a stand-in for the airplane. Although some may believe that this "quasi-transfer" design needs to be validated, others say that high-level simulators have been validated as a stand-in for the airplane by use of the simulator for total flight training. Also, given that the motion-trained group transferred to the same simulator configuration that they had been trained in, whereas the No-Motion group transferred to a configuration that was new to them (i.e., the motion configuration), the Motion group should have had an advantage. It is unlikely that it would have had a greater advantage transferring to an airplane.

Clearly additional steps must be taken to determine the extent to which it is appropriate to draw generalizations from these results. A first step was the comparison of the objective measures from the motion system used in this experiment with such measures taken from other FAA qualified Level C simulators, which suggested that the motion used in the present study was representative. This should be followed by an investigation on whether operational relevant effects of motion would be found with a simulator where the motion is manipulated to assure that it is representative of the airplane for the maneuvers selected. Additional maneuvers that may be

[3] The question on control strategy and technique was not considered in the analysis of the questionnaires. It contained some inadvertently ambiguous language.

diagnostic and a different pilot population should be tested as well. Ideally, some validation of the quasi-transfer design with a real airplane could also be undertaken.

1. INTRODUCTION[4]

1.1 BACKGROUND AND MOTIVATION

1.1.1 Simulators Needed for All Airlines

This research effort is part of the Federal Aviation Administration (FAA) initiative towards more affordable flight simulators for U.S. commuter airline training. The FAA is proposing a rule that would mandate the use of simulators for all air carrier flight training and qualification, limiting the use of the aircraft itself as a training option even for small regional airlines. This research will provide a scientific basis to assure that FAA requirements are commensurate with safety objectives as well as the widest access to simulator training.

Today, the use of airplane simulators in pilot training and evaluation is universal. Simulators not only enable savings in training cost, but they have also practically eliminated training accidents for major airlines. They allow the training of emergency maneuvers, which are inherently unsafe in the aircraft; and they permit crews to gain experience in operationally realistic scenarios that focus on both technical and crew resource management skills. In fact, initiatives such as the Federal Aviation Administration's (FAA) Advanced Qualification Program (Federal Aviation Regulation [FAR], Part 121, SFAR 58, 1990), which heavily relies on crew resource management and need-based proficiency objectives, would be unthinkable without ready access to a full flight simulator.

Nevertheless, some regional airlines elect to do at least their recurrent training in the airplane. In part, this situation is due to a shortage of qualified simulators, especially for certain turboprop airplanes where the flight test data is not readily available. A second, and perhaps even more important, reason can be found in the prohibitive rental and purchase costs for full flight simulators, which, for small turboprops, may even exceed the cost of the airplane. The motion platform in particular represents a substantial portion of the cost. In addition to the cost of the platform and hydraulics themselves, there are costs for the enlarged building and maintenance to house the platform, for the hardware and software to coordinate the visual system with the movements of the cockpit, and for the added wear and tear on all other systems due to the jolting from the motion platform (Roscoe, 1980).

Despite these overwhelming costs for regional airlines, simulator training is at least as necessary for regional pilots as for the pilots of major airlines given that regional pilots tend to have less experience, pilot turnover is higher for regional pilots, certain regional airport environments are less well-equipped with navigation aids, and the pilot's response to power plant malfunction or engine failure in turboprop aircraft can be more demanding than for jet aircraft. Clearly a consideration of what measures can be taken to increase the accessibility of simulator training for regional pilots would be beneficial.

[4] Some of the information and opinions presented here can also be found in Bürki-Cohen, Soja, and Longridge, 1998a and in an extended form in Bürki-Cohen, Soja, and Longridge, 1998b.

1.1.2 Simulator Qualification Standards Review

As a first step in addressing the availability and affordability problems of airplane simulators, a group of experts from industry, academia, and government was convened to review the existing simulator qualification standards contained in Advisory Circular AC 120-40B (Federal Aviation Administration, 1991). The subject of this review was the Level B simulator, which can be used for 100% recurrent training and evaluation of qualified airline pilots. The mandate was to examine the standards for ways to simplify the requirements such as to achieve a reduced-cost simulation without loss of training and evaluation effectiveness. A first symposium focused on the aeromodel validation standards. The participants envisioned a savings of up to 50% on the flight test data package by simplifying the validation test and flight instrumentation and using some predictive modeling for flight data estimation (Joint FAA/Industry Symposium, 1996a).

A second symposium focused on the motion requirements for the Level B simulator. The international panel of experts felt that motion may have an important alerting function in maneuvers entailing sudden motion-onset cuing, such as loss of engine during initial segment climb, where visual references are limited. Currently, the Federal Aviation Regulation requires motion in three unspecified degrees-of-freedom (DOF). The panel recommended that for motion to have a beneficial effect, it should at least encompass four DOF, namely, pitch, roll, sway, and heave. Also, the panel advised that the allowable motion transport delay should be reduced from 300 to 150 ms. Both of these suggestions ran somewhat counter to the FAA's goal of finding a safe way to make simulators more affordable, and are indicative of the conflict that was apparent throughout the symposium discussions between recommending the "best available motion" and "motion good enough" for the intended purpose (Longridge, Ray, Boothe, & Bürki-Cohen, 1996; Joint FAA/Industry Symposium, 1996b).

1.1.3 Simulator Fidelity Requirements

In the past, technical constraints provided a natural limit to the fidelity of a simulation. Today, however, technical capabilities have expanded to a point where they may enable a degree of fidelity that may exceed the one required for a particular purpose. This may lead to a situation where the benefit resulting from increased fidelity no longer justifies its cost. The focus thus needs to shift from ever more sophisticated technology to the level of fidelity required to train and evaluate to a specific safety standard.

When determining the required fidelity of a device for a particular purpose, we have to distinguish between physical and perceptual fidelity (Advisory Group for Aerospace Research and Development, 1980). **Physical fidelity** of a flight simulator is relatively easy to determine. Using carefully calibrated instruments, simultaneous recordings of all pertinent variables of both the airplane simulator and the simulated environment are compared with the corresponding measurements from the pilot's seat in the actual airplane (Ashkenas, 1985). The closer the match, the more the simulator is faithful to the airplane. A more valid measure, however, may be **perceptual fidelity**. It is defined not only as a match between pilots' subjective perceptions of the simulator and the airplane, but also as a match between pilots' performance and control strategy in the simulator and the airplane. Its determination requires carefully controlled pilot-in-the-loop experiments.

As a second step in the FAA's effort to increase the availability and affordability of airplane simulators, the question was addressed as to what level of fidelity is required to simulate airplane motion in pilot recurrent training and evaluation, and how it can best be achieved. The discussion of physical versus perceptual fidelity is especially pertinent in the context of simulator motion, which is inherently limited in its physical fidelity despite substantial technological advances. Even as late as 1989, Brown, Cardullo, and Sinacori state that "[b]arring an unforeseen revolution in the technology of force and motion cuing, it is evident that it is hopeless to attempt to provide realistic force and motion stimuli in the sense that the acceleration forces produced by the aircraft can be replicated in the simulator" (p. 78). In particular, it is impossible to simulate sustained acceleration without sustained displacement, and any direct application of whole body acceleration forces will require inappropriate counter forces. The only way out of this dilemma is to focus not on the reality of the force and motion stimuli, but on the perceptions associated with force and motion, i.e., perceptual fidelity. One of the common misconceptions is that the higher the physical fidelity of the simulation the better the training will be. The level of fidelity of the simulator should be determined, however, by the level needed to support the learning/evaluation of the tasks that will be trained/evaluated using the device (Salas, Bowers, & Rhodenizer, 1998).

1.1.4 Perception of Motion

Motion occurs in space and over time. The most important systems for the perception of self-motion are the vestibular system and the visual system. The vestibular system resides in the inner ear and perceives angular velocity and linear acceleration (Hall, 1989). The visual system perceives motion from changes in position, and perceives velocity and acceleration by additionally taking time into account (Sedgwick, 1986). For the perception of self-motion, the peripheral visual system is especially important (see, e.g., McCauley, 1984; Dichgans & Brandt, 1978). Both the vestibular and visual systems are being stimulated in airplane simulations to induce the perception of self-motion. The tactile or somatic systems also perceive self-motion and have been stimulated in the past with dynamic seatpans (see, e.g., Martin, 1985), but these are currently not under consideration for airline pilot training and evaluation.

The phenomenon of perceiving illusory self-motion from vision alone is called vection. A familiar example of such an illusion may occur while seated in a stationary car. When the neighboring car moves forward, people in the stationary car perceive their car as moving backward. An example from an experimental setting is when subjects are placed in a patterned drum moving around them. At first, subjects correctly perceive the drum to rotate, but very rapidly the illusion of self-rotation develops. It is this illusion that may enable vision to replace physical motion in the simulator. However, the delay in the onset of the illusion may undermine the possibility of using vision alone because a primary role for physical motion in simulators may be to serve as an early alerting system (e.g., Gundry, 1976). In fact, Boldovici (1992), in quoting a written correspondence with Stark, wrote "Non-visual cues are important because (1) they tend to be perceived and used before there is time to analyze correlated visual cues, (2) they do not require a specific focus of attention, (3) they tend to be available when visual cues are temporarily absent" (p. 16).

1.1.5 The Need for Objective Empirical Evidence

Boldovici (1992) reports that "Vestewig and others told me about cases in which instructors turned off motion bases during aircraft simulator exercises without students noticing the difference" (p. 9). The problem with such anecdotal evidence is that we are often not aware of how we are affected by our environment. Background low-level noise, for example, is often ignored when focusing on the task at hand, but may still result in exhaustion at the end of the day. In fact, most of human information processing is subconscious. Just as we are able to prevent unwanted information from distracting us, pilots may also be very good at compensating for cues that are missing perhaps by simply working harder, while remaining unaware of both the missing cues and the extra effort.

Compelling evidence from controlled studies addressing the role of motion in airplane simulations is needed for the FAA, whose primary concern is passenger safety, to consider a change in regulations that have been successfully applied for the past twenty years. Moreover, flight simulators are used for *pilot certification (licensing)* as well as *training*. Not only are simulators used to train pilots, but they are also used to determine whether or not a pilot is ready to fly the respective airplane with passengers in revenue service. That is, the simulator is used for total training and evaluation of airline pilots, with no additional training or evaluation in an airplane without passengers. Unlike many classical transfer-of-training situations, where supplemental training may occur in the transfer device, the simulator-trained air-carrier pilot must perform within satisfactory standards of proficiency in the aircraft from day one. Consequently, the simulator must be capable of supporting 100% transfer of performance to the aircraft. Anything less would compromise the safety of the flying public. To begin this compilation of evidence, the existing research will be reviewed. For a more complete description of each experiment, and a few others, please see Bürki-Cohen, et al., 1998b.

1.2 PREVIOUS RESEARCH

1.2.1 Acceptability of Simulator

The consensus is that pilots prefer simulators with motion. For example, Reid and Nahon (1988) and Hall (1978) used different simulators and different acceptability ratings and yet both found that pilots prefer that the simulator move. Interestingly, however, motion appeared to be most important when there was no visual information available besides instruments. Also, there is still some question as to the impact of motion on acceptability ratings because in most studies the pilots were informed of the simulator's motion state and that knowledge could have influenced their ratings as much as the actual state of the simulators. Indeed, when Lee and Bussolari (1989; Bussolari, Young, & Lee, 1987) did not inform the pilots of the motion state of the simulator, these pilots did not prefer the simulator with six DOF motion. However, this study is not fully conclusive either because of an additional change they made: Instead of turning the motion off completely, they let the simulator vibrate at an amplitude of 1 cm in heave.

Thus, it is not clear whether the general preference among pilots for motion is due to an actual preference or the presence of a preconceived bias towards motion. Additionally, even if there is a real preference, it may be possible to eliminate it by simply adding vibration to the no-motion simulator.

1.2.2 Performance/Control Behavior *in Simulator*

More important than subjective preference is knowing which simulator configuration results in best flying performance and behavior. The vast majority of studies examining this question obtained data in the simulator only, without regard for performance/behavior in the actual airplane.

When investigating the role of motion cues when controlling an airplane, it is important to differentiate between two general kinds of tasks, tracking tasks and disturbance tasks. In a **tracking task,** the crew is asked to track a random signal, such as a specific flight path or the lead airplane in formation flight. In this sort of task, the motion is part of the control loop and provides the pilot with feedback on the appropriateness of his own intended control actions, and thus is unlikely to act as an alerting cue. An exception to this is a task of controlling unstable aircraft, where motion is an essential cue to the pilot.

In a **disturbance task,** the crew needs to correct for a random perturbation of the controlled system, such as stabilizing the airplane in turbulence or compensating for a mechanical failure (e.g., an engine failure). In this case the motion of the airplane arises from outside the control loop and is unexpected by the pilot. Thus, platform motion provides an early alerting cue to the disturbance which could potentially enable a more rapid response with motion than would be achieved with visual cues alone.

Given these differences in the role of motion for the two kinds of tasks, it is important to examine the role of platform motion individually for each. For tracking tasks, Hall (1978) and Hosman and van der Vaart (1981; Hosman, 1996) found only a small effect of motion on performance. Pilots provided with motion cues showed slightly less roll angle error than pilots without. Moreover, control behavior was affected by motion cues only with unstable aircraft. In that case, there was an increase in stability for pilots with motion, but there was a concomitant loss in gain.

In contrast, there was a large effect of motion with disturbance tasks. Hosman and van der Vaart (1981) found that pilots who received motion cues performed much better, in terms of roll angle, than the pilots who did not. Furthermore, the presence of motion cues improved control behavior for all aircraft, whether stable or unstable, by increasing gain without impacting stability.

In addition, Hall (1978) and Hosman and van der Vaart (1981) each examined the interaction between vision (central and peripheral displays) and motion cues. For both kinds of tasks they found that the effect of motion was strongest when there was no visual information available. That is, visual information could compensate for the lack of motion to a certain extent. Even so, vision alone, even with peripheral vision included, was not as good as vision and motion together.

Thus, both Hall (1978) and Hosman and van der Vaart (1981) concur in finding that the presence of motion improved pilot performance and behavior in the simulator, and that this improvement cannot be duplicated by the presence of peripheral vision in the absence of motion. Hosman and van der Vaart (1981) further demonstrated that the effect of motion is mediated by the kind of maneuver, both in terms of the strength of the effect and the type of the effect. That is, the performance results indicate that the need for motion is greater with disturbance maneuvers than with tracking maneuvers; and the control behavior assessment indicates that the effect on disturbance maneuvers is an increase in pilot gain, whereas the effect on tracking maneuvers is an increase in stability (and a loss of gain).

1.2.3 Transfer *to Simulator* (Quasi-Transfer)

The studies discussed so far were only concerned with the potential benefit of platform motion on performance and behavior *in the simulator*. The real task, however, is flying *the airplane*. For simulator motion to be useful in training, it would have to increase the proficiency of the pilot after the training session, in the airplane. Similarly, pilot evaluations conducted in the simulator with motion would have to reflect pilot proficiency in the airplane better than evaluations conducted without platform motion.

Conducting experiments in the simulator, however, shares some of the same advantages as using simulators in training and evaluation, that is, it enables controlled conditions in a safe and cost-effective setting. Consequently, some scientists compared the training effectiveness of competing simulator configurations by evaluating which elicits the best transfer to a stand-in for the airplane, i.e., a configuration that is presumed to represent most faithfully the airplane. This paradigm is called "quasi-transfer" because it does test transfer, but not to an actual airplane.

Levison (1981; Levison & Junker, 1977) used a quasi-transfer paradigm to study the effects on training of different types of motion simulation methods, including vision-only, synchronous vision and motion, and three conditions where motion lagged vision by 80, 200, and 300 ms. Subjects were trained to control roll angle in gust-like disturbances. During training, large roll angle reductions were observed in all conditions, but especially in the shortest lag and no-lag motion conditions. In fact, the synchronous motion group achieved asymptotic performance so early that Transfer testing was omitted for this group. Additionally, all groups, except the group with the longest lag, performed better than the no-motion group.

When testing transfer of this training using the synchronous motion condition as a stand-in for the airplane, all groups showed immediate improvement, but only the 80 ms group caught up with the synchronous motion group on the first post-transition trial, thus indicating transfer of training with non-synchronous motion. The vision-only group caught up eventually, and in fact, performed at the same level as the 200 ms lag group (and continued to outperform the 300 ms lag group), thus repudiating the motion advantage for cases with badly synchronized motion. In sum, the performance benefit of simulator motion experienced during training did transfer to the higher-fidelity simulator, but only when motion and vision were nearly synchronized.

The same was true also for control strategy. Presumably, "subjects trained initially with the 80-msec delayed motion cues were exposed to a perceptual situation more like the transfer task than were subjects trained fixed base, and were therefore able to more quickly learn to process faithful motion cues and adopt the appropriate control strategy in the transfer condition" (Levison, 1981, p. 22).

In sum, the Levison (1981) study significantly extends the findings that motion increases proficiency in the simulator. It showed that this motion advantage transfers to a higher-fidelity device, suggesting that it may transfer to the airplane as well.

1.2.4 Transfer *to Airplane:* Tracking and Disturbance Tasks

Despite the inherent constraints on transfer-to-airplane studies, several people have attempted them (e.g., Koonce, 1974; Jacobs, 1776; Ryan, Scott, & Browning, 1978; Martin, 1981). Not surprisingly, these studies showed that simulator training in general improved proficiency in the airplane, regardless of simulator motion state. Also, nearly all these studies confirmed that simulator motion is associated with superior performance *during training within* *the simulator*, as seen in most simulator-only and quasi-transfer studies. In contrast to Levison's (1981) quasi-transfer study, however, these true transfer studies did *not* support the supposition that the advantage of motion *transfers to real airplanes*.

All of the transfer-to-airplane studies, however, share a number of problems that may have diminished their diagnosticity. Many of these problems also apply to the simulator-only and the quasi-transfer studies and many were beyond the control of the scientists such as the state of technology at the time of the experiment and the problems inherent in using airplanes.

First, many studies used outdated motion systems. The studies succeeded in showing that such relatively crude motion systems improved how the simulator was flown. Perhaps a newer more sophisticated motion system would improve transfer of training to the airplane as well. It should also be noted that the studies similarly employed rudimentary visual systems. The studies showed that the visual systems could compensate somewhat for the lack of motion, at least within the simulator, suggesting the possibility that newer more sophisticated visual systems could fully replace motion at least for some purposes.

Second, most of the experiments used tracking instead of disturbance maneuvers, the latter being both difficult and dangerous to test in the air. Only disturbance maneuvers, however, may be able to diagnose the advantage of exposure during training to the early alerting cues provided by motion. In fact, Gundry (1976, for example) asserts that the alerting function is assumed to exist only for disturbance motion, not maneuver motion.

Third, many of the experiments used non-representative subject samples, both with respect to number of subjects sampled and their flying experience. None of the studies cited so far analyzed the interpilot variability within groups to determine the number of pilots required to find a specific size of effect. In fact, Boldovici (1992) said, “Inadequate statistical power is ubiquitous in military training research. Finding no differences as functions of treatments often seems virtually guaranteed, not because differences are absent, but because experiments lack power and are otherwise deficient” (p. 5). Moreover, most of the studies used student pilots. There is evidence, however, that well-trained pilots may be more sensitive to the presence or absence of motion than beginner pilots (Young, 1967).

Fourth, only some of the studies analyzed both pilot performance and control input behavior. In order to achieve the same performance in different equipment, pilots will attempt to adapt to deficiencies in equipment by changing their control strategy, e.g., they may increase the frequency and/or amplitude of their control interventions. Such differences in control strategy acquired in the simulator may be detrimental in emergency situations in the airplane.

Fifth, pilots and instructors were not naïve regarding the motion condition, which may have allowed bias to affect their performance or performance evaluation, respectively (Ebbinghaus, 1964).

1.2.5 Summary of Empirical Evidence

In sum, there were many benefits of platform motion within the simulator. First, it improved the acceptability of the simulator, at least when pilots were aware of the motion manipulation (but the amount of motion required may be very small). Second, it improved pilot performance and control behavior for disturbance tasks while training. Third, it improved behavior during a tracking task with an unstable vehicle while training. Finally, it was particularly useful when visual information was limited. Some of the benefits of platform motion transferred to a higher fidelity simulator. In contrast, the benefits of motion have not been proven in the critical case of the transfer of training to the airplane.

1.3 CURRENT RESEARCH

1.3.1 The Need to Revisit Motion Fidelity Requirements

Four decades of research did not provide conclusive evidence that vestibular motion cuing in simulators used for recurrent pilot training and checking is beneficial. Technological advances, industry interest, as well as the lessons learned from previous research provide excellent grounds for readdressing this question.

In the wake of "virtual reality"—or rather simulated reality—technology, progress was made especially with visual systems. In particular, the widening of the field of view (FOV), and resultant increase in stimulation of the peripheral visual system, has created "a more compelling visual display of motion" (McCauley, 1984, p. 9). As we have seen, the advantage of motion observed in the simulator was often reduced with improved visual stimulation. In contrast, the last major advances with regard to motion cuing date back at least 15 years. They include the introduction of critical onset cues followed by subliminal washout, and of "gravity align" platform attitudes to simulate sustained acceleration (Brown et al., 1989). But these innovations still do not overcome the limitations resulting from the fact that simulators are stationary. Boldovici (1992) reports that whereas the state of motion technology for *land-vehicle* simulation is coming along, for *aircraft* simulators (quoting from personal correspondence with Stark), "[t]he problem of cue coordination and the associated cue delay problem are both crucial and severe" (p. 10). The question of interest is whether a state-of-the-art visual system would not simulate airplane motion at least as faithfully on a perceptual level as the inherently limited physical simulator motion does.

Additionally, regional airlines in the U.S. are increasingly interested in the question of whether a Level 6 flight-training device (i.e., a fixed-base simulator) with an enhanced visual system could be employed to satisfy FAA requirements for recurrent training and checking. This would permit airlines now conducting such training in the aircraft to take full advantage of the more comprehensive maneuver-oriented and scenario-based training opportunities available in a simulator. The argument is that this would enhance the overall safety of regional airlines, provided that equivalent safety of training and evaluation with visual motion cues alone can be empirically confirmed.

One *caveat* that needs to be raised here regards simulator sickness. A widely accepted explanation of simulator sickness is that it arises from the sensory conflict resulting from discrepancies between visual and vestibular cues (see, e.g., McCauley, 1984; Oman, 1991). As the quality and, in particular, the FOV of the visual system increase disproportionately compared to the motion system, so will the sensory conflict between visual and vestibular motion cues. Guedry (1987) suggests that this, coupled with an overall increase in simulator use, is one of the main reasons for the increase in reports of simulator sickness over the past decade. McCauley, Hettinger, Sharkey, and Sinacori (1990) cite evidence found by McGuiness, Bouwman, and Forbes (1981), indicating that more experienced pilots may be more susceptible to simulator sickness than novice pilots, just as they may be more likely to rely on vestibular motion cues (Young, 1967). Potentially, then, even if a sophisticated visual system alone were to provide sufficient motion cues for recurrent pilot training in the simulator, forgoing physical motion may still be unacceptable due to the effects of the ensuing sensory conflict on pilots. On the other hand, it has been argued that instead of reducing simulator sickness, motion systems add to simulator sickness. Boldovici (1992) reports that Lintern questions whether cue conflict promotes simulator sickness at all and whether motion can reduce it.[5] He added that a policy used by the U.S. Navy for reducing motion sickness has been to turn off the motion platform.

1.3.2 Research Question

The FAA is revisiting the issue of platform motion in the context of regional airline recurrent pilot training and checking. Given a pilot who is already qualified as a crewmember in the aircraft and who has been serving in revenue service in that aircraft for at least six months, the FAA is interested in obtaining data pertinent to the following questions: First, does the training conducted in a fixed-base simulator with a wide FOV, cross-cockpit-view visual system produce results equivalent to those produced in a like system having platform-motion cuing? Second, from a regulatory perspective, do recurrent proficiency checks conducted in a visually equipped fixed-base simulator verify the line-operational readiness of airline pilots to a degree operationally comparable to a simulator with motion and without compromising the safety of the flying public?

1.3.3 Burden of Proof

It is much easier to prove the *existence* of a requirement for motion than its *non-existence*. A single positive finding would *prove* the need for a requirement. In contrast, any number of negative findings would only *support* that it is unnecessary because the single positive finding could always be just around the corner. Along these lines Boldovici (1992) claims, “Regardless of differing opinions about the causes of null results, the argument for not using motion platforms because results do not demonstrate benefits remains untenable. Null results can ensue from factors other than the absence of differences between motion and no-motion treatments” (p. 6). Thus, it is imperative that every effort be made to find any positive evidence that may exist. If a null result is obtained under this approach, then an inference that no differences exist is more tenable. With this in mind, the research strategy used was biased towards finding an effect of motion. Not only is this good research design, but it also is consistent with the FAA’s need to be biased towards keeping motion for the sake of safety unless a watertight case for change can be made.

1.3.4 Design Strategy

Given the burden of proof every effort was made to find an effect of motion. That is, every aspect of the study was geared towards maximum diagnosticity. Therefore, an FAA qualified simulator with a modern six DOF freedom synergistic motion system in current use for recurrent training was used to compare full motion with no motion at all, i.e., when the motion

---

[5] The original Lintern paper was misreferenced and could not be found in either the year or volume that was listed by Boldovici.

system was turned off and the simulator completely motionless. In addition, it had a state-of-the art wide-angle collimated cross-cockpit visual system to stimulate vection.

The maneuvers were chosen to be as diagnostic of the need for motion as possible, given other constraints of the experiment, among them limiting the duration of the session so that the pilots would not have time to adapt to the simulator. The previous literature suggests that pilots should fly disturbance maneuvers that are asymmetric, closed-loop, unpredictable, and high in thrust, gain, and workload. Maneuvers involving engine failures meet all these criteria. The flight tasks chosen were engine failures on take-off, with a low level of outside-world visual cues encountered in recurrent training and checking (cf. Hall, 1978; 1989).

Subjective and objective data were collected during both training and testing. The subjective data included a grade provided by the instructor for each maneuver as well as opinions from the crew and instructor on control precision, control strategy and technique, gaining proficiency, physical and mental workload, comfort, and acceptability. The objective data was collected from the simulator computer and from measurement devices at a high sampling rate. It included variables measuring stimulation of the pilot (e.g., motion, force feedback, instruments, and visual display), pilot behavior (e.g., control inputs, throttle inputs, and brake pedal inputs), and pilot performance (e.g., ground path control precision and flight path control precision).

In order to evaluate these measurements once they were collected, diagnostic criterion measures were developed. They included the earliest period of the performance envelope in order to be diagnostic of the potential alerting function of motion. They defined the smallest operationally relevant differences so as to provide a way to evaluate whether differences found between the two motion conditions were meaningful. Finally, they defined the acceptable risk of reaching the wrong conclusion. In this case the definition took into account that in the name of maintaining optimal safety, motion should be required unless there is excellent evidence to the contrary. The power of the statistical analysis had to be examined to ensure the reliability of the results. The number of subjects required had to be large enough to detect the smallest operationally relevant effect size even given the naturally occurring variability between subjects.

Along with being as diagnostic as possible, the experiment was designed to minimize any possibility of mistaking spurious findings as effects of motion. Thus, there were no differences between the motion condition and the no-motion condition beyond the presence or absence of motion. Accordingly, a homogeneous pilot sample from the population of interest (i.e., regional airline pilots qualified on the simulated airplane) was used. They were “fresh from the airplane” to ensure that they had not recently been adapted to the simulator. Following standard practice, any uncontrollable variables that may have affected performance, such as time of testing and variations among instructors in grading criteria, were at least counterbalanced if not eliminated across the two conditions. Finally, the equipment was carefully calibrated at the beginning and end of the experiment to demonstrate that simulator performance was the same in both motion conditions. To catch any intermediate drift, there were abbreviated daily calibrations (e.g., a visual comparison with the initial motion calibration and a subjective inspection of the visual system).

To assess the simulator as a training tool, a forward-transfer paradigm is desirable. This would measure how well training in the simulator transfers to the airplane. To assess the simulator as an evaluation tool, a reverse-transfer paradigm is desirable (Cross, 1991). This would measure how well pilots’ proficiency in the airplane is reflected in their performance and behavior in the simulator.

As already mentioned, however, experiments in a real airplane are dangerous, impossible to control, and costly. Simulators, on the other hand, are safe and controllable. In addition, the use of high-fidelity simulators (Level C/D) for total training and checking for qualified pilots for almost twenty years (FAR, Part 121, Appendix H, 1980) validates such a simulator as a stand-in for the airplane. Thus, both the forward-transfer and reverse-transfer studies used a quasi structure. That is, in the quasi-forward transfer study, pilots were trained in the simulator, with or without motion, and then tested in the simulator with motion. The assumption was that the simulator-training configuration that produced the best results during testing provides the best training for airplane flying as well. In the quasi-reverse transfer study because pilots cannot be originally evaluated in the airplane, a sample of homogenous, experienced pilots were evaluated in the simulator with and without motion. The assumption here was that if motion has an effect on how accurately real-life pilot skills are reflected in the simulator, there would be a difference in performance and/or behavior between the two groups. Combining these two approaches, quasi-forward transfer and quasi-reverse transfer, could strengthen the validity of results, provided that they are in agreement. Furthermore, Boldovici (1992) makes a good argument for a well-done quasi-transfer experiment: “Those who insist that transfer to parent equipment is the only legitimate test of simulator effectiveness may be accepting unreliable test results that cannot be valid while rejecting reliable test results that may be valid” (p. 5).

2. METHOD

2.1 FLIGHT SIMULATOR

The flight simulator used in this experiment provided a complete flight environment except for air traffic control (ATC) commands that had to be provided by the instructor/evaluator (I/E), as is usually the case in flight simulation (Bürki-Cohen, Kendra, Kanki, & Lee, in press). It received original FAA Level C qualification in early 1995. It has been maintained to retain continuing qualification at Level C.

The simulated flight deck is a complete replica of the airplane flight station. It is arranged and programmed to represent a thirty passengers plus three crewmembers twin-engine turboprop airplane. The engines are mounted on the wings and the propellers are counter-rotating, hence it has no critical engine. Each engine has 1,650 shaft horsepower. The maximum take-off weight of the airplane is 24,000 pounds. The mathematical simulation model was developed by the airplane manufacturer and was implemented by the simulator manufacturer. Other important major components of the simulator include the following.

Table 2.1 Flight Simulator Information

Host Computer	Harris Night Hawk 4400
Control Loading	Electric Digital Control Loading System
Airplane systems	All airplane systems are represented and, from a pilot operator perspective, operate as they do in the airplane.

2.1.1 Visual Cuing

Visual cuing was provided by the IVEX VDS 2000 image generation system with three channels displayed on a SEOS Panorama display system. The system provided a projected collimated image with a continuous field of view of 150 degrees horizontally and 40 degrees vertically.

2.1.2 Motion System

The flight simulator motion system was a six DOF synergistic system utilizing hydraulically actuated legs capable of a 60-inch stroke. The expected and actual measured performances of the system are shown in Table 2.2 for the heave mode. As can be seen, the demonstrated performance of the system is not as good as that specified by the manufacturer.

Table 2.2 Motion System Performance – Heave Mode

Ref. Freq.	Measured Simulator Response (average of legs)		Manufacturer's Specification for System (expected performance)		Data Typical of Several Modern Simulators[6]	
Freq. (Hz)	Amplitude Attenuation (db)	Phase Shift (degrees)	Max Amplitude Attenuation (db)	Max Phase Shift (degrees)	Amplitude Attenuation (db)	Phase Shift (degrees)
0.1	-.086	-5.5	-1.0	-15	0	0
0.5	-1.18	-36	-1.0	-15	0.03	-3.0
1.0			-3.0	-40	0.4	-9
1.07	-3.90	-72				
1.5	-6.02	-76			1.0	-15
1.7			-5.0	-70		
2.0	-6.41	-101			1.5	-23
2.5					1.5	-29
3.0	11.18	-135	-8.0	-110	1.0	-35
4.0					1.0	-45
5.0					0.5	-55
6.0					0.5	-70
7.0					0.0	-82
8.0					-1.0	-105
9.0					-3.0	-128

The data in the above table indicate that the bandwidth in the heave mode for the simulator used in this experiment was on the order of 1.7 hertz. Bandwidth is the frequency at

[6] Values considered "typical" were selected after reviewing several data sets of new or nearly new flight simulators. These data do not represent any particular simulator.

[7] The system performance is determined by measuring its response to a sinusoidal input. The sinusoidal input is varied in order to observe the response at several different frequencies so that magnitude and phase relative to the input signal can be determined. Bandwidth is used here as a measure of system performance as it is frequently used for that purpose in the specification or description of closed loop linear systems. As used in this case, bandwidth identifies the frequency at which the response of the motion system lags the input signal by 90 degrees. Except at low frequencies, the magnitude of the response is usually different from the input as well and is usually quoted as a ratio of the magnitude of the input. The ratio is stated in units of decibels (sometimes it may be simply stated as a quotient of the output divided by the input). The magnitude ratio, phase and bandwidth of the response constitute the necessary and unique statement of system performance.

which the response (output) lags the input by 90 degrees (phase lag). The specification for this simulator indicated an intended bandwidth between 2.0 and 2.5 hertz. Data reviewed for several other systems, however, indicated bandwidths of approximately 7 hertz. The bandwidth of this simulator did, however, include frequencies normally associated with human response times, but did not include some higher frequency vibrations such as airframe buffets.

Simulator motion performance was measured by recording the response at each of the reference frequencies listed in the left column of Table 2.2. The recorded output was then compared to the sinusoidal input signal to determine the gain and phase relationship of the output to the input. No frequency response automatic test equipment was available. Consequently, the frequency response results were measured and calculated using the graphical recorded output. The frequency response measurement was done before and after the data runs. It must be noted that response measurements were done in only the heave mode. The response in other modes may be different from that measured for this example case.

2.1.3 Aural Cuing

In the experiment, the sound level in the simulator was set at 65%, which represents 100% airplane sound. This was intended to provide a realistic aural cue to the pilots. It is well known that this level is often reduced during training to enable communications and reduce fatigue.

2.1.4 System Calibration

Before collecting data, it was necessary to ensure that the simulator met its original qualification standards initially and that there was no day-to-day change during the data acquisition period. To meet this objective, the FAA Qualification Test Guide (QTG), the test document required for FAA qualification of the simulator, was accomplished completely before the data acquisition began and again after the data acquisition was completed to ensure that the simulator did not change during the time period of the experiment. The simulator complied with all requisite FAA requirements as demonstrated by the 119 tests in the QTG.

Additionally, daily calibration tests were devised to assure that there was no variation in the simulator performance or characteristics day to day. Before conducting any data runs in the simulator, the daily calibration was accomplished by inserting a pulse input into the pitch, roll, and yaw channels of the simulator and outputting eight characteristic responses. The responses were compared to the original run by the technician on duty to ensure that there was no change in the output. The original run was done immediately after the complete FAA QTG was accomplished. Recordings of the visual and motion transport delays were done as the daily calibration runs were done.

The Entran Devices Model EGCS3-A-2 three-axis accelerometer, which was mounted near the pilot station, was calibrated by the manufacturer’s recommended method. The 2 g ± 0.01 g accuracy of the device was verified in each axis.

2.2 PARTICIPANTS

All participants were employed by the same regional airline and were qualified for either flying or instructing/evaluating performance on the specific type of turboprop airplane used. The goal was to use pilots and I/Es that were as homogeneous as possible, and to counterbalance any lack of homogeneity across the two motion groups.

2.2.1 Pilots Flying and Pilots Not Flying

The first nine crews (5 Motion crews and 4 No-Motion crews) consisted of Captains of that particular airplane who had just completed six months of revenue service. Each pair of pilots was tested immediately before their six-month LOFT (Line Oriented Flight Training) and therefore neither pilot had been in the simulator for the previous six months. One of the two Captains was randomly chosen to be the Pilot Flying (PF) and the other to be the Pilot Not Flying (PNF) for the duration of the experimental session.

For all of the other crews (16 Motion crews and 21 No-Motion crews), the PFs were Captains with at least 12 months of revenue service.[8] They were tested immediately before their regularly scheduled maneuver validation and had not been in the simulator for the 12 months prior to their participation in the experiment. The PNFs were either Captains or First Officers. Each pilot participated one time only. The sessions lasted an average of 39 minutes, and all crews from both groups were tested during normal waking hours (i.e., during the day or evening), thus avoiding any group differences due to fatigue.

Equipment failures, data recording failures, and human errors led to the loss of some data. Forty-six crews participated (Motion: 21, No-Motion: 25), but only 42 crews (Motion: 20, No-Motion: 22) provided any useable data and only 23 crews (Motion: 14, No-Motion: 9) provided complete data sets (see Tables 2.3 and 2.4 for details on how many crews completed each type of maneuver and questionnaire).

[8] There were actually two additional ways in which the treatments given to the first nine crews differed from the subsequent crews. One involved the pattern of the side of engine failure, the other the outside temperature (both to be described in more detail later). None of the three differences affected the results. The difference in engine failure side pattern should not have mattered, whereas both the experience and the temperature differences could theoretically have resulted in lower grades for the initial pilots compared to the subsequent pilots (the temperature used with the initial crews could have caused some crews to select an incorrect procedure and thus to receive lower grades). However, the difference was in the wrong direction and was not significant (Fisher Exact probability, all $p > .17$). Additionally, the Motion group/No-Motion group break-down was fairly even for both initial crews and subsequent crews (initial: 5 and 4; subsequent: 15 and 18—these counts include the crews that provided at least some useable data, but not necessarily useable grades data). Thus, the data from the initial and subsequent crews were analyzed together.

Table 2.3 Number of Crews Providing Data for Each Maneuver

		Grades		Objective Data	
		Motion	No-Motion	Motion	No-Motion
Completed Maneuvers	**First Look**				
	V_1 cuts	19	19	18	19
	RTOs	19	19	16	14
	Transfer				
	V_1 cuts	16	16	16	18
	RTOs	16	16	16	16
	Additional Training and Normal Take-Offs	77	71	75	83
Complete Crews		16	16	14	10

Table 2.4 Number of Crews Completing Each Questionnaire

	Motion Crews	No-Motion Crews
After Training	19	19
After Transfer	16	18

The average flight hours for PFs and PNFs are listed in Table 2.5. With a minimum of 3,000 flight hours, the PFs far exceeded the hours required for an airline transport pilot's license (1,500 hours). Similarly, with a minimum of 1,200 hours, the PNFs far exceeded the hours minimally required for a commercial pilot's license (250 hours). No differences in flight hours between the Motion and No-Motion groups were found for either the PF or the PNF positions when considering either all 42 crews providing any data or only the 23 crews providing complete data sets (all $t < 1.3$, all $p > .21$).

Table 2.5 Average Flight Hours

	Motion	No-Motion
Crews Providing Any Useable Data		
PF	7,221	7,845
PNF	3,956	4,260
Crews Providing Complete Data Sets		
PF	7,121	6,000
PNF	4,085	4,694

2.2.2 Instructor/Evaluators

Fourteen different I/Es participated in this experiment. Each tested between one and nine crews. In order to avoid group differences due to individual differences between the I/Es, crews were to be counterbalanced across motion groups so that the difference between the number of crews tested by an individual I/E in the Motion and No-Motion groups would not exceed one. However, in the end the difference for three of the 14 I/Es was two crews and the difference for one I/E was three crews. This would only represent a problem if any of these four I/Es were to be extremely low or high graders. This was not the case for three of the I/Es: Their mean grades were within one standard deviation of the overall mean grade (averaged across all I/Es). The mean grade for the fourth I/E was low, but still within two standard deviations of the mean. Because this I/E tested more No-Motion than Motion crews (3 vs. 1), a disadvantage of the No-Motion group could be explained to a small part by this difference. As will be seen later, this was not an issue.

2.3 MANEUVERS

Test maneuvers had to maximize satisfaction of criteria mentioned in the literature as diagnostic for detection of a motion requirement, while minimally interrupting the host airline's training and evaluation program. These criteria included the following:

1) closed loop, to allow for motion to be part of the control feedback loop to the pilot;
2) unpredictable and asymmetric disturbance, to highlight an early alerting function of motion;
3) high gain and high thrust, to magnify any motion effects;
4) high workload, to increase the need for redundant cues such as those provided by motion, out-the-window view, instruments and sound; and
5) short duration, to prevent pilots from adjusting to a lack of cues.

Engine failures on take-off with either rejected (RTO) or continued (V_1 cut) take-off were deemed as fulfilling most of these criteria. The engine failures occurred, to ensure an RTO, at ten knots below V_1 (i.e., 95 KIAS) and, to induce pilots to continue the take-off, at five knots above V_1 (i.e., 110 KIAS).

2.3.1 Airport, Weather, and Aircraft Information

The full airport, weather, and aircraft information for the conditions used in this study can be found in Appendix A. All conditions were kept constant throughout the study, with the exception of outside temperature, which differed for the first nine crews (see below).

Each maneuver started in take-off position at the University Park Airport, State College, PA. The same airport database was always used. The runway length in that database was 5,000 feet.

A dusk time of day with low clouds (indefinite ceiling) and low visibility (one quarter mile) was simulated to increase reliance on physical motion. A 10 knots crosswind at 90 degree from the left increased the difficulty of control.

The airplane was configured at a relatively lightweight and aft center of gravity (cg) in a cool atmosphere in order to permit a larger dynamic response of the simulated airplane to an engine failure. In consultation with supervisory personnel of the host airline operating the airplane, chosen weight and cg values were normal, or at least not unusual, in routine operations. The take-off related speeds are a function of weight and environmental conditions. A temperature of 55 degrees Fahrenheit was chosen as the lowest temperature practicable that would avoid any ambiguity regarding the take-off configuration, given all other variables. This was recognized after the temperature of 40 degrees chosen for the first nine crews led to uncertainty within some crews and I/Es on whether the take-off configuration should have reflected the possibility of ice accumulation in the visible moisture environment (as required by a recent procedural change within the host airline) (for details on why this did not affect grades, see Footnote 8).

2.4 EXPERIMENTAL DESIGN

Two experiments were combined into one experimental session in order to minimize the disruption to the host airline's training and evaluation program, as well as to reduce pilot adaptation to a simulator configuration. Both experiments investigated the need for platform motion in simulators, focusing on different functions of the simulator. The first experiment, First Look evaluation, examined the use of simulators as *evaluation* tools of pilots' aviating skills. In other words, it assessed the degree to which a pilot's existing skills transferred *from* the airplane *to* the simulator, and whether this was affected by the motion state of the simulator. This assessment needed to occur during the very initial exposure of the crew to the simulator, so that pilots' behavior and performance would reflect their actual skills in the airplane with as little contamination as possible from potential adaptation to a particular simulator configuration. The second experiment, Training and Transfer testing, examined the use of simulators as *training* tools for aviating skills, skills that would eventually need to be transferred to the airplane. That is, the experiment assessed the degree to which motion affected the training of skills and, most importantly, the transfer of those skills to the airplane. This means that there were three major periods of the combined experiment: First Look evaluation, Training, and Transfer testing.

2.4.1 First Look Evaluation

Ideally for First Look evaluation, the crews would be evaluated in the airplane and then again in the simulator, some crews with motion and some without. Then it could be determined whether evaluation in the simulator with or without motion better matched evaluation in the airplane. However, in addition to the danger and expense of flying engine failures in the airplane, it is also impossible to keep conditions constant across the two groups (e.g., weather, traffic, ATC vectors, etc.). Thus, there was no airplane evaluation, with the assumption that the simulator configuration that elicited the better performance with lower workload was the one that more accurately represented the airplane than the other. In order to support this assumption, a homogeneous group of experienced Captains who were at a similar stage in their career and who showed no differences in flight hours across the two motion groups was chosen.

2.4.2 Training and Transfer Testing

For Training and for Transfer testing, the crews were trained in the simulator either with or without motion and then, ideally, would have been tested in the airplane to see which simulator configuration resulted in better behavior and performance in the air. However, the same safety, experimental control, and cost concerns that led to First Look evaluation in the simulator instead of in the airplane also forced the assessment of the effect of motion on Training and Transfer to be done in a simulator with motion—as a stand-in for the airplane.

Using the highest-level configuration of the simulator as a stand-in for the airplane can be justified for two reasons. First, this procedure (often called "quasi-transfer") has been used for this purpose many times in research (see Introduction). Second, the FAA has been allowing the use of a Level C simulator for 100% recurrent training and evaluation of qualified airline pilots for nearly two decades, with no apparent loss in safety (FAR, Part 121, Appendix H, 1980). The success of this substitution over many years of training and evaluation supports the use of the simulator as a stand-in for the airplane in quasi-transfer experiments.

2.4.3 Structure of Experiment

Thus, the combined experiment had the following format. A PF and a PNF flew the simulator with an I/E observing and evaluating. The crew first flew one V_1 cut followed by one RTO, either with or without motion. The pilots did not know that the engine failures would occur. This constituted both the First Look evaluation and also the initial Training for the Training period of the experiment. Then, if the crew did not perform adequately on either maneuver, the pilots received one or two additional Training trials for that maneuver, first any additional RTOs and then any additional V_1 cuts, with the motion platform set the same as before. Because these maneuvers were for Training and because training is often enhanced if crews know what to expect, the pilots were informed about the engine failures for these trials. This concluded the Training period of the experiment. Both pilots and the I/E then filled out a questionnaire. To prevent the pilots from guessing which maneuvers were to come during the final testing, they were given two normal take-offs without being informed about the lack of engine failures, with the motion platform still in its original configuration. Finally, with motion on for all crews, the pilots flew one more V_1 cut followed by one RTO, but were not told about the upcoming engine failures. This constituted the Transfer testing period of the experiment. When this was completed, both pilots and the I/E completed one final brief questionnaire. In summary, there were two groups of crews. The crews in the Motion group had motion on during First Look, Training and Transfer, whereas the crews in the No-Motion group had motion off during First Look and Training but on during Transfer. Collectively these two groups were called the two motion groups because they differed with regard to motion state.

2.4.4 Maneuver Sequence

Given the sequence of maneuvers just described, the only room for variation was in how many RTOs and V_1 cuts each crew had during Training. This was determined by the crew's behavior and performance as evaluated by the I/Es in the form of grades given at the completion of each maneuver. The choices for grades were 1 (unsatisfactory), 2 [FAA Practical Test

Standards (PTS) (FAA, 1995)], 3 (company standards), or 4 (excellent). During the Training period of the experiment, a grade of 1 or 2 resulted in the crew having to repeat that maneuver, for up to a total of three Training maneuvers of each type. Thus, the total number of maneuvers including normal take-offs each crew performed ranged from a minimum of 6 to a maximum of 10.

2.4.5 Engine Failure Side

The side for each engine failure followed a roughly alternating order. Assuming that the crews did all 10 possible maneuvers, half of the crews would have had the order L, R, L, R, R, L, (two normal take-offs), R, L and the other half would have had the inverse order (i.e., R, L, R, L, L, R, (two normal take-offs), L, R). If any maneuvers were omitted, then the engine failure for that maneuver was omitted as well and the next engine failure used the side given next. This sequence had the advantage of having either an equal number of engine failures on each side or only one more on one side than the other, no matter which combinations of maneuvers were run. It also had the property that either side failed at most three times in a row, given any possible sequencing of maneuvers. However, for the first nine crews, a programming error in the host computer caused a slight change in this sequence. It was either R, R, L, R, L, R, (two normal take-offs), L, L or the inverse. This sequencing allows for the possibility that a crew could have had four engine failures in a row on the same side with only two failures on the other side, and in fact, this did happen to one crew. There is no reason, however, to assume that this slight change in counterbalancing engine failure side would have affected the overall results, given that overall variability of failure side was maintained (see Footnote 8).

2.4.6 Motion Group Assignment

The motion assignment for each crew was determined by the following string of decisions. If the I/E for the current crew had tested one more crew from one motion group than the other, the crew was assigned to the motion group with fewer crews tested by this I/E. If the I/E had tested an equal number of crews in each motion group, the crew was assigned to the motion group with fewer crews overall. If the two groups had the same number of crews overall, the current crew was assigned to the Motion group for the first 21 crews (13 Motion and 8 No-Motion) and then to the No-Motion group for the next 21 crews (7 Motion and 14 No-Motion). This resulted in more Motion crews being tested during the first part of the experiment and more No-Motion crews in the second part, which should not have mattered because simulator variables were proven not to have drifted during the data collection period. No other factors potentially affecting pilot performance or I/E grading criteria are known to have shifted during this period.

2.5 EXPERIMENT CONTROL AND RECORDING OF GRADES

A laptop computer was programmed to control the simulator and to record events with minimal I/E intervention, eliminating the need for the presence of an experimenter that might have contaminated the regular training/evaluation environment. The laptop, thus, enabled the I/E to focus on the behavior and performance of the crew. Even more importantly, it eliminated any need to inform the I/E (or the crew) of the interest in motion and the motion state of the simulator for each maneuver, thus minimizing any bias.

Once the I/E had entered his identification, at the start of a session, the laptop assigned the crew to either of the two motion conditions, attempting to counterbalance I/Es across conditions while keeping the numbers in each group even (see above). This minimized any potential differences resulting from variability across I/Es (or across an individual I/E on different days).

After the I/E entered that he and the crew were ready, the laptop triggered initialization of the simulator. Initialization included the rising of the motion platform regardless of whether the motion logic would be enabled, the setting of the motion logic state, the positioning on the runway, the setting up of the environment (airport, whether, airplane variables), etc.

The laptop then presented the I/E with briefing information for the first maneuver (always a V_1 cut), including instructions on what information was to be shared with the pilots. After the I/E indicated readiness, the laptop commanded the simulator to preset the malfunction (engine side and speed) and to unfreeze. Data recording from the simulator started as soon as the speed exceeded a low value and terminated when preset maneuver end criteria were reached. For V1 cuts, the termination point was 10 seconds after the low pressure (LP) shutoff valve was actuated (the LP shutoff valve is the last item on the engine failure memory checklist); for RTOs, it was when a full stop was reached; and for normal take-offs, it was when a crew reached 500 feet.

As soon as the I/E indicated that he was ready to grade, the laptop commanded the simulator to freeze and presented the I/E with a grading screen. The next maneuver was determined based on a preprogrammed sequence and the I/E grade. After reaching the training criterion or maximum number of allowable trials, the laptop cued the I/E to hand out questionnaires before going on to the final test phase, which was also followed by a questionnaire. All events were time stamped and recorded, including maneuver type, engine failure side, and grade.

2.6 INSTRUCTOR/EVALUATOR AND PILOT BRIEFINGS

The pilots and I/Es were briefed at several different times. Each briefing explained that the focus of the experiment was the evaluation of simulators, not pilots, and therefore that pilot performance was being used as a measure of the simulator, not of the pilots. Pilots were assured, for example, that this was a non-jeopardy event for them and that any bad performance would provide useful information about the simulator without reflecting on the pilot.

The first briefing was for the I/Es only and was presented orally to a group of I/Es before they tested any pilots for this experiment. Their role in the experiment was described and the laptop was demonstrated, which also gave them all maneuver and airport, weather, and airplane configuration information. However, they were not told about the interest in motion or the motion manipulation of the simulator.

The second briefing was given to I/Es and pilots in written form before each session. It described generally what would be happening during the session and, like the oral briefing, did not discuss the role of motion at all. The pilots were not informed about the kinds of maneuvers they would be flying (see Appendix B and C for the full briefings[9]). Finally, before each maneuver the I/Es were given (in writing on the laptop) the environmental conditions (e.g., airport information, weather, and airplane information), which they were to convey to the pilots.

[9] Any references to specific airlines and training facilities have been deleted and replaced with general descriptions.

They were also given maneuver information, which they were instructed to share with the pilots only during the Training sessions, so that they could focus their attention on the relevant performance variables during First Look evaluation and Transfer testing.

If any pilot had chosen to withdraw from the study, he or she would have been given a withdrawal form (see Appendix D); however, no pilots chose to do so.

2.7 QUESTIONNAIRES

The pilots and I/Es were given a detailed questionnaire at the end of the Training period, which was before all pilots had experienced motion. An abbreviated version of the same questionnaire was administered at the completion of all maneuvers, after all pilots had experienced motion (see Appendix E for complete questionnaires[10]). Pilots and I/Es were asked about control precision/performance, control strategy and technique, physical and mental workload, ease of gaining proficiency, and comfort (absence of nausea and disorientation). In addition, the two pilots only were asked about the acceptability of the simulator. Because neither the PNF nor the I/E were actually flying the simulator, they were asked to answer questions relating to aircraft control, proficiency gain, and workload about the PF.

The full questionnaire asked some of the questions in more detail. For example, the control precision/performance question asked about various controlled variables (e.g., altitude control, heading control, etc.) and the control strategy and technique question asked about various controls (e.g., rudder inputs, aileron inputs, etc.). Workload was divided in mental and physical workload.

Another consideration was to what standard to compare the variables of interest. For control precision/performance, control strategy and technique, and workload, the pilots were asked to compare flying these maneuvers in the simulator with flying the same maneuvers in the airplane. Because they may never have experienced engine failures in the airplane, they were asked to imagine what it would be like to do so based on their extensive flying of other maneuvers in the airplane. For gaining proficiency, comfort, and simulator acceptability the pilots were asked to compare flying the maneuvers in the simulator during that session to flying them the last time they were in the simulator.

For I/Es, the comparisons were as follows. First, they were asked to give a grade to the crew for performance in controlling different variables. Then, they were asked to compare PFs to an average PF for control strategy and technique, how much physical and mental workload they appeared to experience, and how hard or easy it was for them to gain proficiency. Finally, they were asked to compare their comfort that day to the comfort they normally felt during the 12-month maneuver validations.

Each question was arranged so that the respondent simply had to check a box indicating a rating ranging from 1 to 5. A rating of 3 indicated that the simulator was the same as the airplane or the simulator last time, whereas a rating of 1 and 5 indicated that the simulator this time was much worse or better, respectively. The specifics of the type of scale (e.g., whether from worse to better or harder to easier, etc.) depended on the question and can be seen in Appendix E. The only exception was the I/Es' ratings of performance, which indicated a grade of either unsatisfactory (1), FAA Practical Test Standards (2), company standards (3), or excellent (4) (see Appendix E). It took the respondents, on average, 6.8 minutes to fill out the full questionnaire.

---

[10] The type of simulator used for this experiment has been omitted in this report.

The abbreviated questionnaire administered at the end of the experiment allowed a more condensed grouping of the questions. Also, the respondents were already familiar with the questions and comparisons. It therefore only took, on average, 1.3 minutes to complete this questionnaire.

2.8 OBJECTIVE DATA RECORDING

Parameters to assess simulator performance (i.e., fidelity of stimulation of the pilot by the simulator), pilot performance, and pilot behavior/workload (i.e., pilot control inputs) were sensed directly from the simulator systems. Additionally, a three-axis accelerometer and a two-axis rate gyroscope were installed to acquire additional information on the activity of the motion system. These instruments were attached to the motion base directly below the pilot station. The accelerometer output was a direct indication of the accelerations experienced laterally, vertically and longitudinally by the subject pilot. The rate information was used only as a secondary reference to aid understanding of the motions experienced in the simulator.

All objective data were recorded on a data acquisition computer (DAC). The DAC, a personal computer, was installed especially for the purpose of this experiment. The data were sampled and recorded at a rate of 50 samples per second. Seventy-eight parameters were recorded in this manner (see Appendix F).

3. RESULTS

Both subjective and objective data were collected. I/Es graded the PFs for each maneuver flown. In addition, PFs, PNFs and I/Es filled out questionnaires after the Training phase and again after Transfer testing. Objective data were recorded directly from the simulator and covered both simulator and pilot performance and behavior. The main interest was always whether there was a statistically significant difference between the Motion and the No-Motion groups.

In order to facilitate the reading of this section, a review of the sequencing of events is provided.

1) First Look (and first Training): The crews did one V_1 cut followed by one RTO (motion on or off depending on group).
2) Training: Any additional Training on RTOs came next, as needed, followed by any additional Training on V_1 cuts, as needed (at most there were two additional Training trials of each kind, motion on or off depending on group).
3) All participants filled out the first questionnaire.
4) The crews did two normal take-offs (motion on or off depending on group).
5) Transfer: The crews did one last V_1 cut followed by one last RTO (motion on for all crews).
6) All participants then completed the final questionnaire.

Thus, the crews in the Motion group had motion on during First Look, Training, and Transfer, whereas the crews in the No-Motion group had motion off during First Look and Training but on during Transfer.

3.1 GRADE AND QUESTIONNAIRE RESULTS

3.1.1 Grades

The analyses of the grades were used to examine whether I/Es perceived any difference in performance between the two groups for First Look evaluation (when only one group had motion), and for the Transfer of training test (when both groups had motion). The effect of motion on Training progress was also examined. Additionally, the effect of transitioning from the last Training trial to the Transfer of training test was examined. Finally, I/Es perception of performance for normal take-offs was analyzed, although these take-offs served only to reduce the crews' anticipation of failures during the Transfer of training trials. Thus, the analysis of grades was divided into four parts. First, the grades during First Look evaluation and Transfer trials were examined. Second, the improvement in grades across the Training period was examined. Third, the relationship between the grades on the last Training trial and the grades during Transfer was examined. Fourth, the grades of the normal take-offs were examined. A summary of the results of all the analyses described below is provided in Table 3.8.

The two maneuvers, RTOs and V_1 cuts, were always analyzed separately, using always the same analysis sequence and procedures described below. First, a group analysis showed whether there were any overall differences between groups, whether there was improvement across trials (e.g., between First Look and Transfer), and whether improvement depended on motion group. For this analysis, the mean grades for the relevant trials for each motion group were compared using a 2 x 2 (two motion groups by two types of trials) analysis of variance (ANOVA).

Second, individual analyses were done to confirm any findings from the group analyses. The reason for this is that any significant effects found by a group analysis could be due to a small number of crews having very extreme scores, instead of an effect carried by all (or most) crews. The individual crews were studied in two different ways. First, the number of crews in each group getting low grades (i.e., 1 or 2) was compared to the number of crews in each group getting high grades (i.e., 3 or 4) to confirm the (lack of a) main effect of motion group from the group analysis. Second, the number of crews to improve across the relevant trials was counted *separately for each motion group*, in order to confirm the (lack of an) interaction between motion group and improvement found in the group analysis. Third, the number of crews to improve across the relevant trials was counted, *combining across the two motion groups*, in order to confirm the (lack of a) main effect of improvement found in the group analysis. These data were analyzed nonparametrically, with most of them being represented by frequencies in a 2 x 2 contingency table (e.g., motion group by low/high grade) and evaluated with a Fisher Exact probability test (Siegel, 1956).[11] In some cases, however, there were only two numbers to be compared (e.g., when the interest was overall improvement regardless of motion group membership), and these were analyzed with a Chi-Square test.

It could be argued that the analyses of low/high grades should have compared grades of 1 with higher grades because only grades of 1 represent performance below FAA standards. Consequently, all low/high analyses were done in both ways. Using the one vs. higher comparison, there was only one case in which a significant difference between motion groups was found. This case is reported in the appropriate section; but otherwise, to simplify the description, only the analyses using the median split (i.e., grades of one and two vs. three and four) are reported.

First Look and Transfer

Group Analyses

The ANOVAs comparing the grades for First Look evaluation and Transfer across the two motion groups found no differences between the Motion and No-Motion groups for both RTOs and V_1 cuts (all F's ≤ 1, all p values $\geq .33$; see Figure 3.1[12]).

[11] The Chi-Square test was considered, but rejected because in many cases the n was too small for that test. Because the Fisher Exact probability test had to be used in many cases, it was used in all cases to facilitate comparison of different analyses.

[12] The decrease in the number of crews from 19 at First Look to 16 at Transfer was due to loss of some crews due to technical problems during Training.

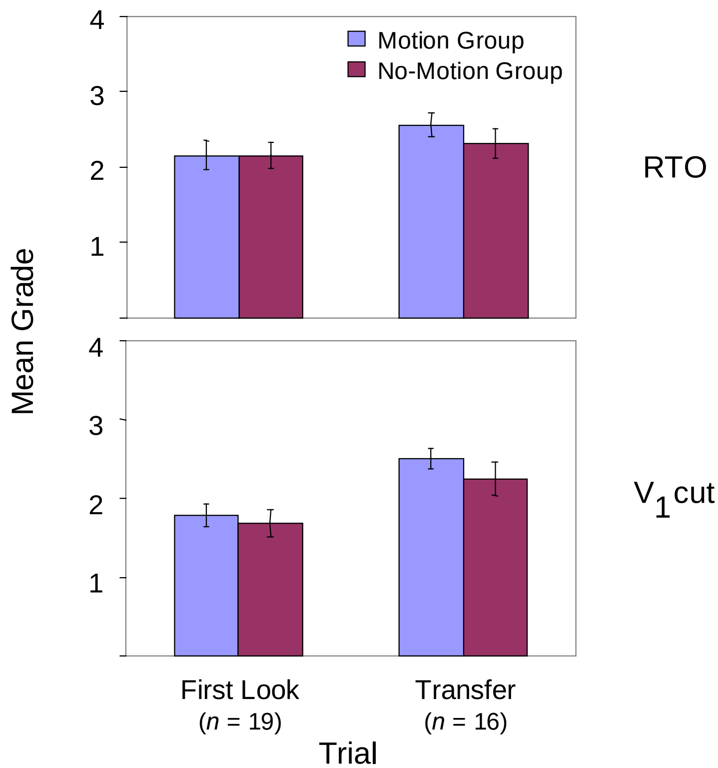


Figure 3.1 Mean grade as a function of maneuver, trial, and motion group (Error bars indicate standard error).

The group analyses did show that the crews' performance for V_1 cuts improved from First Look to Transfer ($F(1, 30) = 9.79$, $p = .004$), but that there was no change for RTOs ($F(1, 30) = .93$, $p = .34$). Importantly, improvement did not depend on motion group for either RTOs (F(1, 30 = .41, $p = .53$) or V_1 cuts ($F(1, 30) = .30$, $p = .59$).

Individual Analyses

<u>Motion vs. No-Motion.</u> Next, the individual analyses were performed. The distribution of grades at both First Look evaluation and Transfer testing is shown in Figure 3.2. To determine whether the analysis of individual crews supports the group findings of no difference between the two motion groups at both First Look (when only the Motion group had motion) and Transfer (when both groups had motion), the number of low vs. high grades for the two motion groups were compared (see Table 3.1). In addition, the number of crews who improved from the First Look trial to the Transfer trial was compared for the two groups (see Table 3.2). No significant differences were found for either maneuver.

However, the only grade analysis that did find a motion group difference occurred here. It was the analysis using the one vs. higher grade comparison (i.e., instead of the median split comparison). The two motion groups did differ on performance of V_1 cuts at Transfer ($N = 32$, Fisher Exact $p = .05$). The crews who were trained without motion received more grades of 1 than the crews who were trained with motion, even though all crews had motion at the time of evaluation. Figure 3.2 reveals that both groups received an equal number of grades of 3 and neither group received any grades of 4. This shows that the difference was all in the number of grades of 1 and 2. That is, the No-Motion group received more grades of 1 whereas the Motion group received more grades of 2, but both groups received an equal number of higher grades.

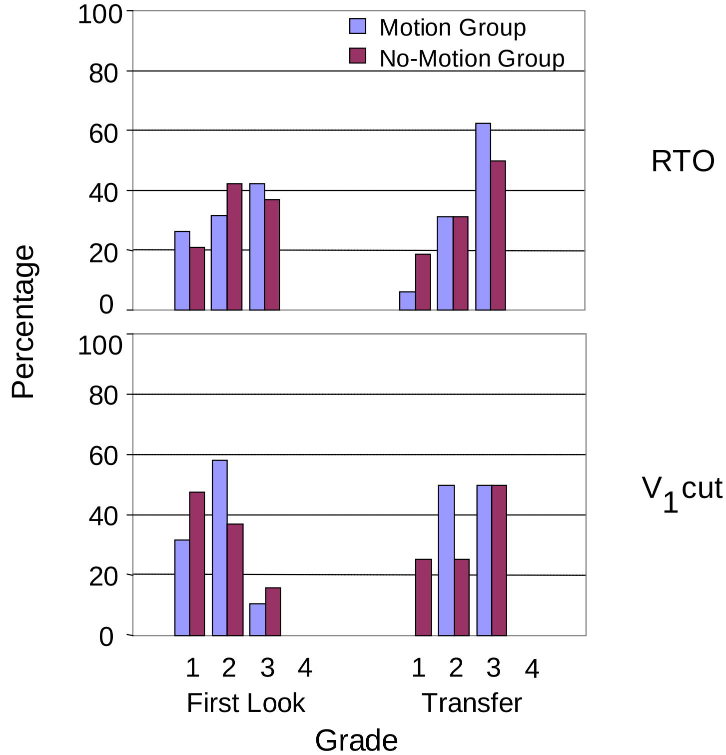


Figure 3.2 Percentage of grades in each grading category as a function of maneuver, trial, and motion group.

Table 3.1 Fisher Exact Probability Values for Test of Motion Group Differences in the Number of Low (1, 2) and High (3, 4) Grades for Each Maneuver and Trial

Maneuver	**Trial**	**Motion Group**		**Fisher Exact Probability**
		Motion Group	**No-Motion Group**	
RTO	First Look	11 low, 8 high	12 low, 7 high	$N = 38$, $p > .25$
	Transfer	6 low, 10 high	8 low, 8 high	$N = 32$, $p > .22$
V_1 Cut	First Look	17 low, 2 high	16 low, 3 high	$N = 38$, $p = .5$
	Transfer	8 low, 8 high	8 low, 8 high	$N = 32$, $p > .28$

Table 3.2 Fisher Exact Probability Values for Test of Motion Group Differences in Improvement From First Look to Transfer for Each Maneuver

Maneuver	Motion Group[13]		Fisher Exact Probability
	Motion Group	No-Motion Group	
RTO	8 improved, 4 got worse	7 improved, 4 got worse	$N = 23$, $p = .61$
V_1 Cut	9 improved, 2 got worse	9 improved, 2 got worse	$N = 22$, $p = .71$

Improvement. To determine if the individual analyses support the finding of improvement from First Look to Transfer for V_1 cuts, the number of crews who improved between those two trials was compared to the number of crews who deteriorated, combining across the two motion groups (i.e., instead of comparing whether the number of Motion crews who improved differed from the number of No-Motion crews who improved). Because these analyses were specifically comparing the grade each crew received on First Look with the grade that that crew received on Transfer, only the crews who actually did both trials could be included (some crews did not participate in the Transfer trials due to technical problems). Table 3.3 shows that, just as with the group analyses, the individual analyses support that both the Motion crews and the No-Motion crews improved on V_1 cuts from the First Look trial to the Transfer trial, but not on RTOs.

Table 3.3 Probability Values for Test of Improvement Between First Look and Transfer for Each Maneuver

Maneuver	Motion Groups Combined[14, 15]	Chi Square and Probability
RTO	15 improved, 8 got worse	$\chi^2(1, N = 23) = 2.13$, $p = .14$
V_1 Cut	18 improved, 4 got worse	$\chi^2(1, N = 22) = 8.91$, $p = .003$

Training Progress

In order to determine if the Motion and No-Motion groups differed in how well they progressed during Training, the crews' grades on their first Training trial were compared to their grades on their last Training trial. Only the crews who had had at least two Training trials were included in these analyses. Thus, only a subset of the First Look trials examined in the previous section is included here as first Training trials. For a First Look trial to be also a first Training trial, the crew must have also had at least one other Training trial. This means that when a crew achieved a grade of three or higher during First Look and thus did not require further training, it was excluded from the Training progress analysis.

[13] For both RTOs and V_1 cuts, the remaining crews either obtained the same grade on First Look and Transfer or did not have a Transfer trial due to technical problems.
[14] For both RTOs and V_1 cuts, the *n*s are too small to do the two motion groups separately.
[15] For both RTOs and V_1 cuts, in addition to omitting any crew who did not have a Transfer trial, any crew who obtained the same grade for both trials was also omitted.

Group Analyses

Figure 3.3 shows the mean grades for both motion groups on the first and last Training trials. I/Es perceived no differences in performance between the two motion groups for either RTOs (F(1, 19) = .14, p = .71) or V_1 cuts (F(1, 29) = 1.54, p = .23). However, the crews did improve across the Training for both RTOs (F(1, 19) = 57.02, p < .001) and V_1 cuts (F(1, 29) = 22.32, p < .001). The improvement was equal for the two motion groups (RTOs: F(1, 19) = 1.92, p = .18; V_1 cuts: F(1, 29) = .45, p = .51).

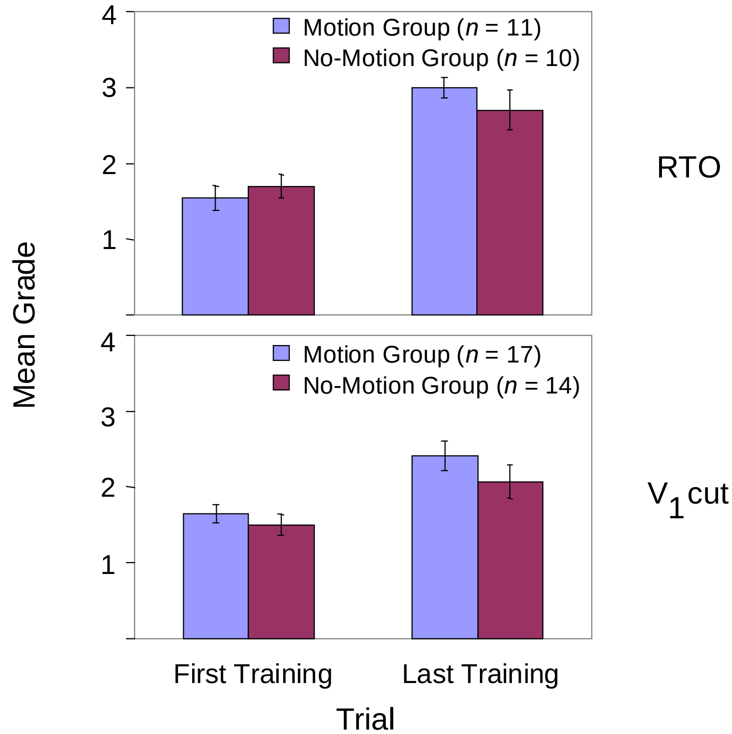


Figure 3.3 Mean grade as a function of maneuver, trial, and motion group (Error bars indicate standard error).

Individual Analyses

Motion vs. No-Motion. For the individual analyses, the distribution of grades at both the first Training trial and the last Training trial is shown in Figure 3.4. The low/high analyses and the improvement analyses confirm that the presence of motion did not affect grades at either the first or the last Training trials (see Table 3.4), or improvement between those trials (see Table 3.5). Thus, with the group data, there were no differences between the two motion groups.

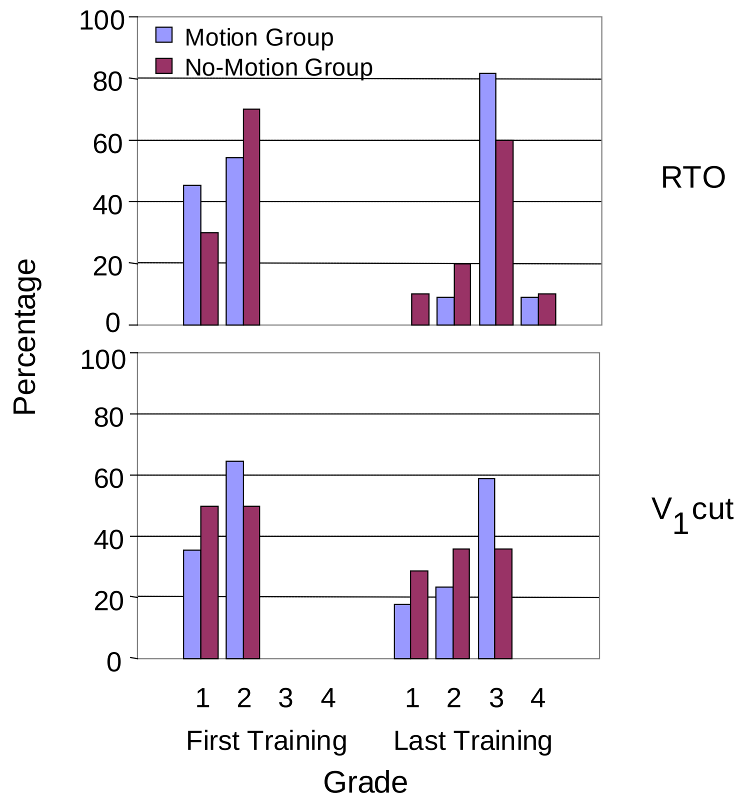


Figure 3.4 Percentage of grades in each grading category as a function of maneuver, trial, and motion group.

Table 3.4 Fisher Exact Probability Values for Test of Motion Group Differences in the Number of Low (1, 2) and High (3, 4) Grades for Each Maneuver and Trial

Maneuver	**Trial**	**Motion Group**		**Fisher Exact Probability**
		Motion Group	**No-Motion Group**	
RTO	First Training	11 low, 0 high	10 low, 0 high	$N = 21$, $p = 1$
	Last Training	1 low, 10 high	3 low, 7 high	$N = 21$, $p = .26$
V_1 Cut	First Training	17 low, 0 high	14 low, 0 high	$N = 31$, $p = 1$
	Last Training	7 low, 10 high	9 low, 5 high	$N = 31$, $p > .17$

Table 3.5 Fisher Exact Probability Values for Test of Motion Group Differences in Improvement From First Training Trial to Last Training Trial for Each Maneuver

Maneuver	Motion Group[16]		Fisher Exact Probability
	Motion Group	No-Motion Group	
RTO	10 improved, 0 got worse	8 improved, 0 got worse	$N = 18$, $p = 1$
V_1 Cut	12 improved, 1 got worse	7 improved, 1 got worse	$N = 21$, $p = .63$

Improvement. To determine if the individual analyses support the finding that the crews improved across the first and last Training trials, for both RTOs and V_1 cuts, the number of crews to improve from the first to the last Training trial was examined, combining across the two motion groups (i.e., instead of comparing the two groups). Indeed, the individual analyses do show the same pattern (see Table 3.6).

Table 3.6 Probability Values for Test of Improvement Between First Training Trial and Last Training Trial for Each Maneuver

Maneuver	Motion Groups Combined[17, 18]	Chi Square and Probability
RTO	18 improved, 0 got worse	See footnote[19]
V_1 Cut	19 improved, 2 got worse	$\chi^2(1, N = 21) = 13.76$, $p = .0002$

Two additional individual analyses of motion group differences for Training progress were done. All maneuvers that came before the normal take-offs, regardless of the total number of Training trials for that crew, were included here. The first analysis compared the total number of Training trials required by each motion group, and the second compared the number of crews to reach the criterion grade of 3 or 4 in each group. Again, there were no differences between the Motion and No-Motion groups in either of these comparisons for either of the two engine failure maneuvers (see Table 3.7).

[16] For both maneuvers, the remaining crews did the same on their first and last Training trial (or were omitted because they only did one Training trial, as in the other analyses in this section).
[17] For both RTOs and V_1 cuts, the *n*s are too small to analyze the two motion groups separately.
[18] For both maneuvers, the remaining crews did the same on their first and last Training trial (or were omitted because they only did one Training trial, as in the other analyses in this section).
[19] For the RTOs in this analysis, the *n*s were too small even with the two motion groups combined to do a Chi-Square test. Therefore, the test of significance was omitted, but the numbers themselves strongly suggest that the crews improved.

Table 3.7 Fisher Exact Probability Values for Test of Motion Group Differences as a Function of Maneuver and Type of Comparison

	Number of Training Trials Required[20]	Number of Crews To Reach Criterion Grade of 3 or 4
RTO	**No difference**	**No difference**, $p = .26$
Motion	19 crews took 35 trials	18 crews out of 19 = 95%
No-Motion	17 crews took 31 trials	14 crews out of 17 = 82%
V1	**No difference**	**No difference**, $p > .24$
Motion	18 crews took 45 trials	12 crews out of 18 = 67%
No-Motion	16 crews took 40 trials	8 crews out of 16 = 50%

Transfer from Last Training Trial to Transfer Trial

One question was what happened when crews transitioned (or transferred) from the last Training trial in the simulator with or without motion to the Transfer trial in the simulator with motion. This could be looked at as the true test of transfer of skills learned during Training. Even if I/Es perceived no difference between the two motion groups in Training progress and on the Transfer trials *per se* (as already discussed), there may still be a difference in how well crews' skills transfer to the simulator with motion from the very last Training trial. To test this, the last Training trial was compared with the Transfer trial. Because the purpose of this comparison was to compare pre-Transfer with post-Transfer performance, the last trial during the Training period was used as the pre-Transfer trial, even if it was the only Training trial (and thus also the First Look trial). In addition, only crews who actually completed the Transfer period of the experiment were included.

Group Analyses

The group analyses, once again, showed no difference between the Motion and No-Motion groups for the grades for RTOs ($F(1, 30) = 1.55$, $p = .22$) or V_1 cuts ($F(1, 30) = 1.34$, $p = .26$). One significant effect, however, was that the crews obtained significantly worse grades on the RTO during the Transfer trial compared to the last Training trial ($F(1, 30) = 11.76$, $p = .002$). This did not happen with the V_1 cuts ($F(1, 30) = .19$, $p = .67$) (see Figure 3.5). As before, motion group did not affect improvement (or deterioration, as it may be) for either RTOs ($F(1, 30) = .05$, $p = .82$) or V_1 cuts ($F(1, 30) = 0$, $p = 1$).

[20] Although no statistical method was deemed appropriate for comparison, it appears to be evident from inspection that there are no significant differences between the two motion groups.

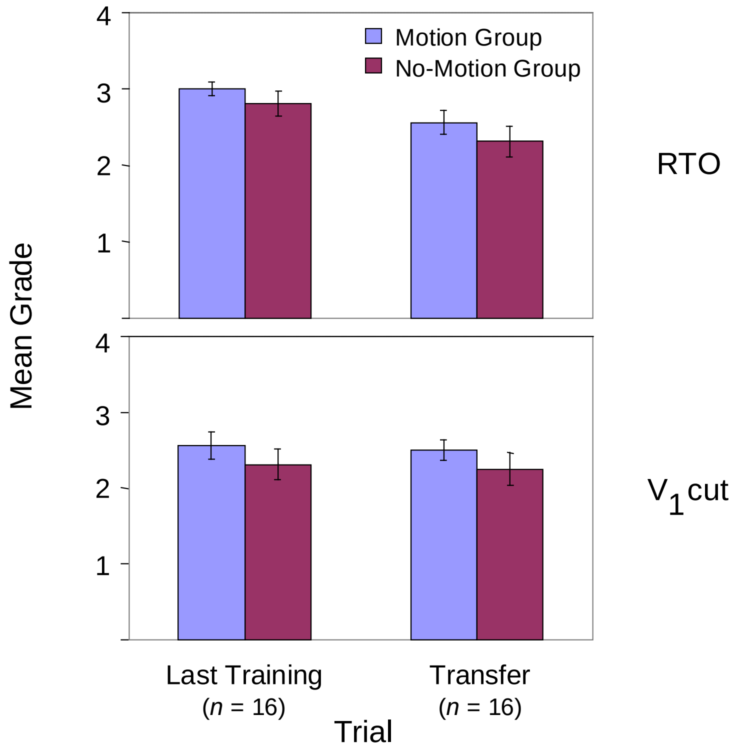


Figure 3.5 Mean grade as a function of maneuver, trial, and motion group (Error bars indicate standard error).

Individual Analyses

Motion vs. No-Motion. To determine if this lack of an effect of motion on grades during the last Training trial and the Transfer trial and the lack of an effect of motion on improvement between those two trials is supported by the individual data, the two groups were compared again looking at the low vs. high grades and individual improvement. Figure 3.6 shows the distribution of grades across these two trials. Once again, there was no effect of motion group membership on either performance at the last Training trial (see Table 3.8) or the Transfer trial or on the change between the last Training trial and Transfer trial for either maneuver (see Table 3.9).

However, the analysis of the low/high grades of the Transfer trial reported here is the same as the one reported in the section on First Look and Transfer. Therefore, the low/high analysis using the one vs. higher grades that revealed a significant difference between the motion groups, described in that section, applies here as well.

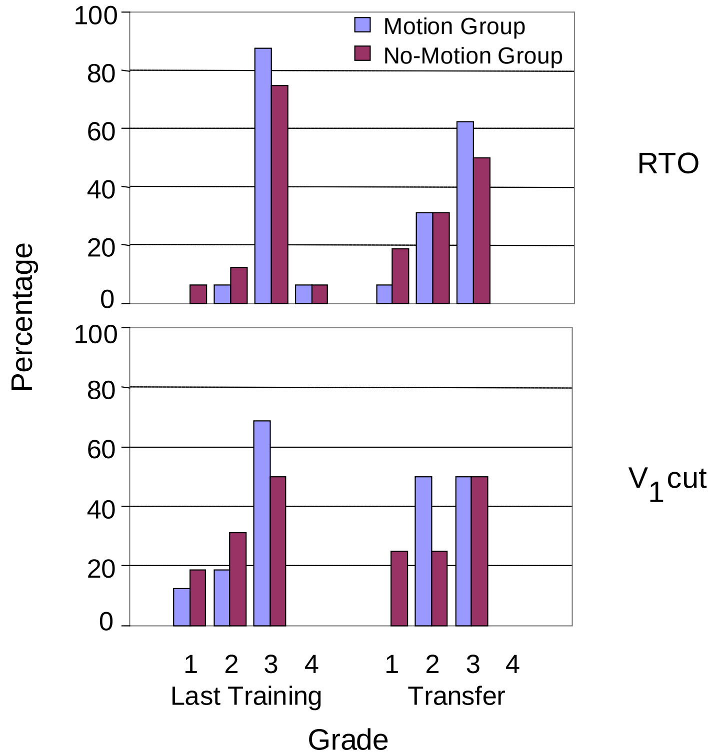


Figure 3.6 Percentage of grades in each grading category as a function of maneuver, trial, and motion group.

Table 3.8 Fisher Exact Probability Values for Test Of Motion Group Differences in the Number of Low (1, 2) and High (3, 4) Grades for Each Maneuver and Trial

Maneuver	**Trial**	**Motion Group**		**Fisher Exact Probability**
		Motion Group	**No-Motion Group**	
RTO	Last Training	1 low, 15 high	3 low, 13 high	$N = 32, p = .3$
	Transfer	6 low, 10 high	8 low, 8 high	$N = 32, p > .22$
V_1 Cut	Last Training	5 low, 11 high	8 low, 8 high	$N = 32, p > .22$
	Transfer	8 low, 8 high	8 low, 8 high	$N = 32, p > .28$

Table 3.9 Fisher Exact Probability Values for Test of Motion Group Differences in Improvement From Last Training Trial to Transfer Trial for Each Maneuver

Maneuver	Motion Group[21]		Fisher Exact Probability
	Motion Group	No-Motion Group	
RTO	1 improved, 7 got worse	1 improved, 7 got worse	$N = 16$, $p = .77$
V_1 Cut	2 improved, 4 got worse	4 improved, 4 got worse	$N = 14$, $p = .47$

Improvement. To see if the deterioration in performance (for both motion groups) from the last Training trial to the Transfer trial for RTOs, but not for V_1 cuts, is supported by the individual data, the number of crews to improve, combining across the two motion groups (i.e., instead of comparing the two motion groups), was examined (see Table 3.10). Indeed, these analyses do support the finding that performance deteriorated for RTOs, but not for V_1 cuts.

Table 3.10 Probability Values for Test of Improvement Between Last Training Trial and Transfer for Each Maneuver

Maneuver	Motion Groups Combined[22, 23]	Chi Square and Probability
RTO	2 improved, 14 got worse	See footnote[24]
V_1 Cut	6 Improved, 8 got worse	See footnote [24]

Summary: V1 cuts and RTOs

In all cases, but one, there were no differences between the Motion and No-Motion groups using any analyses examining any aspect of this experiment. That is, the group analyses, the low/high analyses (with the median split), and the improvement analyses revealed no differences between the groups for First Look evaluation, Transfer of training, improvement between First Look and Transfer, Training progress, and improvement between the last Training trial to Transfer. This was shown for both RTOs and V_1 cuts.

However, one difference was found for crews performing V_1 cuts during Transfer (when all crews had motion). There was a difference in the number of the lowest grades (i.e., grades of 1) vs. the next lowest grade (i.e., grades of 2) assigned to crews dependent on whether they had had motion or not during Training. Apparently, the addition of motion for crews who had not had motion increased the probability that mediocre performance on a V_1 cut would become unacceptably poor. This might indicate that simulators without motion provide less good training

[21] For both maneuvers, the remaining crews did the same on their last Training trial and their Transfer trial (or were omitted because they did not have a Transfer trial as in the other analyses in this section).
[22] For both maneuvers, the *n*s are too small to do the two motion groups separately.
[23] For both maneuvers, the remaining crews did the same on their last Training trial and the Transfer trial (or were omitted because they did not have a Transfer trial as in the other analyses in this section).
[24] For the RTOs and the V_1 cuts in these analyses, the *n*s were too small even with the two motion groups combined to do a Chi-Square test. Therefore, the tests of significance were omitted, but the numbers themselves strongly suggest that the crews got worse on the RTOs but stayed the same on the V_1 cuts.

than simulators with motion. However, the single motion group advantage must be interpreted within the abundance of grade results (i.e., all of the parametric analyses and all of the other nonparametric analyses) showing no differences between the two motion groups.

Although the motion state had nearly no impact on the crews' performance, as measured by I/E grades, the motion state may have had an impact on the I/Es' grading criteria, given that the I/Es are subjected to the same motion (or lack of motion) as the pilots. If so, differences in crew performance might not be reflected in the grades assigned by the I/E. This concern is addressed and dismissed in Section 3.2.4.

With regard to motion-independent effects, the crews did improve for both RTOs and V_1 cuts. The pattern of improvement, however, differed slightly for the two maneuvers. For both RTOs and V_1 cuts, the crews improved across the Training period. For RTOs, however, this improvement was lost during the Transfer period, resulting in no overall improvement. For V_1 cuts, the crews maintained their performance level during the Transfer period, resulting in overall improvement. One possible explanation is that the crews had a certain degree of complacency with regard to RTOs, for which 89% of the crews had reached the criterion grade of 3, as opposed to V_1 cuts, where only 59% had reached criterion during training.

<u>Normal Take-Offs</u>

Although the normal take-offs were never considered diagnostic for the need for motion and were only included to reduce the expectation of an engine failure during the Transfer period, motion group differences were still assessed. However, because there was no attempt to train performance on the normal trials, improvement across the two trials was not analyzed. The grades from the two normal trials for a single crew were averaged into one mean grade for that crew.

Group Analysis

The group analysis of the normal take-off maneuvers did not reveal a motion group difference ($F(1, 31) = 0.96$, $p = .34$), as can be seen in Figure 3.7.

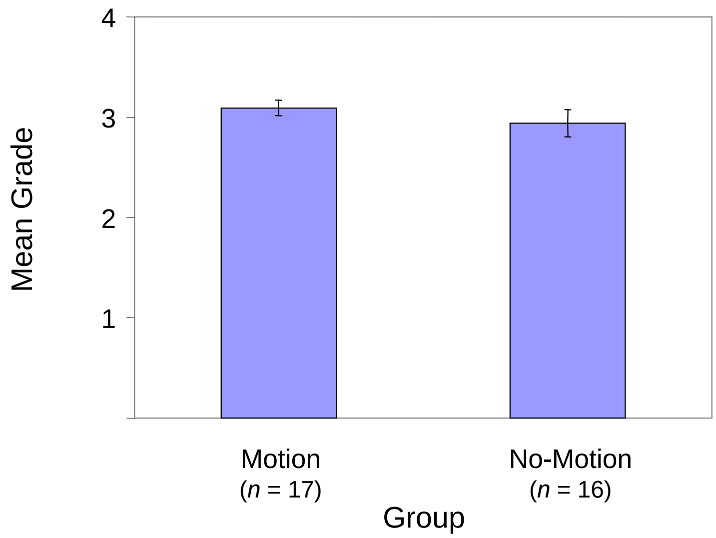


Figure 3.7 Mean grade on normal take-offs as a function of motion group (Error bars indicate standard error).

Individual Analysis, Motion vs. No-Motion

Figure 3.8 provides the grade distribution for the mean grades for the normal trials for both the Motion and No-Motion crews. The individual analysis of low vs. high grades supports the lack of a motion group difference ($N = 34$, Fisher Exact $p = .27$).[25]

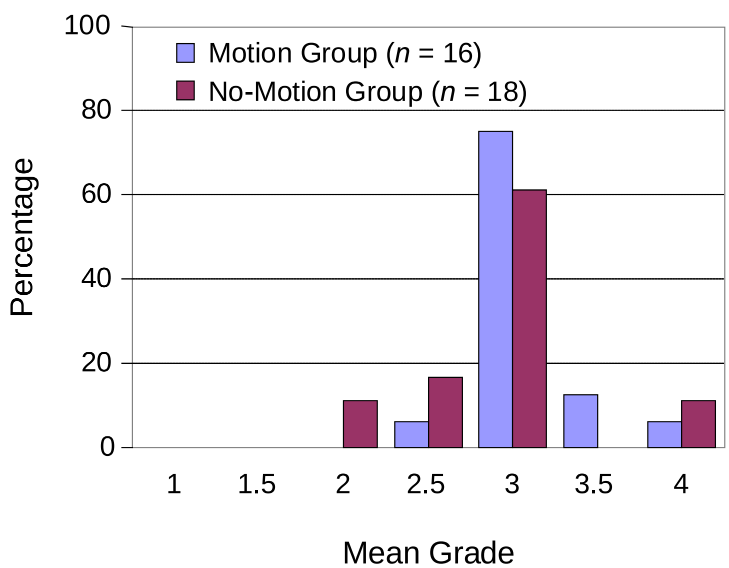


Figure 3.8 Percentage of grades in each grading category for normal take-offs.

[25] All mean grades of 2.5 were grouped with the high grades. If they are grouped with the low grades, there is still no significant effect of motion ($N = 34$, Fisher Exact $p = .12$)

Overall Summary of Grade Results

The grade results are summarized in Table 3.11. Only one single effect of motion was found in all the analyses performed, and only when resorting to a special case of nonparametric analysis after having found no effect with an ANOVA or one type of nonparametric analysis. Moreover, no effects of motion were found for total number of Training trials or for number of crews to reach criterion.

Table 3.11 Summary of Grade Analyses ("n.s." indicates that all of the relevant analyses were not significant, "sig." indicates that all of the relevant analyses were significant, bold font indicates that the result is statistically significant**)**

Within-Subjects Variable: Trial 1	Trial 2	Task	*p* Values: Trial ANOVA	Trial Indiv.	Motion ANOVA	Motion Indiv.[26]	Motion x Trial ANOVA	Motion x Trial Indiv.
First Look	Transfer	RTO	.34	.14	.50	>.22	.53	.61
		V_1 Cut	**.004**	**.003**	.29	> .28 & **.05**[27]	.59	.71
First Training	Last Training	RTO	**< .001**	**sig.**[28]	.71	> .26	.18	1
		V_1 Cut	**< .001**	**.0002**	.23	> .17	.51	.63
Last Training	Transfer	RTO	**.002**	**sig.**[28]	.22	> .22	.82	.77
		V_1 Cut	.67	n.s.[28]	.26	> .22 & .05[27]	1	.47
Normal Take-Off					.34	.27		
Summary Description			General improvement and, for RTOs, a loss of improvement		Nearly no differences between the Motion and No-Motion groups		No differences in improvement between the Motion and No-Motion groups	

3.1.2 Questionnaires: Ratings

To see whether the motion state of the simulator affected participants' opinions of the simulator, three different questionnaires were administered to the PF, PNF, and I/E. They were administered at two different times, i.e., first after completion of the Training period (i.e., before the No-Motion group had experienced motion) and then again at the end of the experiment (i.e., after all participants had experienced motion). Each questionnaire asked about control precision/performance, control strategy and technique, workload, gaining proficiency, comfort, and acceptability (only for the PF and PNF) (see Appendix E). Generally, the PFs rated themselves, and the PNFs and I/Es rated the PFs, with the exception of comfort where everybody

[26] Two analyses were done for each of these comparisons. The lowest *p* value is reported here. Both *p* values can be found in the specific tables within each section.
[27] This was the single case in which the low/high comparison using the 1 vs. 2, 3, 4 split resulted in a significant motion group difference even though the other comparisons did not.
[28] The *n*s were too small to use the Chi-Square test. Therefore, the test of significance was omitted, but the numbers themselves are clear.

rated their own, and acceptability, which referred to the simulator. For the first three questions the PFs and PNFs were asked to base their ratings on comparisons with performance in the airplane. For the other three, the ratings were to be based on comparisons with the simulator last time. The I/E gave grades or compared the PF with the average PF or with the simulator last time, as appropriate. Questionnaire ratings always ranged from 1 to 5, except for the I/Es grades for performance, which ranged from 1 to 4. For the visual representation of the questions, which used a 1 to 5 rating scale, the scale was translated to -2 to +2 rating scale. Thus, any negative data points indicate that the simulator was rated worse than either the airplane or the simulator last time whereas any positive data points indicate that the simulator was rated better than either the airplane or the simulator last time The specifics of the scales (e.g., whether from worse to better or harder to easier, etc.) depended on the question and can be seen in Appendix E.

The questionnaires were analyzed in a variety of ways. First, for each question an ANOVA tested whether there was a significant difference in responses between the two motion groups. Second, for each question the response made after Training was compared to the response made after Transfer.[29] In the after-Training questionnaire the questions about control precision/performance and control strategy and technique had several subquestions. Also, the last subquestion asked for an overall rating. It was this overall rating that was compared to the rating from the after-Transfer questionnaire (because there were no subquestions in that questionnaire). The workload question had two parts (i.e., mental and physical) but no overall question; therefore for the comparison with the after-Transfer questionnaire, the mean of the two parts was used. The after-Training vs. after-Transfer comparison also addressed the motion issue because the first questionnaire was filled out when the motion system was still in its original state (i.e., the No-Motion group had still not had any motion), but the second questionnaire was filled out after all crews had had motion. Therefore, any changes in responses between the first and second questionnaire for the No-Motion group, but not for the Motion group, would point to an impact of motion.

An unanticipated problem surfaced with the control strategy and technique question. It was worded in a way that could have been interpreted to be asking about simulator control loading. As shown in Appendix E, the page was clearly labeled “Control Strategy and Technique.” Then, however, the pilots were asked specifically to indicate whether operating the controls was harder or easier than expected (compared to the airplane). This ambiguity was even worse on the final questionnaire, because there was no specific reference to control strategy and technique, but participants were simply asked again whether operating the controls was harder or easier than expected. Because interpreting any pattern of responses from this question would be ambiguous at best, this question was omitted from analyses and interpretation.

<u>Motion Group Differences</u>

Four differences were found between the ratings of the Motion and No-Motion groups. In three cases, involving ratings of the PFs and PNFs, *absence* of motion led to higher ratings (although in one of the three, this was only after motion had been added) whereas in the fourth case, involving I/Es, the *presence* of motion led to higher ratings (but also only when motion was on for both groups).

[29] For all of these analyses if a participant had a response for one part of the analysis (e.g., a response from the after-Training questionnaire), but not a response for another part of the analysis (e.g., a response from the after-Transfer questionnaire), the data that exists was used.

Pilot Flying

The two groups, Motion and No-Motion, rated control precision equivalently after Training ($F(1, 35) = .03$, $p = .86$), but after Transfer, the No-Motion group rated control precision higher than the Motion group ($F(1, 32) = 3.21$, $p = .08$) (interaction: $F(1, 31) = 5.03$, $p = .03$) (see Figure 3.9). This is due to the Motion group giving significantly worse ratings to control precision after Transfer than after Training ($F(1, 15) = 7.74$, $p = .01$), but the No-Motion group giving equivalent ratings at both times ($F(1, 16) = .88$, $p = .36$). This suggests that the PFs who were exposed to motion through the entire experimental sequence, perceived a decline in control precision during Transfer. This decline was not found for the PFs who did their First Look evaluation and Training with the motion system turned off. These pilots rated their control precision both before and after Transfer as high as the Motion PFs did before Transfer. Whether their ratings would also have declined had they continued without motion, indicating that it was the change to motion that kept their ratings level, cannot be answered from this experiment.

Moreover, when comparing the simulator today to the simulator last time, the No-Motion PFs found gaining proficiency easier than the Motion PFs, combining across both questionnaires ($F(1, 35) = 3.43$, $p = .07$) (see Figure 3.9). This suggests that the PFs believed that the motion was somehow a hindrance, possibly a distraction.

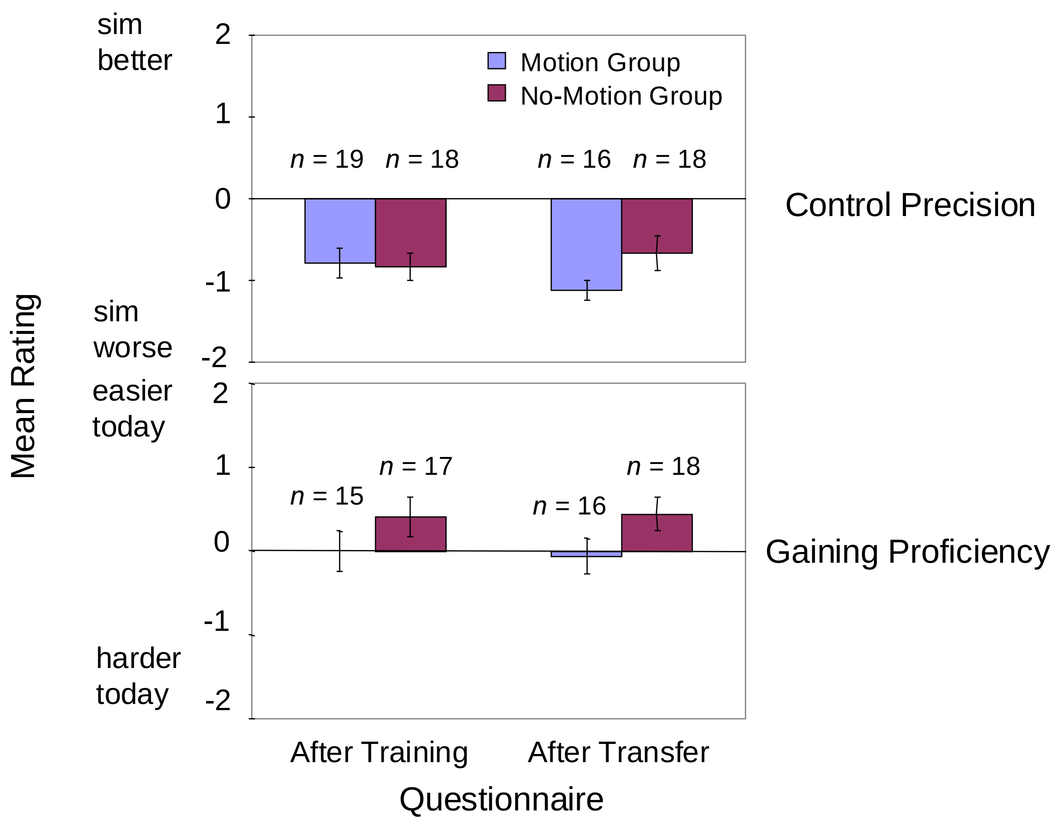


Figure 3.9 Mean rating given by the PFs as a function of specific question, questionnaire, and motion group (Error bars indicate standard error, the comparison for control precision was the airplane and for gaining proficiency was the simulator).

Pilot Not Flying

The No-Motion PNFs rated the PFs as having better control precision after Training than the PNFs in the Motion group did ($F(1, 36) = 5.76$, $p = .02$) (see Figure 3.10). This may suggest that control precision is actually better without motion than with motion. However, this opinion was not shared by the PFs or the I/Es, although after *Transfer* the PFs from the crews who were trained without motion did rate control precision higher than the PFs from the crews who had had motion during Training (see above).

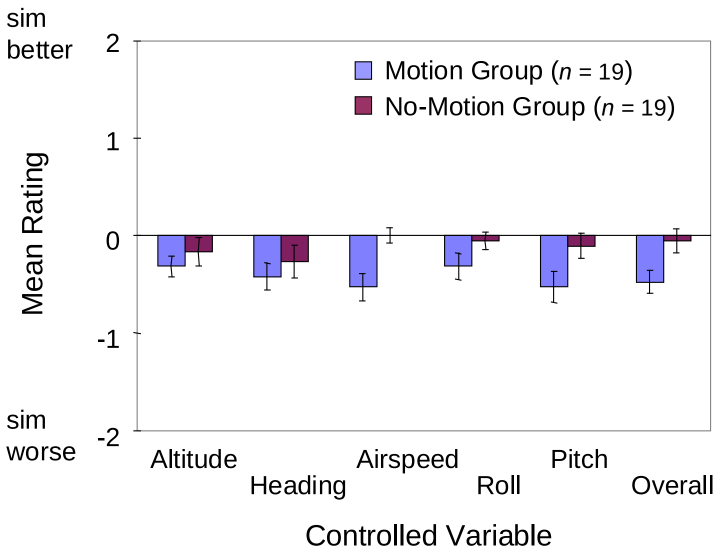


Figure 3.10 Mean rating given by the PNFs of the control precision of the PFs as a function of controlled variable and motion group (Error bars indicate standard error, the comparison was the airplane).

Instructor/Evaluator

The responses from the I/Es were analyzed slightly differently than the responses from the pilots. Each I/E may have tested more than one crew, thus filling out more than one pair of questionnaires. Therefore, for each motion group, a mean set of ratings was calculated for each I/E based on the questionnaires filled out by that I/E for the crews in that particular motion group. The underlying reason for this procedure was not to confound variability within a single I/E with variability between I/Es in the statistical analyses.

The I/Es gave equivalent ratings for control performance to the two groups of PFs after *Training* ($F(1, 7) = .18$, $p = .69$), but after *Transfer* they gave higher ratings for performance to the PFs who had had motion than to the PFs who had not had motion ($F(1, 6) = 10.84$, $p = .02$) (interaction: $F(1, 6) = 10.15$, $p = .02$) (see Figure 3.11). However, their ratings for both groups of PFs did not change significantly between the two questionnaires (Motion: $F(1, 10) = 1.77$, $p = .21$; No-Motion: $F(1, 8) = .37$, $p = .56$). This may suggest that even though motion had no impact on performance during Training, the history of having had motion results in better performance later during Transfer testing. This conclusion, however, was not supported by the ratings given for control precision by either the PFs or the PNFs. On the other hand, this result is somewhat consistent with the grades that the I/Es assigned, with regard to the single motion group difference found in grades. That is, the I/Es gave more extremely low grades (i.e., 1) than moderately low grades (i.e., 2) during Transfer for V_1 cuts.

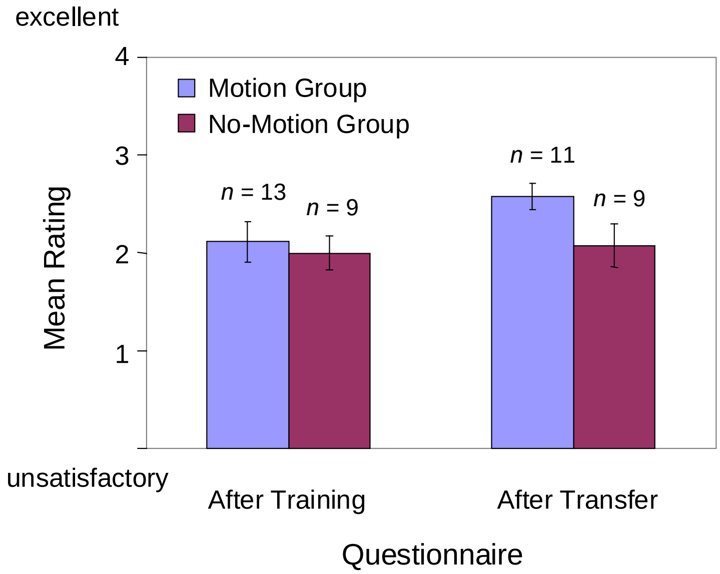


Figure 3.11 Mean rating given by the I/Es of the performance of the PFs as a function of questionnaire and motion group (Error bars indicate standard error, the comparison was the average PF).

In sum, the PFs trained without motion rated control precision better after Transfer and gaining proficiency easier throughout. This finding is reinforced by the higher PNF ratings of PF control precision during Training without motion. However, the I/Es found that having been trained with motion resulted in better PF performance during Transfer than having been trained without motion. It may be meaningful that three of the four effects of motion concerned control precision or performance; however, the importance of this convergence is diminished because the findings are contradictory.

Other (Non-Motion) Effects

There were two other significant effects that were independent of the presence or absence of motion. The PFs found workload to be lower after Transfer than after Training ($F(1, 32) = 7.05$, $p = .01$), but equally for both motion groups ($F(1, 32) = .109$, $p = .74$). This suggested that they were getting used to performing these failures in the simulator (Figure 3.12). In contrast, the I/Es rated workload for the PFs as higher after Transfer than before ($F(1, 12) = 3.37$, $p = .09$) (again, regardless of motion group membership, $F(1, 6) = 1.49$, $p = .27$), suggesting that the PFs had to work harder and harder as they continued through the session (Figure 3.12). Again, the ratings of the I/Es contradict the ratings of the pilots.

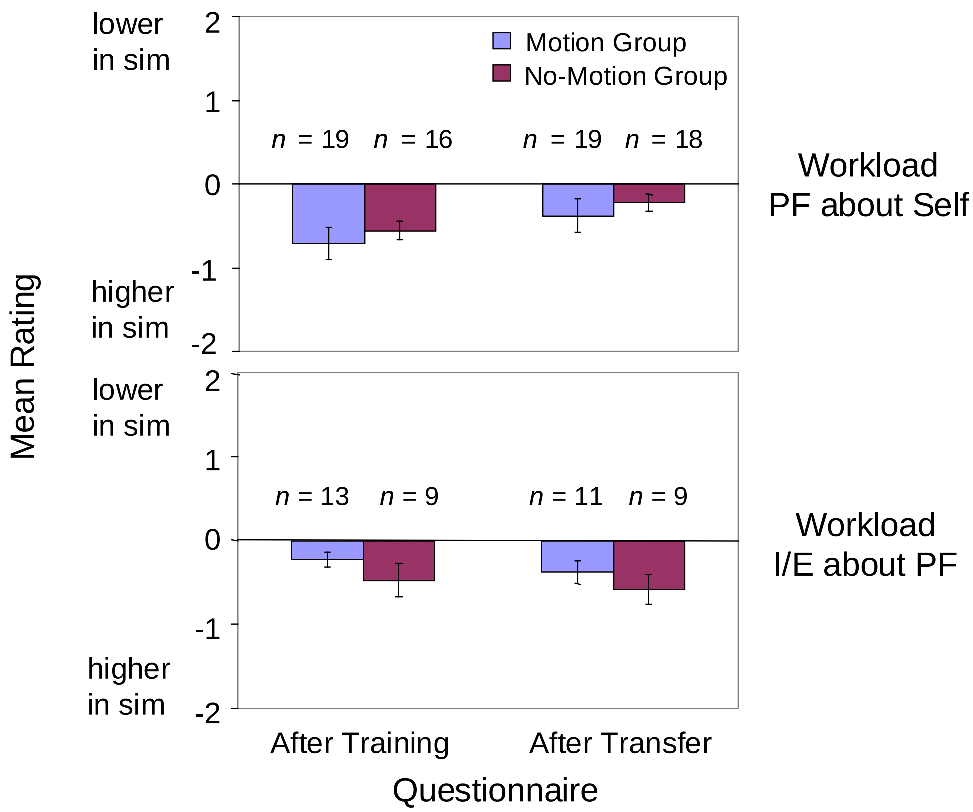


Figure 3.12 Mean rating of workload as a function of respondent, questionnaire, and motion group (Error bars indicate standard error, the comparison was the airplane for the PFs and the average PF for the I/Es).

3.1.3 Questionnaires: Additional Comments

At the end of the second questionnaire, the participants were asked if they had any additional comments (see Table 3.12). Most importantly, only one of the 13 No-Motion PFs that had any comments explicitly mentioned simulator motion (and none of the six Motion PFs providing comments). However, three of the ten No-Motion PNFs and one of the six Motion PNFs with comments mentioned motion. Moreover, two of the No-Motion PNFs mentioned disorientation, although one of these was actually referring to the last RTO Transfer Testing trial with motion, where the crew had obtained an unsatisfactory grade, "The last take-off roll was disorienting a few more of those would definitely make me vomit." Nevertheless, this may support anecdotal evidence that the less pilots are actively engaged in flying, the less complete is the illusion of motion provided by peripheral vision. None of the eight I/Es, when commenting on the Motion crews, mentioned motion, whereas two of the nine I/Es, when commenting on the No-Motion crews, did so. Note that an I/E may have previously evaluated a crew from the other group. The full comments on motion provided by participants are provided below (all quotes in this section are verbatim, but capitalization was adjusted and emphasis was added).

Motion Crews

PNF: "I felt I really needed to watch my instruments to get a sense of what was going on. **Something didn't feel quite right with motion at times**. Today, it felt like I needed to adjust to the simulator a little more than normal."

No-Motion Crews

PF: "While doing the first set of maneuvers **without motion**, when left Rudder trim was selected it dident work at first and then would give full trim all of a sudden."
PNF 1: "**The simulator motion is still somewhat different from the aircraft**."
PNF 2:"Unrealistic with no motion. Things felt more unnatural & uncommon without motion. I was more comfortable with my duties & the A/C with motion, although the **sim's motion is <u>Not</u> like it is in a real airplane** & probably never will be. I does (& did) cause orientation that you wouldn't get in the same situation in an airplane."
PNF 3: "1. **The yaw in the sim is much less than in the airplane (motion)**. 2. **The motion in the** sim needs to be more pronounced. Its much less than you feel in the airplane."
I/E 1: "PF controlled A/C more accurately with motion on although adherence to airspeed tolerances was poor."
I/E 2: "Try putting motion on first! Then off."

In addition to the comments about motion and disorientation, there were also comments about the degree of simulator fidelity. Overall, the No-Motion PFs complained almost twice as often as the Motion PFs that the simulator was not like the airplane. Most often, they found the simulator controls too sensitive. This specific concern was explicitly shared by three of the six Motion PFs pointing out differences between the simulator and the airplane, but by eight of the 13 No-Motion PFs. The sensitivity issue was identified by three of the No-Motion PNFs, but by none of the Motion PNFs. Finally, one Motion PNF mentioned that "[t]he noise level was much better as far as realism was concerned." Three I/Es, however, found it "way" or "much too loud." As mentioned earlier, the sound level had been set at 65%, which does represent 100% airplane noise. It is well known that this level is often reduced during training to enable communications and reduce fatigue.

Table 3.12 Summary of Additional Comments as a Function of Respondent and Motion Group

		Total Number of Comments	Comments About Simulator Motion	Comments About Disorientation	Comments About Simulator Controls Being Too Sensitive	Comments About Sound Level
PF						
	Motion	6	0	0	3	0
	No-Motion	13	1	0	8	0
PNF						
	Motion	6	1	0	0	1
	No-Motion	10	3	2	3	0
I/E						
	Motion	8	0	0	0	3
	No-Motion	9	2	0	0	0

3.2 OBJECTIVE RESULTS

Four types of objective results are presented. First, erroneous take-off decisions are reported. These occurred very rarely and are reported for completeness only. Second, to put any effect or lack of an effect of platform motion in perspective, the motion stimulation provided by the simulator is discussed in light of what is known about the motion stimulation in the real airplane. Third, the criterion measures to decide whether motion has an effect on training and evaluation of the test maneuvers were determined. To complement the standards provided by the FAA Practical Test Standards (FAA, 1995) and the host airline itself, an attempt was made to capture the criteria used by the I/Es when they were grading the pilots on the respective maneuvers by performing correlation and regression analyses between I/E grades and objective measures. These analyses also showed whether the presence or absence of motion affected which measures I/Es considered for grading. Finally, the effect of motion on a) First Look evaluation, b) Transfer of training to the simulator, c) Training progress and d) improvement from last Training trial to Transfer testing was examined.

3.2.1 Incorrect Take-Off Decisions

In the experiment, the crews occasionally made the incorrect take-off decision. That is, at times they decided to take off on an RTO or to reject the take-off on a V_1 cut. This occurred very rarely in this experiment and only during the First Look maneuvers.

Table 3.13 shows the number and percentage of incorrect take-off decisions that occurred in the experiment. There was no significant difference in the number of errors in the two groups for either RTOs (N = 101, Fisher Exact p = .49) or V_1 cuts (N = 123, Fisher Exact p = .49), suggesting that the platform motion did not affect take-off decisions.

Table 3.13 Number of Incorrect Take-Off Decisions for Each Motion Group and Maneuver

	Crew Rejected Take-Off At $V>V_1=V_R$	**Crew Took Off at $V<V_1=V_R$**
Motion Group	0 (0%)	1 (1.96%)
No-Motion Group	1 (1.67%)	2 (4%)

3.2.2 Motion Stimulation Provided by the Simulator

As mentioned previously, 78 simulator state and control input variables were recorded (see Appendix F). These included linear accelerations and angular rate information used to assess simulator motion performance. Three types of information were collected: 1) the outputs of the equations of motion of the simulator, 2) the outputs of the motion drive equations, and 3) direct measurements provided by accelerometers and a rate gyro installed directly under the pilot station. Representative examples that allow comparison of the accelerations from equations of motion, motion drive equations, and accelerometer and rate gyro measurements are shown in Appendix G.

As shown in Appendix G, for roll rate and longitudinal acceleration, the directly measured motion followed the airplane model fairly well, given the limitations inherent to all simulators. These limitations account for the magnitude attenuation of the measured response compared to the outputs from the equations of motion. The immediate changes in longitudinal acceleration and roll rate following an engine failure can be seen in the measured response as commanded by the equations of motion.

For vertical acceleration, however, the motion system of the test simulator did not respond much to the command provided by the equations of motion. This was especially true for V_1 cut maneuvers. However, because the engine failures used in our experiment did not produce much vertical acceleration, the lack of vertical acceleration cuing might not have been very important.

More important, however, was the finding that the failure induced lateral acceleration was not well represented by the motion system of the test simulator. Not only was it greatly attenuated, but also visual inspection of the measured response did not lead to an easy distinction of failure-induced lateral acceleration, unlike the response derived from the equations of motion (relatively high peak shortly after engine failure). This may represent a significant deficiency in pilot stimulation, because lateral acceleration may act as a useful cue for proper failure recognition and for initiation of appropriate pilot response. To the best of the authors' knowledge, however, the importance of lateral versus other cues in the failure recognition has not been systematically examined in the literature. Research for developing motion-cuing criteria based on human perception has been recommended or initiated only very recently [see, for example, Hosman (1999) and White and Rodchenko (1999)].

3.2.3 Criterion Measures

From the 78 variables recorded in the experiment, a set of measures was derived, which were categorized into performance and workload/behavior measures. Performance measures reflect a pilot's control precision and efficiency in handling the airplane by measurements such as flight path deviations and reaction time. Workload/behavior measures describe how a pilot uses the controls by measurement of control inputs.

The question that needed to be answered was how the variables recorded would help discriminate between different levels of performance and workload. A first cut was provided by the PTS (FAA, 1995) and the standards provided by the host airline itself, which were in fact somewhat stricter than the FAA standards (see Appendix H). For RTOs, the PTS and company standards focus on directional measures (i.e., measures related to yaw, heading, and lateral deviation). For V_1 cuts, the standards include, in addition to directional measures, longitudinal (pitch and airspeed) and lateral (roll and bank angle) measures.

An additional goal was to capture performance and workload immediately after the engine failure, because disturbance motion was expected to act as an alerting cue to the pilots that would enhance early performance. Measures that may be sensitive to an early alerting effect of motion include reaction times and Integrated Failure Induced Bank Angle, among others. The lists of measures extracted from the experiment data for rejected take-offs (RTO) and V_1 cut maneuvers are given in Appendix I and J, respectively. In general, lower numerical values of the measures indicate better performance or lower workload.

Most of the measures selected were computed over the 15-second time period following an engine failure. Exceptions include measures of reaction time and time to reach 400 ft altitude. For RTO, reaction time was defined as the time it took pilots to reduce the power of the good engine instead of the time it took pilots to activate the brake. See Appendix K for a detailed explanation of this issue.

To complement the PTS and company standards, an attempt was made to determine the standards used by the participating I/Es themselves when grading the maneuvers. Although PTS and company standards are largely based on subject matter expert (SME) opinion, this analysis may reveal additional performance or behavioral aspects considered by I/Es when grading; aspects of which the SMEs were not consciously aware. A second purpose of this analysis was to identify measures that are closely related to each other and can therefore be treated as one. The third and last purpose of this analysis was to examine whether the exposure of I/Es to motion (or lack of motion) at the instructor station influences their grading criteria. This may have implications on how to interpret the subjective grade data for First Look evaluation and Training.

3.2.4 Relationship between the I/E Grades and the Objective Measures

To examine the relationship between I/E grades and the objective data as well as between the objective data with each other as a function of motion status, separate correlation and regression analyses were performed for the motion-on and motion-off conditions. Details of the analyses and the results will be reported in Go, Bürki-Cohen, DiSario, and Jo (in preparation). Only the results pertinent to this report are presented here.

From the correlation analysis, it was found that for RTOs, Mean Absolute Lateral Deviation and Root Mean Square (RMS) Lateral Deviation are highly correlated to each other ($r \geq 0.95$). Therefore, Mean Absolute Lateral Deviation is used to represent lateral deviations in the subsequent discussion of RTOs.

For V1 cuts, two pairs of measures were highly correlated to each other ($r \geq 0.94$), namely, RMS Heading Deviation and Integrated Heading Exceedance, and Integrated Roll

Activity and Total Integrated Roll Yaw Activity. Therefore, only the results for RMS Heading Deviation and Integrated Roll Activity are presented for V_1 cuts.

Next, stepwise logistic regression analyses were performed to find measures that significantly predict I/E grades. Logistic was preferred over linear regression because it is more suitable for cases involving ordinal data (like the grading system used here). Regressions where the model was forced to include one of the measures even if it had not been found a significant predictor when running the analysis without forcing, were also performed to obtain some alternative predictive models. The Statistical Analysis Software (SAS) output summaries for the models where all measures, including the forced ones, were found to significantly predict I/E grades are presented in Appendices L and M. Note, however, that the regression models obtained were not meant to model I/E's decision process in determining the grades, which is actually very complex. They were only used to examine if the available measures related to I/E's grading criteria.

For RTOs, regardless of whether the motion system was on or off, the measures of lateral and heading deviations played an important role in predicting the I/E grades. All the predictive models obtained from the logistic regression analysis included at least one such measure. As can be seen from Appendix L, Standard Deviation (STD) of Pedal Position might also affect the I/E grades, however, only when the motion system was off.

For V_1 cuts, none of the measures was found to *consistently* predict I/E grades regardless of the motion system status. When the motion system was on, the predictive model for I/E grades always included longitudinal measures (i.e. Time to Reach 400 ft Altitude, STD Pitch Angle, Integrated Airspeed Exceedance, and STD Column Position). Some lateral measures, such as Integrated Roll Activity, Integrated Bank Angle Exceedance, and Maximum Bank Angle, also appeared in some models. However, when the motion system was turned off, I/E grades for V_1 cuts were more strongly predicted by lateral measures, such as Integrated Bank Angle Exceedance, STD Wheel Position, and Maximum Bank Angle. The predictive models also suggested measures, such as Time to Reach 400 ft Altitude and Pedal Reaction Time, might affect the I/E grades without motion.

The result for RTO maneuvers was not surprising, considering the airplane was still on the ground and variables such as the lateral and heading deviations could easily be assessed by the instructors based on visual information alone, without depending on motion information. For V_1 cut maneuvers, the results indicated that the platform motion status might affect grading. In both motion-on and motion-off conditions, some (but not all the same) lateral measures seemed to play a role in determining the I/E grades. However, longitudinal measures appeared to matter mainly when the motion system was on (except time to reach altitude).

Given that I/Es may have used different grading criteria dependent on motion status, does this imply that an effect of motion on grades before Transfer to the simulator with motion may have been masked? This appears to be a possibility at least for V_1 cuts, where the No-Motion pilots would have been able to get away with worse performance on longitudinal measures. The decisive question is whether or not the objective results will confirm the subjective results, specifically, whether the No-Motion crews did indeed perform worse on longitudinal measures for V_1 cuts before Transfer. To anticipate a result described in Section 3.2.6, the Motion pilots did indeed control pitch angle marginally better than the No-Motion pilots (p = .096); however, the difference was very slight (less than one degree in STD) and not accompanied by improvements in any of the other longitudinal measures. Moreover, there was practically no simple correlation between V_1 cut Pitch Angle STD of Motion pilots and grades (R^2 = .01), and even the stepwise regression model yielding all four (and only) longitudinal measures (Time to Reach 400 ft Altitude, STD Pitch Angle, Integrated Airspeed Exceedance, STD Column Position) accounts for no more than 30% of the variance in the grades (R^2 = .30).[30] Based on these findings, and on the very minor differences in grading criteria for RTOs, it appears highly unlikely that there are masked differences in the grades assigned to the two motion groups.

3.2.5 Statistical Power to Detect an Effect of Motion

One very important consideration when considering experimental results is whether the quality of the data gathered is sufficient to reveal an existing effect. This is commonly referred to as "power." The power of an experiment to reveal (or the power of a statistical analysis to detect) an operationally relevant effect is directly proportional to the size of this effect and the number of subjects in each group, and indirectly proportional to the variability between subjects within each group. Therefore, a finding of no difference between two groups does not necessarily mean that the difference is absent, but that the difference may be so small that it is masked by the variability between subjects within each group. In this study, the resolution of the analysis was inferred by calculating the effect size to reach the power of .80 for each measure (i.e., the minimum difference between the standardized means that will lead to the rejection of the null hypothesis with a probability of .80). Power $\geq$ .80 is conventionally considered as sufficient (Cohen, 1988).

For each measure, the size of the effect that could be detected with sufficient power will be indicated in the results below. Already at First Look evaluation, the detectable effect sizes were considered small enough to include any operationally relevant effects. At Transfer, the power was increased (i.e., the detectable effect size reduced) for many measures due to a reduction of variability between subjects within each group, presumably due to training.

3.2.6 First Look Evaluation

The purpose of this section is to examine whether platform motion had an effect on pilot performance and workload during First Look evaluation (i.e., the very first time they flew the simulator after having flown only the airplane for the past six months). They immediately performed first the V_1 cut followed by the RTO, without being given a chance to adapt to the simulator.

RTO

The resolution obtained with 16 crews in the Motion group and 14 crews in the No-Motion group for each criterion measure is shown in Table 3.14 (i.e. the smallest effect of motion that could be detected with a power of .80). The number of crews included in the analysis here and on the rest of the report really indicates the number of complete data gathered in the experiment for the type of trial and maneuver. There might be more crews participating in the experiment, but some of the data gathered from some crews was incomplete due to technical difficulties. For the measures involving integration (Integrated Yaw Activity and Sum of Pedal Spectrum f>.5 Hz), it should be kept in mind that the integration was done during the 15 seconds

[30] This is the R^2 of the linear stepwise regression model with all measures offered, which yielded the same results as the logistic regression analysis (Go et al., in preparation). Logistic regression analyses do not provide R^2.

following an engine failure. Hence, for Integrated Yaw Activity, e.g., the detectable effect size was about 1.73 degrees per second. The table shows that the experiment should have been able to capture all effects that would significantly impact operations, and in some cases much more than that (e.g. Power Lever and Pedal Reaction Time of less than 1 second). Maximum Heading Deviation, with a resolution of slightly over four degrees, was perhaps the weakest indicator.

Table 3.14 RTO First Look: Criterion Measures with Effect Size for Power = .80

Measures	Effect Size
Mean Absolute Lateral Deviation (ft)	10
RMS Lateral Deviation (ft)	13
Integrated Yaw Activity (deg)	26
RMS Heading Deviation (deg)	1.54
Maximum Heading Deviation (deg)	4.15
Power Lever Reaction Time (sec)	0.81
Pedal Reaction Time (sec)	0.91
STD Pedal Position (in)	0.585
Number of Pedal Reversals	1.72
Sum of Pedal Spectrum f>0.5 Hz (in.sec)	0.325

Figure 3.13 shows the First Look evaluation results for the directional measures, which have been identified as criterion measured both by the PTS and by the regression analyses with I/E grades. In this and all subsequent figures, the mean for each group is shown with the standard error bars. The presence of platform motion significantly improved Integrated Yaw Activity of pilots ($F(1, 28) = 5.028$, $p = 0.033$). Visual inspection suggests that the Motion group also performed better in RMS and Maximum Heading Deviations; however these differences were not significant ($p > 0.1$). For lateral deviation, however, which, as indicated both by the PTS and the regression analyses, may be the most important performance measure for RTO, both groups performed equivalently.

Figure 3.14 shows that Power Lever and Pedal Reaction Time was not affected by motion status ($F(1, 28) = 0.829$, $p = 0.370$), despite the assumption of an early alerting function of motion and the excellent resolution for this measure.

None of the other performance measures showed any effects of motion, nor did any of the workload measures (see Appendix N).

Thus, for First Look evaluation of RTOs, motion affected one aspect of the directional performance of the pilots. This effect, however, was not found on measures having strong influences on the I/E grades, i.e. heading and lateral deviations (see Section 3.2.5). Therefore, it appears that First Look evaluation would not likely be affected by the presence of platform motion, as was also confirmed by the subjective I/E grades.

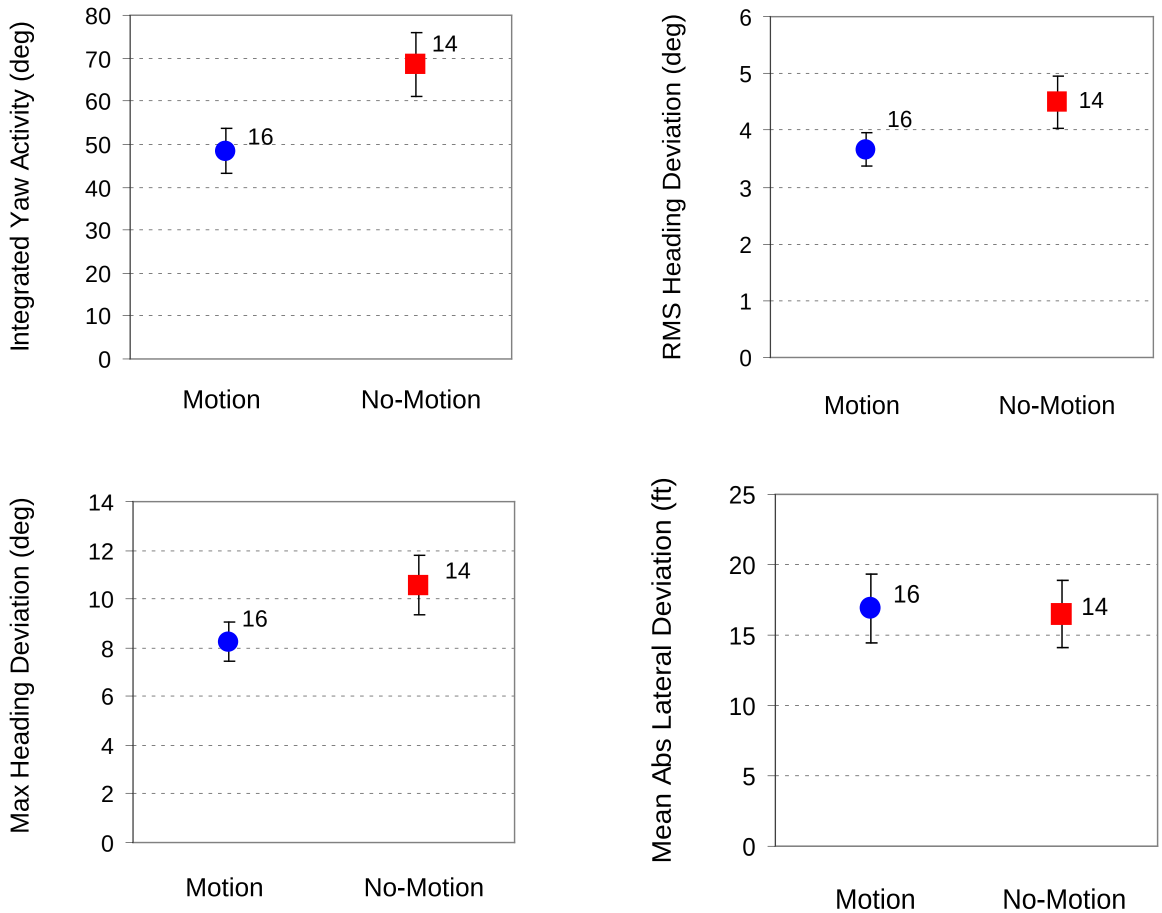


Figure 3.13 Directional performance on RTOs during First Look as a function of measure and motion group (Numbers indicate *n*, error bars indicate standard error).

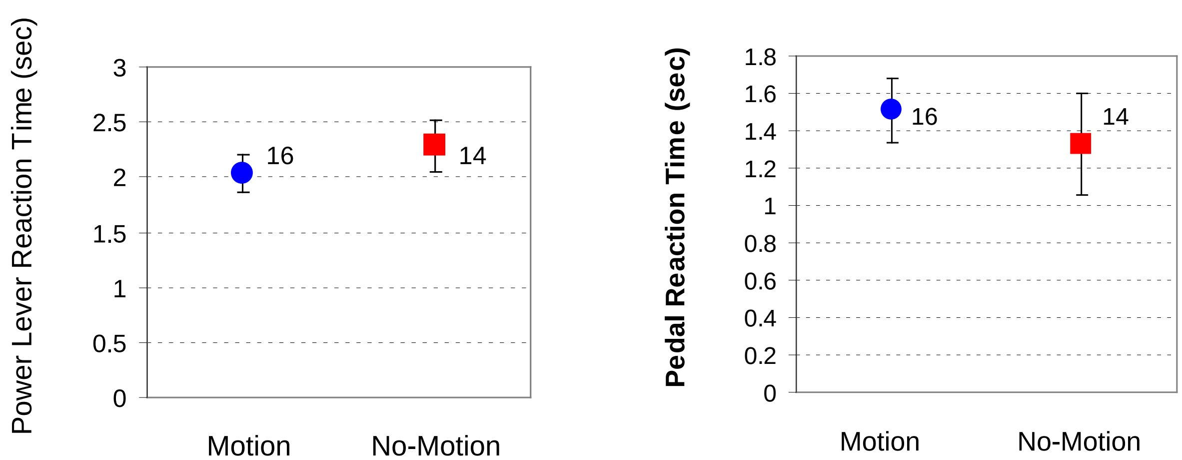


Figure 3.14 Power Lever Reaction Time and Pedal Reaction Time for RTOs during First Look as a function of motion group (Numbers indicate *n*, error bars indicate standard error).

V_1 Cut

The resolution found with 18 crews in the Motion group and 19 in the No-Motion group for the criterion measures is presented in Table 3.15. The smallest effect sizes that could be detected by the analysis should be sufficient to capture any operationally relevant differences.

Table 3.15 V_1 Cut First Look: Criterion Measures with Effect Size for Power = 0.80

Measures	Effect Size
Maximum Absolute Bank Angle (deg)	3.95
Integrated Bank Angle Exceedance (deg.sec)	14.9
Integrated Failure Induced Bank Angle (deg.sec)	26.4
Integrated Roll Activity (deg)	21.5
Integrated Yaw Activity (deg)	7.3
RMS Heading Deviation (deg)	2.77
Maximum Heading Deviation (deg)	4.05
Integrated Heading Exceedance (deg.sec)	28
Integrated Airspeed Exceedance (kts.sec)	43
Time to Reach 400 ft Altitude (sec)	7
STD Pitch Angle (deg)	1.15
Wheel Reaction Time (sec)	1.12
Pedal Reaction Time (sec)	0.97
STD Column Position (in)	0.203
STD Wheel Position (deg)	2.85
STD Pedal Position (in)	0.26
Number of Column Reversals	1.28
Number of Wheel Reversals	1.4
Number of Pedal Reversals	0.82
Sum of Column Spectrum f>0.5 Hz	0.209
Sum of Wheel Spectrum f>0.5 Hz	0.614
Sum of Pedal Spectrum f>0.5 Hz	0.899

No statistically significant differences for either performance or workload measures were found between groups as a function of motion for First Look evaluation of V_1 cuts (Appendix O), although the Motion group was found to control pitch angle marginally more steadily than the No-Motion group ($F(1, 35) = 2.923$, $p = .096$). Physically, however, this difference was less than one degree in average STD (Figure 3.15). Moreover, this slight advantage in pitch control was not accompanied by improvements in any of the other longitudinal performance measures.

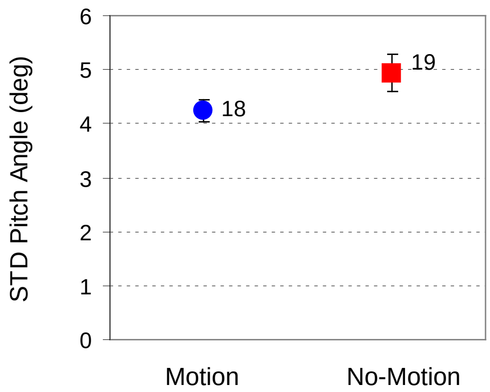


Figure 3.15 STD Pitch Angle for V_1 Cuts during First Look as a function of motion group (Numbers indicate *n*, error bars indicate standard error).

3.2.7 Transfer

The purpose of the analysis was to examine whether there was any difference in the transfer of skills to the airplane between the two groups of pilots trained with and without motion. Transfer was tested by having the two groups, one trained with motion and the other without, perform RTO and V_1 cut maneuvers in the simulator with the motion system turned on, and then comparing their performance and workload for the criterion measures. In this "quasi-transfer," the simulator with the motion system activated served as a stand-in for the airplane.

RTO

Table 3.16 RTO Transfer: Criterion Measures with Effect Size for Power = 0.80

Measures	Effect Size
Mean Absolute Lateral Deviation (ft)	8.7
RMS Lateral Deviation (ft)	11.6
Integrated Yaw Activity (deg)	20.5
RMS Heading Deviation (deg)	2.52
Maximum Heading Deviation (deg)	2.94
Power Lever Reaction Time (sec)	0.87
Pedal Reaction Time (sec)	0.46
STD Pedal Position (in)	0.54
Number of Pedal Reversals	2.2
Sum of Pedal Spectrum f>0.5 Hz (in.sec)	0.42

The smallest effect sizes that could be detected with the 16 Motion and 14 No-Motion crews are given in Table 3.16. The statistical resolution was again sufficient to capture any operationally relevant effects; in fact, for many of the measures it was actually higher than in the

First Look evaluation. Apparently, some of the variability within groups was reduced during Training, thus increasing power.

No significant differences between the two groups were found for any of the performance or workload measures, indicating that the motion status of the simulator during Training had no effect on either performance or workload of pilots once they transferred to the simulator with motion as a stand-in for the airplane (all $p > .19$; see Appendix P for summary data and graphs). This lack of a difference between the two groups for RTO measures is especially significant when considering the fact that the Motion crews were trained and tested on the same simulator configuration. Thus, the motion status during Training did not affect the transfer of skills for RTO maneuvers in this quasi-transfer experiment.

V_1 Cut

The statistical resolution obtained with 18 Motion and 19 No-Motion crews for V_1 cut Transfer can be seen in Table 3.17.

Table 3.17 V_1 Cut Transfer: Criterion Measures with Effect Size for Power = 0.80

Measures	Effect Size
Maximum Absolute Bank Angle (deg)	4.18
Integrated Bank Angle Exceedance (deg.sec)	10.7
Integrated Failure Induced Bank Angle (deg.sec)	18
Integrated Roll Activity (deg)	14.5
Integrated Yaw Activity (deg)	9
RMS Heading Deviation (deg)	3.15
Maximum Heading Deviation (deg)	5.1
Integrated Heading Exceedance (deg.sec)	32
Integrated Airspeed Exceedance (kts.sec)	23.7
Time to Reach 400 ft Altitude (sec)	8.8
STD Pitch Angle (deg)	0.75
Wheel Reaction Time (sec)	0.95
Pedal Reaction Time (sec)	1.26
STD Column Position (in)	0.16
STD Wheel Position (deg)	2.37
STD Pedal Position (in)	0.293
Number of Column Reversals	1.16
Number of Wheel Reversals	1.45
Number of Pedal Reversals	0.62
Sum of Column Spectrum f>0.5 Hz	0.43
Sum of Wheel Spectrum f>0.5 Hz	33.2
Sum of Pedal Spectrum f>0.5 Hz	0.152

As for RTOs, the power to differentiate between the two groups was increased especially for the measures that strongly influenced I/E grades.

In terms of performance, the most notable differences between the two groups were on Integrated Airspeed Exceedance and STD Pitch Angle. The Motion group controlled airspeed better ($F(1, 32) = 8.859$, $p = .006$) at the expense of increased STD Pitch Angle ($F(1, 32) = 5.508$, $p = .025$) (see Figure 3.16). Physically this can be interpreted as the Motion group controlling airspeed more successfully by adjusting pitch angle more aggressively than the No-

Motion group. Note that speed control is critical in V_1 cuts, because it involves safety (e.g. for obstacle clearance). The Motion group also displayed higher Integrated Yaw Activity compared to the No-Motion group ($F(1, 32) = 5.621$, $p = .024$) (see Figure 3.17). However, this did not appear to result in any differences in heading control or other directional performance measures. No other statistically significant performance differences were found (see Appendix Q for remaining summary statistics and graphs).

The workload measures revealed some interesting differences between the Motion and No-Motion groups. The Motion group had fewer Wheel Reversals than the No-Motion group ($F(1, 32) = 3.825$, $p = .059$), whereas the No-Motion group had fewer Pedal Reversals than the Motion group ($F(1, 32) = 8.038$, $p = .008$) (see Figure 3.18). Note that although these differences were statistically significant, they represented an average increase or decrease in less than one reversal within the 15-second period following an engine failure.

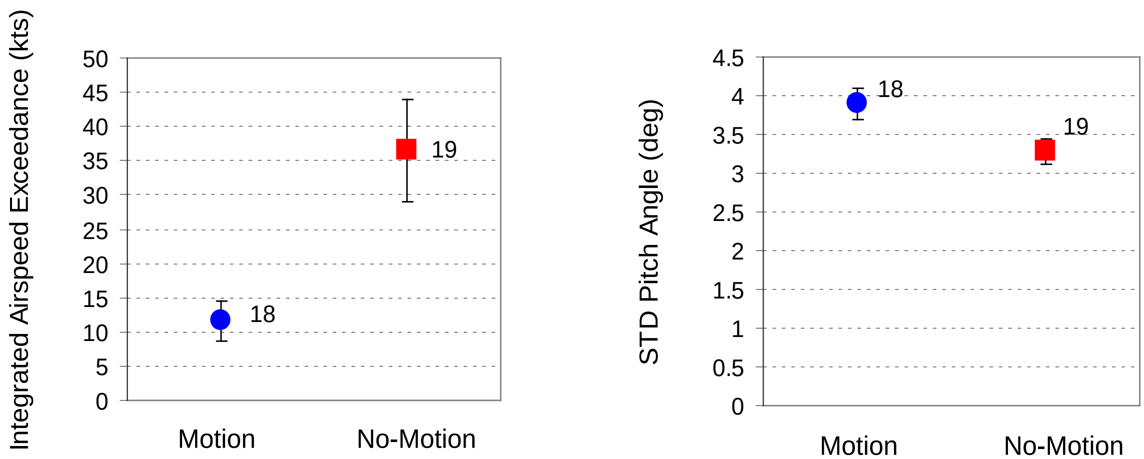


Figure 3.16 Longitudinal performance on V_1 Cuts during Transfer as a function of measure and motion group (Numbers indicate *n*, error bars indicate standard error).

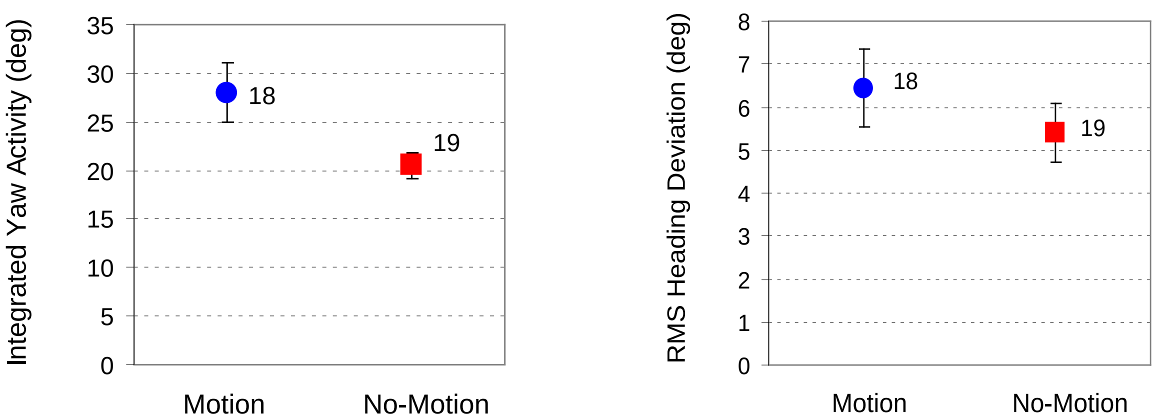


Figure 3.17 Directional performance on V_1 Cuts during Transfer as a function of measure and motion group (Numbers indicate *n*, error bars indicate standard error).

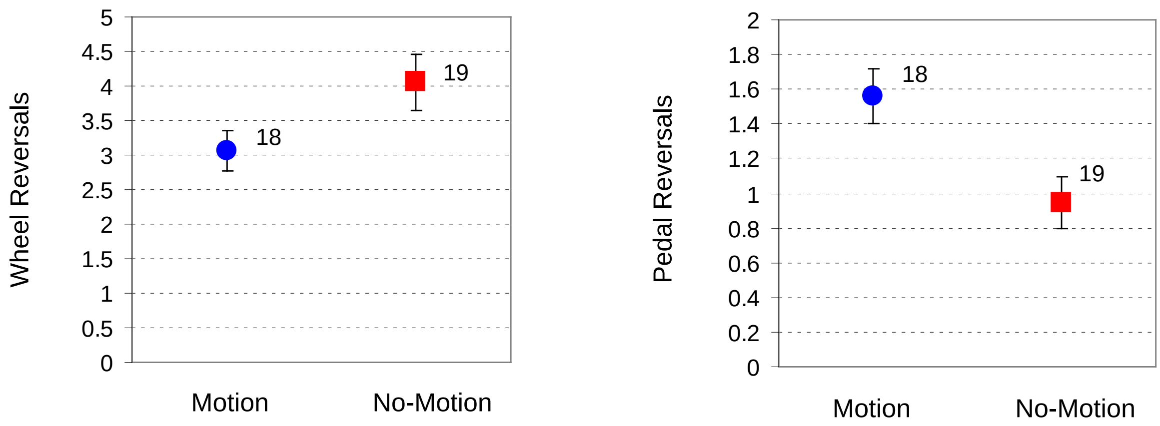


Figure 3.18 Wheel and Pedal Reversals for V_1 Cuts during Transfer as a function of motion group (Numbers indicate *n*, error bars indicate standard error).

The increased number of Wheel Reversals of the No-Motion group was not accompanied by any lateral performance differences. The increased number of Pedal Reversals of the Motion group, however, was accompanied by an increase in Integrated Yaw Activity, as was discussed earlier. This fact suggests that at least after Training, the group trained with motion worked harder to achieve lateral-directional control than the group trained without motion. These differences were not statistically significant during First Look evaluation. The graph for Pedal Reversals shows that this may be due to the variability within the Motion group (see Appendix O). As can be seen in Figure 3.18, this variability is reduced and the resolution was slightly increased during Transfer.

Because of the group difference in Wheel and Pedal Reversals, the pilots' responses to the related questions of rudder and aileron in the questionnaires were examined (see Figure 3.19). Indeed, the pilots' opinions on the control of the rudder and aileron were consistent with their use of the wheel and pedal. The Motion PF rated rudder as significantly harder to control than aileron ($F(1, 36) = 4.840$, $p = .034$), while the No-Motion PF rated the rudder and aileron as about the same. The difference in how the two groups perceived rudder and aileron might have driven them to use a different control strategy.

Although a few statistically significant differences between the groups trained with and without motion were found during Transfer testing of V_1 cut maneuvers with motion, the size of these differences raises questions about their operational relevance. On average, they were only about 1.5 knots exceedance per second for airspeed, half a degree RMS for pitch angle deviation, and half a degree per second for yaw rate. Such differences are very small compared to about 110 knots desired nominal airspeed and about 10 degrees nominal pitch angle.

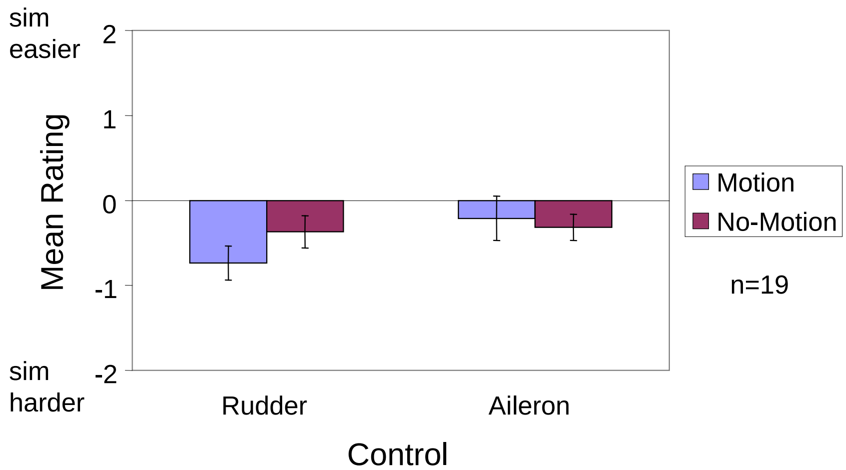


Figure 3.19 Mean rating given by the PFs of the controls as a function of motion group (Error bars indicate standard error, the comparison was the airplane).

3.2.8 Training Progress

The effect of the platform motion on the course of Training was analyzed by using nonparametric analyses. Comparison of the data from the first and the last Training trials was used as a basis to evaluate the Training progress. The eligible crews for this comparison were the ones having complete objective data for both the first and the last Training trials. The difference in percentage of crews that improved between groups was examined for each measure. Because in some cases, the sample size was less than 20, Fisher Exact probability test was utilized for the analyses (Siegel, 1956). Within each group, the significance of the number of crews who improved was also evaluated, using a goodness-of-fit test, which is basically a Chi-Square one-sample test (Siegel, 1956).

RTO

The statistics and graphs for the comparison of the RTO measures between the first and the last Training trials are presented in Appendix R. No statistically significant differences in improvement from first to last Training trial were found between groups for any of the measures (all $p \geq .1$). This suggests that the platform motion did not affect the Training progress of the pilots.

Also, as can be observed from the data, the overall number of crews (Motion and No-Motion) improving in most lateral performance and workload measures was significant for most measures, with the exception of Integrated Yaw Activity with no overall improvement and Pedal Reversals, which actually increased after Training. When looking at the groups separately for these two measures, neither of the groups shows any improvement or deterioration. This shows that the pilots generally did improve during Training regardless of the motion status of the simulator.

V_1 Cut

For longitudinal control during V_1 cuts, motion improved Training progress for speed control (Integrated Airspeed Exceedance), but at the cost of pitch angle control (STD Pitch

Angle). Progress in directional control (i.e., RMS Heading Deviation, Integrated Heading Exceedance, and Maximum Heading Deviation) was also negatively affected by the presence of motion during Training ($p < .1$; see Figure 3.20). The Training progress on lateral control was not affected by the presence or absence of motion. Also, there was no difference for workload between the two motion groups. The statistics and graphs for the comparison of the measures between the first and the last Training trials are given in Appendix S.

The appendix indicates that the No-Motion group improved on more measures than the Motion group. While Motion crews improved in Integrated Airspeed Exceedance and STD Column Position only, the No-Motion crews improved in Integrated Bank Angle Exceedance, Heading Deviation, Time to Reach 400 ft Altitude, and STD Pitch Angle. During Transfer, however, the No-Motion group surpassed the Motion group only with steadier pitch angle and yaw activity; and the actual size of these differences was very small.

The above discussion indicates that the Training without motion was at least as effective the Training with motion, and the earlier results on Transfer show that although some differences were found in Training progress between the two groups, they did not translate into operationally relevant differences during Transfer.

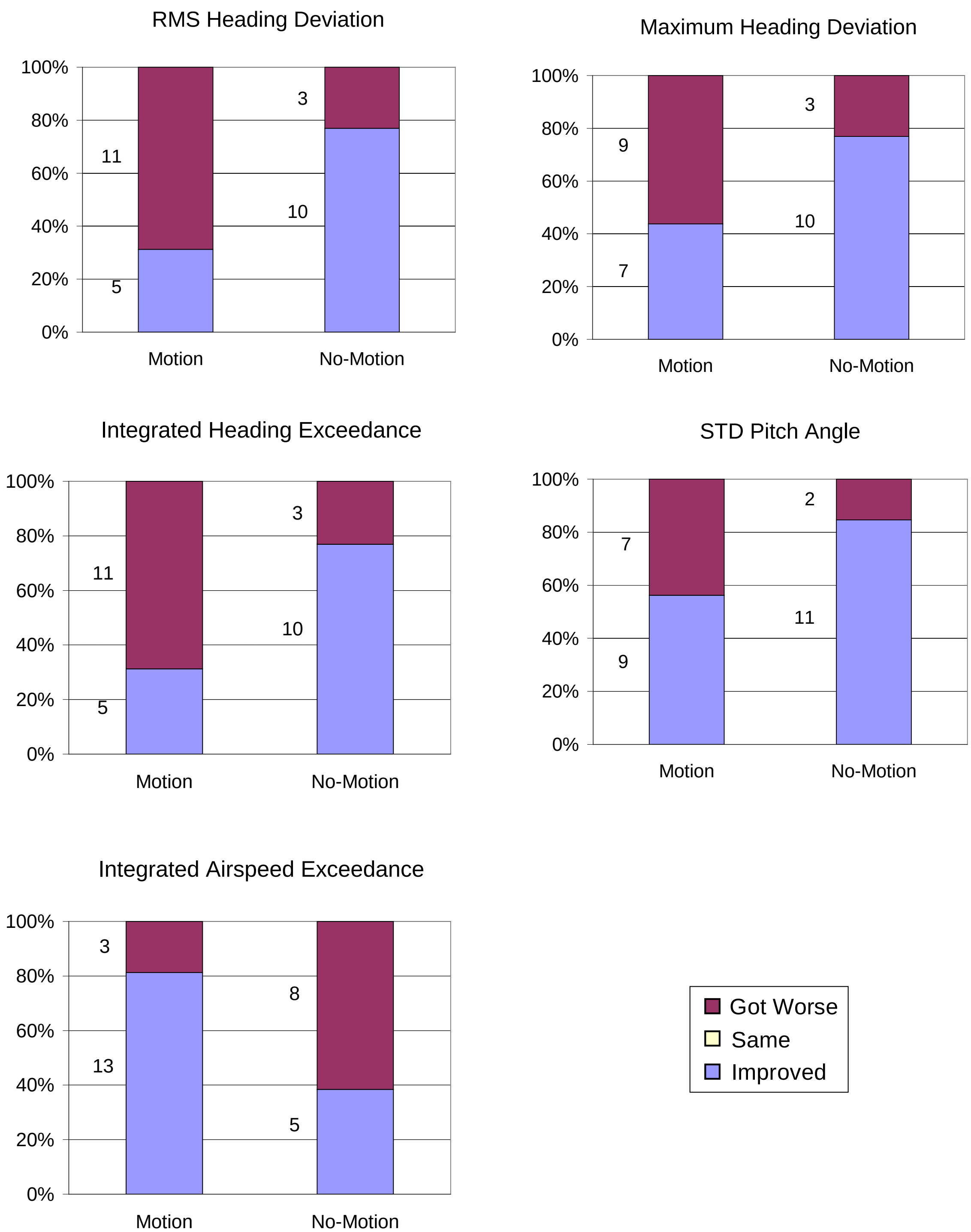


Figure 3.20 Percentage of crews getting worse, staying the same, or improving from first Training to last Training in performance on V_1 Cuts as a function of measure and motion group (Numbers indicate *n*).

3.2.9 Improvement From Last Training Trial to Transfer Testing

The effect of adding motion from the last Training trial to Transfer testing was again examined by using nonparametric analyses. The main interest here is whether there was any difference between the two groups in the percentage of crews who improved from the last Training trial to Transfer testing for each measure. As for the subjective results, this may be regarded as a true test of transfer, because even without operationally relevant difference between the two groups during the Transfer trials *per se*, there may still be a difference between the groups in how they transition from the last Training (with or without motion) to the simulator with motion. Note that in the counting for the nonparametric analysis, only crews who had complete objective data on Transfer testing and on the last Training trial were included. Fisher Exact probability test was utilized in the analysis.

RTO

For RTOs, the differences in the number of crews improved from the last Training trial to Transfer testing between groups were significant only for some directional performance measures: Integrated Yaw Activity (N = 30, Fisher Exact p = .014) and Maximum Heading Deviation (N = 30, Fisher Exact p = .033) (see Figure 3.21).

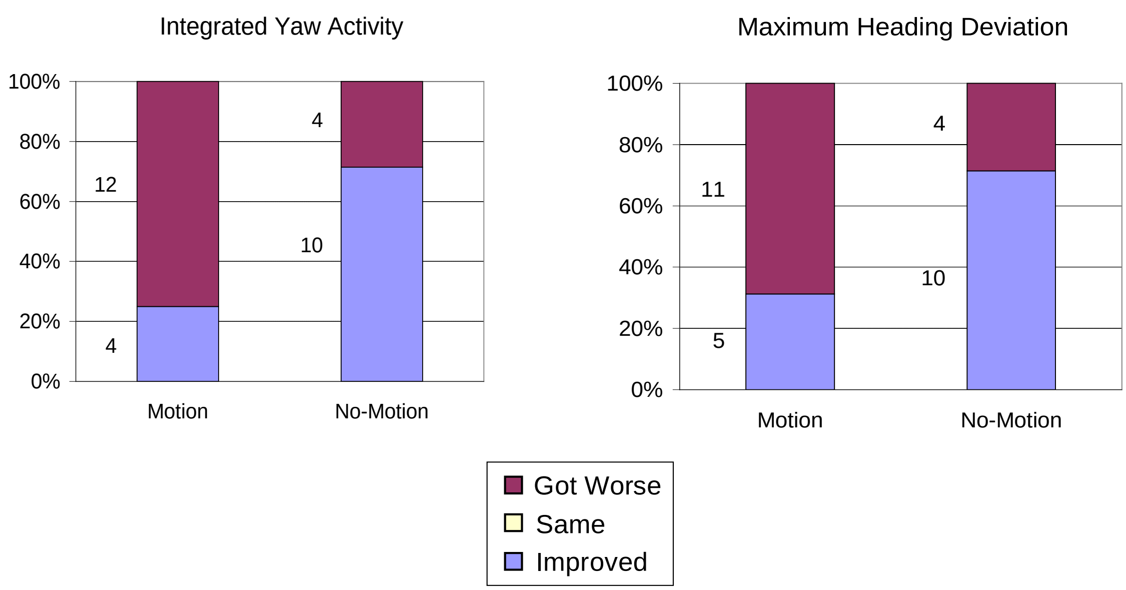


Figure 3.21 Percentage of crews getting worse, staying the same, or improving from last Training to Transfer in directional performance on RTOs as a function of measure and motion group (Numbers indicate *n*).

For both of these measures, more No-Motion crews improved from last Training trial to Transfer testing than Motion crews. No other differences were found for RTO maneuvers, as can be seen from the statistics and graphs presented in Appendix T.

This result indicated that the addition of motion was beneficial to the No-Motion group. However, this benefit did not carry through to the Transfer testing, where no statistical difference in the performance level between the two groups was found, as has been reported previously.

V$_1$ Cut

For V$_1$ cuts, statistical differences in improvement from last Training to Transfer between groups were found for several measures (i.e., Integrated Yaw Activity, Wheel Reaction Time, STD Pedal Position, and Sum of Column Spectrum f > .5 Hz) (see Figure 3.22). Specifically, from the last Training trial to Transfer testing, there are more Motion crews who improved in Wheel Reaction Time (N = 34, Fisher Exact p = .087) and Sum of Column Spectrum f > .5 Hz (N = 34, Fisher Exact p = .02) compared to No-Motion crews. For Integrated Yaw Activity and STD Pedal Position, however, more No-Motion crews improved from last Training to Transfer than Motion crews (N = 34, Fisher Exact p < .02). No other differences were found between the two motion groups for performance or workload measures (see Appendix U).

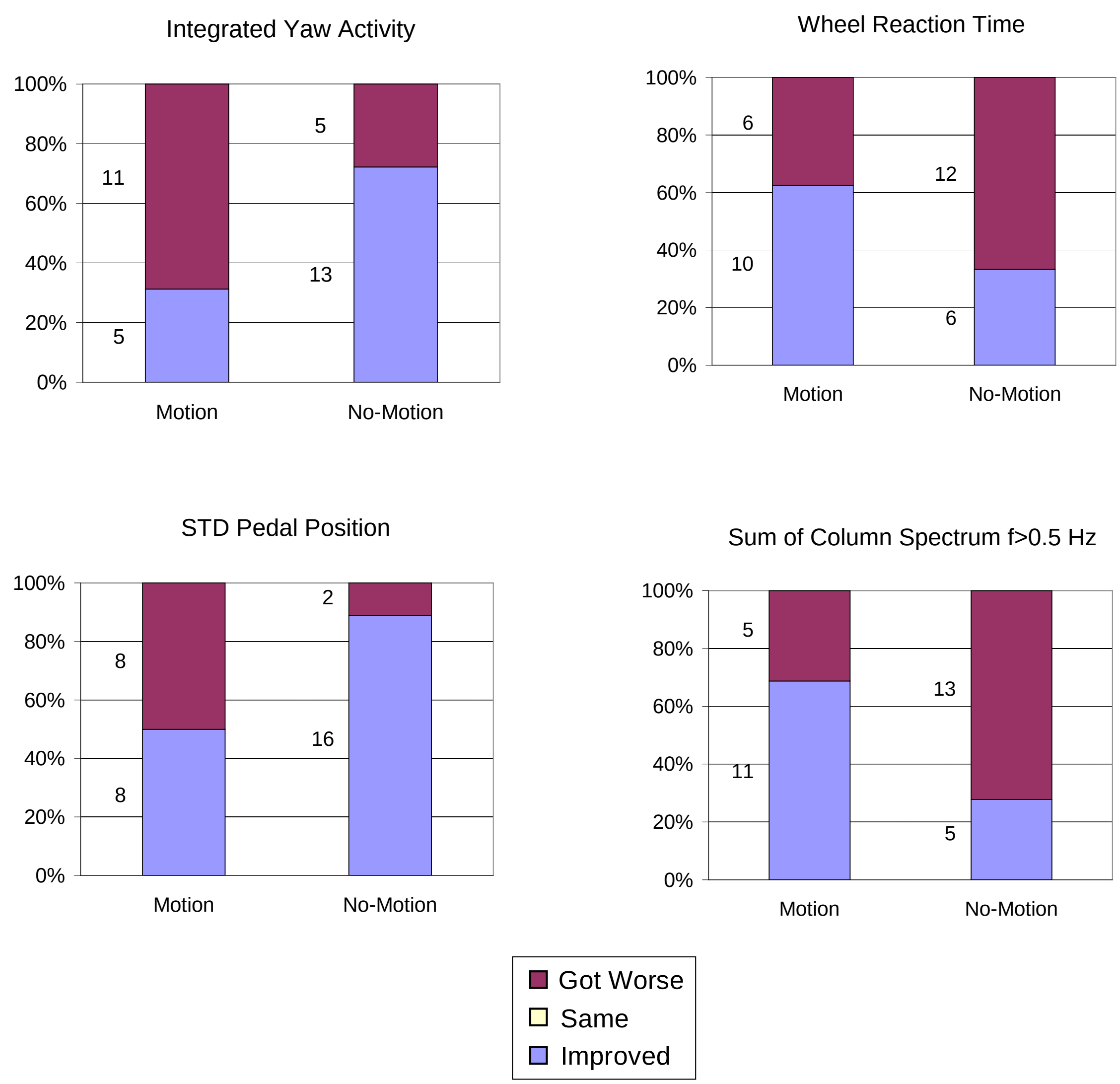


Figure 3.22 Percentage of crews getting worse, staying the same, or improving from last Training to Transfer in lateral directional measures on V$_1$ Cuts as a function of measure and motion group (Numbers indicate *n*).

Unlike for the RTO maneuvers, the difference in improvement from the last Training trial to Transfer for Integrated Yaw Activity did translate into a difference in Transfer testing performance. It should be kept in mind, however, that the difference, although significant statistically, was quite small and may not matter operationally.

4. ASSESSMENT OF THE REPRESENTATIVENESS OF THE TEST SIMULATOR

4.1 INTRODUCTION

As mentioned in Chapter 3, a relatively large discrepancy was observed between the magnitude of the lateral acceleration produced by the test simulator and the lateral acceleration from the equations of motion during the first few seconds following the engine failure. Visual inspection of the measured lateral acceleration response, or the motion drive equation outputs in this matter, did not easily reveal failure-induced acceleration--unlike inspection of the response obtained from the equation of motion. Because lateral acceleration cues might be important for flying the type of maneuvers tested, this finding leads to the question whether the simulator used in the experiment was typical in regard to the generation of the lateral motion. The answer to this question will also determine the scope of the applicability or generalizability of the current study's results.

One the one hand, some simulator experts suggest that it may be difficult for a hexapod motion platform, such as the one of the test simulator, to generate sufficient lateral acceleration in the pitch-up position during take-off, assuming that the travel available to the associated motion platform actuators is greatly restricted in this position. Moreover, motion drive algorithms generally tend to increasingly attenuate the motion as travel limits are approached. It is hard to know, however, whether and how this generally applies, given that much of the information necessary to perform this analysis is proprietary. On the other hand, because the current simulator qualification procedures do not provide a means to objectively assess the quality of the produced motion (Lahiri, 2000), it is possible that the motion provided by the test simulator may be atypical of other similarly qualified simulators.

To investigate whether the test simulator was representative of other FAA Level C simulators, especially in regard to its lateral acceleration during V_1 cut, data was requested from a representative sample of other FAA qualified level C and D simulators and compared to similar data from the test simulator. The purpose of the comparison is twofold: to examine whether the level of lateral acceleration generated by the test simulator is typical and to see whether the discrepancy between the lateral accelerations from the equations of motion and from the motion drive equations is common.

4.2 MOTION DATA COLLECTION PROCESS

With the help from the National Simulator Program Office (NSPO), in January 2000, the regular attendees of the Simulator Technical Issues Group (STIG) meeting, who are mostly the simulator operators, were briefed on the FAA need to gather simulator data. They were told that the data would be used to improve the simulator standards. It was hoped that this briefing would discourage the simulator operators from the common practice of reducing the gains of the motion system while recording the requested data. On February 15, 2000, a formal letter requesting the recording of accelerations from simulator tests was sent to the meeting attendees by the NSPO Manager.

The data of requested were the acceleration values from the equations of motion (EOM) output and the motion drive acceleration commands just prior to the application of the washout algorithm. The EOM values should reflect the accelerations at the simulated airplane cg, as corrected for cg location relative to the aerodynamic center, with the aircraft at a normal angle of attack for the flight condition, that is, rectilinear and rotational accelerations along the x, y, and z body axes. The motion system drive accelerations should be the commanded rectilinear and rotational accelerations at the pilot station. Hence, these values should represent the EOM values after being operated upon by the transformation algorithm that translates the motion parameter values from the cg to the pilot station. They should roughly reflect the amount of motion experienced by the pilot in the simulator. Because the operators were briefed on the purpose of the data collection, it can be assumed that full motion gain was used in recording the data. The maneuvers to use for the recording of the six accelerations above were V_1 cut and Dutch roll in the approach and landing configuration. The selection of V_1 cut for the data recording was to reproduce one of the maneuvers used in the experiment. The Dutch roll was chosen to provide accelerations in the free response mode and would be used to examine the motion system performance while the simulator actuators were unrestricted by the pitch-up position.

The letters requested acceleration data from 30 flight simulators, which were selected randomly from a pool of 116 eligible simulators satisfying the following criteria:

1) AC120-40B (or 40C) qualification basis
2) Simulator operator has not asked to be excluded from the data collection effort
3) Flight simulator located within North America
4) The flight simulator represents an airplane with wing-mounted engines.

The fourth criterion indicates that only flight simulators representing airplanes with similar engine configurations to the airplane simulated during the study were of interest. Due to slow response from the simulator operators, a follow up request was made in August 2000. Only 9 simulator data sets have been gathered by the NSPO to date. Initial analysis was performed on 8 of those sets (Boothe, 2000). Unfortunately, some data sets were lost as a consequence of personnel changes at the NSPO, and only 4 sets were available for further examination.

4.3 MOTION DATA

The motion systems of the simulators generally performed well in representing the aircraft motion on the longitudinal axis, as indicated by the closeness of the longitudinal acceleration traces from the equations of motion and from the motion drive equation. Because of the physical limitation of the simulator motion system, motion in the vertical axis was usually limited to the high frequency only up to the dynamic bandwidth of the motion system. This fact is well known and common to all simulators with the hexapod-type motion platforms commonly used in commercial 6 DOF simulators, including the group of simulators analyzed here. For these reasons, the focus of the investigation was mainly on the lateral motion performance.

4.3.1 Test Simulator

Examination of the motion drive equations output data of the test simulator, which reflect the measurement data taken at the pilot station, revealed that the simulator motion system only generated a maximum peak lateral acceleration of about 0.1 g during the RTO and V_1 cut maneuvers. This level of lateral acceleration was often much less than what the equations of motion (at the airplane cg) suggested. For example, for V_1 cut maneuvers, the equations of motion at cg might produce peak lateral acceleration of about 0.2 or 0.3 g. Moreover, one could not easily distinguish the peak of failure-induced lateral acceleration from the measurement data. Such peak was easily distinguishable in the output of the equations of motion.

4.3.2 Other Simulators

Table 4.1 and 4.2 summarize the analysis of the lateral acceleration data for V_1 cut and Dutch Roll maneuver. It should be noted that due to simulator time constraints, the operators obtained the data by running the maneuvers with the automatic testing feature of the simulators. Lateral acceleration data from the test simulator are also included in Table 4.1, obtained both from automatic testing and the experiment. There are several aspects that need to be kept in mind when reading the table. First of all, the simulators came from various manufacturers and as such the capability of accessing the data at various points in the simulation loop varied between simulators. Second, the engine failure profiles used in the automatic testing vary greatly from one simulator to another, and these profiles affect the rate of change of momentum of the aircraft, which in turn, influences the resulting yaw and lateral acceleration.

Table 4.1 V_1 Cut Data from Several Simulators

Simulator	Aircraft Weight (lbs)	Engine Type	Engine Failure Speed (kts)	Failed Engine Power Decay Time (s)	Maximum Failure-Induced Lateral Acceleration from EOM at cg (g)	Maximum Failure-Induced Lateral Acceleration from motion drive equations at pilot station (g)
B737-200	99330	Turbofan	135	7.6	0.078	
B737-800	151699	Turbofan	129	8.9	0.062	
A-320#1	141975	Turbofan	131	14.3	0.04	0.04
B747-400	626400	Turbofan	125	7.0	0.071	0.070
B737-300*		Turbofan	118		0.062	0.047
A-330-300*		Turbofan	135		0.065	0.059
B757*		Turbofan	120		0.002	0.003
SAAB 340*		Turboprop	117		0.078	0.025
Test sim (auto test)	17893	Turboprop	84	1.2	0.1	0.06
Test sim (from experiment)#	20500	Turboprop	110	1.2	0.21	0.069

*From initial analysis only (Boothe, 2000).
#For this comparison, a V_1 cut maneuver with grade 3 is used.
Note: Blank cells on the table indicate the data are not available.

Table 4.2 Dutch Roll Data from Several Simulators

Simulator	Aircraft Weight (lbs)	Airspeed (kts)	Maximum Lateral Acceleration from EOM at CG (g)	Maximum Lateral Acceleration from motion drive equations at pilot station (g)
B737-200	87557	120	0.11	
B737-800	133622	160	0.11	
A-320#1	132094	145	0.12	0.02
B747-400	626400			
B737-300*			0.12	0.093
A-330-300*			0.093	0.031
B757*			0.19	0.078
SAAB 340*			0.2	0.001

*From initial analysis only (Boothe, 2000)
Note: Blank cells indicate the data are not available.

As can be seen from Table 4.1, the values of the maximum failure-induced lateral acceleration from the equations of motion at cg as obtained from the automatic testing, mostly fall within .04 and .1 g. An exception is the value reported for the B757 simulator, which is unusually low (.002 g). Unfortunately, this information could not be studied further due to the data loss. The values of the maximum failure-induced lateral acceleration from the equations of motion of the test simulator obtained in the experiment are in general higher than the values from the automatic testing (.2 to .3 g). The values of the failure-induced lateral acceleration from the motion drive equations are in general about the same or lower than the respective values from the equations of motion. With the exception of the B757 simulator, these values range from .025 to .07 g. Again, the value of the lateral acceleration from the motion drive equations of the B757 simulator is unusually low. The discrepancy between the maximum failure-induced acceleration from the equations of motion and from the motion drive equations is relatively large for the Saab 340 simulator and the test simulator, especially for the values obtained from the experiment.

As can be seen from Table 4.2, the values of the maximum lateral acceleration from the equations of motion for Dutch roll are in general higher than the values for V_1 cut. They range from .093 to .2 g. The values of the maximum lateral acceleration values from the motion drive equations, however, are not consistently higher than those of the V_1 cut. The motion drive equations commanded higher lateral acceleration values during Dutch roll only for the B737-300 and the B757 simulators.

4.4 DISCUSSION

It was true that the level of peak lateral acceleration produced by the test simulator was above the human perception threshold, which is about .005 g for the frequencies typical of piloting (Peters, 1969; Zaichik, et al. (1999)). However, it was unknown whether such level of lateral acceleration was effective in generating the sensation approaching that experienced in the airplane, especially in light of the discrepancy observed with the outputs of the equations of motion from the experiment. Also, it is unknown whether the level of acceleration generated is perceptible by pilots in the presence of background acceleration noise and concurrent tasks competing for their attention. Although the current effort does not provide answers to all these questions, it should be clear that more research is needed to fully understand the necessary level of simulation physical fidelity, regardless of whether the test simulator motion is representative.

As mentioned earlier, V_1 cut and Dutch roll data from eight other level C and D simulators have been analyzed for comparison. The following observations can be drawn from the V_1 cut analysis in Table 4.1:

- The peak of the failure-induced lateral acceleration demanded by the equations of motion of the test simulator (.21 g), as obtained from the experiment, is much higher than the others in the table, which were obtained from automatic testing (<.09 g). There are two possible explanations for this. First, this large peak might be due to the combination of the relatively large ratio of the engine yawing torque to the yaw moment of inertia of the airplane and the fast decay time of the failed engine power of the test simulator. Without further airplane data, however, it is not possible to check whether the values on the table are consistent with this conjecture. Second, the control response in the automatic testing feature might reflect near ideal response to engine failure, hence preventing the occurrence of large lateral acceleration. In the experiment, the pilot might not have responded to the engine failure fast enough, resulting in high failure-induced lateral acceleration. The data shown in Table 1 suggest that this is likely the case.
- The test simulator is not the only simulator that shows a relatively large discrepancy between the magnitude of the lateral accelerations from the equations of motion and the motion drive equations for V_1 cut maneuver (about 3 to 1 ratio). The Saab 340 simulator in the above list exhibits a similar trend, with a ratio of about 3 between the lateral acceleration from the equations of motion and from the motion drive equations. The discrepancy might indicate the attenuation applied to the motion system commands to keep the actuator travels within their limit.

The magnitude of the peak lateral acceleration from the motion drive equations of the test simulator is comparable to the others in the table. It may not follow the peak suggested by the equations of motion well, but the level commanded by the motion drive equations is definitely typical in comparison to V_1 cut maneuver data from the other simulators. Generally, the outputs of the motion drive equations reflect the actual motion produced by the motion system after some filtering.

The collection of the Dutch roll data was intended to see whether larger lateral motion commands are given by the motion drive equations during Dutch roll simulation than during the V_1 cut. This is because Dutch roll usually results in larger lateral acceleration than V_1 cut, as can be seen from the equations of motion readings in Table 4.2. However, except for the B737-300 and B757 simulators, the lateral motion drive outputs during Dutch roll are more suppressed than during V_1 cut. The suppression might be because the Dutch roll frequencies are below the frequencies passed by the actuator travel limiter in the form of high pass filter, which is commonly used in a hexapod motion system. These Dutch roll data support the finding that relatively large discrepancy between lateral acceleration from the equations of motion and from the motion drive equations is not uncommon.

4.5 CONCLUDING REMARKS

As has been mentioned previously, this comparison is limited by the difficulties experienced in gathering the necessary data and the constraints associated with the information collected. Nevertheless, within these limitations, the comparison suggests that the level of lateral acceleration produced by the test simulator is not atypical. Even the relatively large discrepancy found between the magnitude of the lateral accelerations from the equations of motion and the motion drive equations is not unique to the test simulator. These findings suggest that in terms of the magnitude of the lateral acceleration produced, the motion of the test simulator may be considered to be representative of other FAA qualified Level C and D simulators.

5. FINAL REMARKS AND CONCLUSIONS

5.1 FINDINGS

5.1.1 From the Analysis of Grades and Questionnaire Data

Platform motion had no effect on the grades that were provided by the I/Es at any time for either the RTO or the V_1 cut or for the normal take-offs. That is, platform motion did not affect First Look evaluation in the simulator, nor did it affect the grades at Transfer to the simulator with motion. The latter was true when comparing the group means or the number of low vs. high grades in each group (i.e., grades of 1 and 2 vs. grades of 3 and 4). However, for V_1 cuts and at Transfer only, the crews that had previously had motion did receive more grades of 2 than the crews who had not previously had motion, and fewer grades of 1 (none actually). Despite this single effect of motion, there was no effect of motion on the course of Training or on the amount of Training required before reaching the criterion needed to move onto Transfer. Overall, the grades did show that the pilots improved across the maneuvers, but similarly for the two groups.

The pilots' and I/Es' questionnaire responses showed only a few and mostly contradictory motion group differences. The No-Motion PF rated their training progress higher than the Motion PFs rated theirs, but only when combining both questionnaires. They also rated their control precision higher than the Motion PFs did theirs, but only after Transfer (before Transfer, they were rated higher by their PNF than the Motion PFs were by theirs, but not by themselves). However, from the point of view of the I/Es, after Transfer the control performance of the Motion PFs was better than the one of the No-Motion PFs (rated by their respective I/Es). This was not reflected in the I/E grades, however, with the exception of the fewer unsatisfactory grades obtained by the Motion pilots after Transfer. Due to their contradictory nature, these findings do not support an advantage of either of the two configurations. All together the subjective responses of the pilots and the I/Es did not indicate that the motion used in this study had any impact on the pilots' performance. It also had very minimal impact on the pilots' perception of their own performance, workload, ability to gain proficiency, comfort, or their acceptability of the simulator.

In fact, the lack of a significant difference in ratings of acceptability supports Lee and Bussolari's findings (Lee & Bussolari, 1989) that pilots' acceptance of a simulator is not affected by the presence or absence of platform motion when the pilots are not made aware of the lack of motion, in contrast to their preference for motion when they are aware. This study even extends Bussolari's findings somewhat in that Bussolari did not have a real no-motion condition, but instead a condition in which a small amount of vibration (1 cm amplitude in heave) was used.

Finally, the subjective responses of all participants also suggest that the lack of platform motion does not increase simulator sickness (i.e., there was no difference across groups in ratings of comfort), despite the conflict between vestibular and visual sensations. Thus, the concern that the lack of motion might cause simulator sickness even if it does not help in training or evaluation has not been substantiated.

5.1.2 From the Analysis of Objective Data

First Look Evaluation

For RTO maneuvers, the Motion group had less Integrated Yaw Activity than the No-Motion group. No effect of motion, however, was found on any other performance or workload measures, including performance in heading and lateral deviations, which were strongly related to I/E grades. This suggests that the presence of motion did not affect First Look evaluation of RTOs in any operationally relevant manner.

For V_1 cut maneuvers, the only effect of motion found was slightly better pitch control, as manifested in a marginally lower STD Pitch Angle for the Motion pilots. This measure was used by I/Es as a grading criterion, especially when the motion was on, but only in combination with other longitudinal measures (see Section 3.2.4). Therefore, because the effect of motion on this measure was physically small and not accompanied by any other effects, the I/E grades would unlikely be affected by the motion system status, and in fact, they were not.

The above remarks on First Look evaluation suggest that the platform motion might have an effect on certain aspects of performance of the pilots. However, this effect was not operationally relevant nor did it manifest itself for the measures instructors focus on when grading. Thus, the presence or absence of platform motion does not appear to matter for pilot evaluation.

Transfer Testing

For RTO maneuvers, the performance and workload of the Motion and No-Motion groups did not exhibit any statistically meaningful differences, despite the fact that unlike the No-Motion group, the Motion group was trained and tested on the same simulator configuration. Also, there were more No-Motion crews than Motion crews who improved in heading control between the last Training and the Transfer testing, although during Transfer testing the two groups performed at the same level (as just described).

For V_1 cut maneuvers, the Motion crews had better speed control (i.e., Integrated Airspeed Exceedance) compared to the No-Motion crews. This came at the price of increasing pitch angle standard deviation, but it was still advantageous because of the critical role speed plays in safety, e.g. for clearing obstacles. There was also an increase in Integrated Yaw Activity shown by the motion-trained group, which did not appear to affect heading or other lateral performances. Note that these differences, although statistically significant, were actually quite small. With regard to workload during V_1 cuts, the motion-trained group had more rudder pedal reversals but fewer wheel reversals than the fixed-base trained group (with only minor differences in directional performance, however). The reason for this is not clear. The difference was not there at First Look, and a combined ANOVA of Motion/No-Motion by First Look vs. Transfer did not find a significant interaction. The questionnaire data indicated that this difference might be because the Motion group felt the pedal was less like the airplane than the No-Motion group did. This difference, although significant, represented less than one reversal over a 15 second time period.

The above discussion indicates that the platform motion had an effect on some aspects of pilot performance and workload in Transfer. It was not clear, however, whether these differences did in fact benefit pilot training in an operationally relevant manner, given the small physical size of the effect.

Training Progress

For RTO maneuvers, there were no differences in performance or workload improvement from first to last Training between the two motion groups.

For V_1 cut maneuvers, the results reflected the Transfer results. Motion did improve Training progress for Integrated Airspeed Exceedance, but hindered Training for STD Pitch Angle (the price for reduced Integrated Airspeed Exceedance) and for the heading variables. For workload variables, there were no differences between the two groups.

The above remarks suggest that the differences on the Training progress due to platform motion were found only on some measures and were not consistently beneficial. Moreover, these differences did not translate into operationally significant differences in Transfer testing, as discussed previously.

5.2 CONCLUSIONS

The objective findings indicate that the motion provided by the test simulator does not, in an operationally significant way, affect either pilot evaluation or pilot training in the simulator. That is, there was no effect of motion on First Look evaluation, Training progress, or Transfer of training acquired in the simulator with or without motion to the simulator with motion. As has been discussed, effects of motion were detected on some measures. However, these effects were mostly minimal, inconsistent and not always beneficial. Therefore, the effects of the motion provided by the test simulator for the RTO and V_1 cut tasks do not appear to carry any operational significance.

This conclusion is supported by the subjective data, which show that motion does not consistently affect the PFs', PNFs', and I/Es' subjective perception of the PFs' performance, workload, and Training progress, or of their own comfort in the simulator for the tasks tested. Neither does it affect the acceptability of the simulator by the PF and the PNF. As noted, however, the motion provided by the test simulator may or may not be typical of other FAA qualified Level C flight simulators. Comparison with data from some other FAA qualified Level C and D simulators, however, indicate that the motion of the test simulator may be considered representative.

5.3 LIMITATIONS AND RECOMMENDATIONS

This study was designed to maximize experimental power to detect any effects of motion. However, all research investigations have practical constraints. It is important that the potential users of research have a clear understanding of how they may limit both the generalizability of the results and their implications for the determination of regulations. The factors potentially limiting the generalizability of this study are, in order of importance:

1) the stimulation of the pilot by the motion of the test simulator;
2) the use of a quasi-transfer design; and
3) the particular maneuvers, pilot population, and airplane used during testing.

With regard to motion stimulation, the simulator used in this study may not have provided sufficient lateral motion to be effective. The measurements indicate that the lateral acceleration cuing provided by the simulator may not be enough to represent that of a real airplane for the test maneuvers. For the maneuvers considered, lateral acceleration may act as an important cue for proper failure recognition. Thus, the lack of it may represent a significant deficiency in pilot stimulation. However, this simulator is in current use as an FAA approved Level C flight simulator by an approved training center and the comparison study suggests that its motion platform performance may be considered representative of other similarly FAA qualified simulators. Therefore these results still suggest that the current requirement for platform motion, which does not specify the quality of motion cuing, is unnecessary for the tasks tested. They also suggest that the initial and life cycle costs associated with the requirement for motion may unnecessarily preclude certain sectors of the aviation community from the benefits of simulators.

Although it could be argued that the quasi-transfer design needs to be validated, the use of the simulator as a stand-in for the airplane was warranted for several reasons. First, the maneuvers that the subjects were trained and tested on (engine failures on take-off) were selected because they are safety-critical and diagnostic for a need of motion. However, repeated performance of this type of maneuver in an actual airplane could result in damage to the test airplane and injury to the crew and I/E. Second, to ensure that the study would find an effect of motion if there were one and to prevent motion-independent variables from causing spurious effects, it was important to reduce variability resulting from extraneous factors (e.g., weather, traffic, ATC vectors, etc.). Only by using a flight simulator was it possible to control these conditions appropriately. Third, the use of a Level C simulator as a stand-in for the airplane has been validated by the successful use of such simulators for total training and evaluation for many years. Additionally, the quasi-transfer paradigm has been used in prior research and as such facilitates data comparisons (e.g., Levison, 1981). Finally, given that the motion-trained group transferred to the same simulator configuration that they had been trained in, whereas the No-Motion group crews transferred to a configuration that was new to them (i.e., the motion configuration), the Motion group crews should have had an advantage. However, Training with motion did not improve their performance/behavior or reduce their workload for the tasks tested compared to the group trained without motion, and it is unlikely that they would have had an advantage had they transferred to an airplane.

With regard to the generalizability of the results to other maneuvers, pilot populations, and airplanes, every effort was taken to satisfy the criteria mentioned in the literature as diagnostic for the need of motion while preventing pilot adaptation to a particular simulator configuration. The pilot population was chosen for high proficiency with the airplane tested and as representative for pilots working for regional airlines. The airplane was chosen as representative for regional turboprop airplanes with wing-mounted engines maximizing the asymmetry experienced during engine failures. Nevertheless, any study attempting to prove the non-existence of a phenomenon may be subject to the criticism that a different configuration of the test, such as using different maneuvers, pilots, or airplane, would have revealed the existence of the phenomena.

Clearly, additional steps must be taken to validate the current study and to determine the extent to which it may or may not be appropriate to draw generalizations from these results. A first step, the comparison of the objective measures from the motion system used in this experiment with such measures taken from other FAA qualified Level C and D simulators, suggested that the motion used in the present study was representative. Another step should be an experiment in which the motion will be manipulated to assure that it is representative of the airplane for the maneuvers selected. Such an experiment should also test whether the results generalize to other potentially diagnostic maneuvers and other pilot populations. Ideally, some validation of the quasi-transfer design using a real airplane could also be undertaken.

6. REFERENCES

Advisory Group for Aerospace Research and Development. (1980). Fidelity of simulation for pilot training (AGARD-AR-159). France: Neuilly sur Seine.

Ashkenas, I. L. (1985, September). Collected flight and simulation comparisons and considerations. Paper presented at the Flight Mechanics Panel Symposium on Flight Simulation, Cambridge, United Kingdom.

Boldovici, J.A. (1992). Simulator Motion (ARI Technical Report 961). Alexandria, VA: US Army Research Institute.

Boothe, E.M. (email with attached spreadsheet, October 26, 2000).

Brown, Y. J., Cardullo, F. M., & Sinacori, J. B. (1989). Need-based evaluation of simulator force and motion cuing devices (AIAA 89-3272-CP). American Institute of Aeronautics and Astronautics.

Bürki-Cohen, J., Kendra, A. J., Kanki, B. G., & Lee, A. T. (in press). Realistic radiocommunications in pilot simulator training (DOT/FAA report).

Bürki-Cohen, J., Soja, N., & Longridge, T. (1998a). Simulator fidelity requirements: The case of platform motion. Paper presented at the International Training and Education Conference and Exhibition, Lausanne, Switzerland.

Bürki-Cohen, J., Soja, N., & Longridge, T. (1998b). Simulator platform motion - The need revisited. The International Journal of Aviation Psychology, 8(3), 293-317.

Bussolari, S. R., Young, L. R., & Lee, A. T. (1987, September). The use of vestibular models for design and evaluation of flight simulator motion (AGARD-CP-433, 1988). Paper presented at the Aerospace Medical Panel Symposium on Motion Cues in Flight Simulation and Simulator Induced Sickness, Brussels, Belgium.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Cross, Kenneth D. (1991, April). Training effectiveness assessment: Methodological problems and issues (DOT/FAA/RD-92/2). Paper presented at the NASA/FAA Helicopter Simulator Workshop, Santa Clara, CA.

Dichgans, J., & Brandt, T. (1978). Visual-vestibular interactions: Effects on self-motion perception and postural control. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), *Handbook of Sensory Physiology: Vol. 8 Perception* (pp.755-804). Berlin, Germany: Springer-Verlag.

Ebbinghaus, H. (1964). Memory: A contribution to experimental psychology. New York: Dover Publications.

Federal Aviation Administration. (1991, July 29). Airplane Simulator Qualification (Advisory Circular 120-40B). Washington, DC: US Government Printing Office.

Federal Aviation Administration. (1995, July). Airline Transport Pilot and Type Rating Practical Test Standards (FAA-S-8081-5B). Washington, DC: US Government Printing Office.

Federal Aviation Regulation, Part 121. (1980). Advanced Simulation Plan (14 C.F.R., Part 121, Appendix H). Washington, DC: US Government Printing Office.

Federal Aviation Regulation, Part 121. (1990). Advanced Qualification Program (14 C.F.R., Part 121, SFAR 58). Washington, DC: US Government Printing Office.

Guedry, F. E. J. (1987, September). Technical evaluation report (AGARD-CP-433, 1988). Paper presented at the Aerospace Medical Panel Symposium on Motion Cues in Flight Simulation and Simulator Induced Sickness, Brussels, Belgium.

Go, T., DiSario, R., Bürki-Cohen, J., & Jo, Y. (in preparation). Relationship between objective measures of pilot performance/behavior and instructor grades (DOT/FAA report).

Gundry, J. (1976, April). Man and motion cues. Paper presented at the Third Flight Simulation Symposium, London.

Hall, J. R. (1978, April). Motion versus visual cues in piloted flight simulation (AGARD-CP-249). Paper presented at the Flight Mechanics Panel Specialists' Meeting on Piloted Aircraft Environment Simulation Techniques, Brussels, Belgium.

Hall, J. R. (1989). The need for platform motion in modern piloted flight training simulators (Technical Memorandum Tech Memo FM 35). Bedford, United Kingdom: Royal Aerospace Establishment.

Hosman, R. J. A. W. (1996). Pilot's perception and control of aircraft motions. Delft, The Netherlands: Delftse Universitaire Pers.

Hosman, R. (1999). Are Criteria for Motion Cueing and Time Delays Possible? (AIAA 99-4028). American Institute of Aeronautics and Astronautics.

Hosman, R. J. A. W., & van der Vaart, J. C. (1981). Effects of vestibular and visual motion perception on task performance. Acta Psychologica, 48, 271-287.

Jacobs, R. S. (1976). Simulator cockpit motion and the transfer of initial flight training (Technical Report ARL-76-8/AFOSR-76-4). Urbana-Champaign: University of Illinois, Institute of Aviation, Aviation Research Laboratory.

Joint FAA/Industry Symposium. (1996a, March). Symposium on Level B Airplane Simulator Aeromodel Validation Requirements, Washington Dulles Airport Hilton. (Transcript available in electronic format from Thomas Longridge, Advanced Qualification Program Manager, FAA AFS-230, phone: 703-661-0275)

Joint FAA/Industry Symposium. (1996b, June). Symposium on Level B Airplane Simulator Motion Requirements, Washington Dulles Airport Hilton. (Transcript available in electronic format from Thomas Longridge, Advanced Qualification Program Manager, FAA AFS-230, phone: 703-661-0275)

Koonce, J. M. (1974). Effects of ground-based aircraft simulator motion conditions upon prediction of pilot proficiency. Part II (AFOSR-TR-74-1292). Savoy: University of Illinois.

Lahiri, A. (2000). Significance of a qualitative handling assessment approach towards motion system requirements for flight simulators. Proceedings of the Royal Aeronautical Society conference on Flight Simulation – A Decade of Regulatory Change: Has the Situation Changed for the Better? London, UK.

Lee, A. T., & Bussolari, S. R. (1989). Flight simulator platform motion and air transport pilot training. Aviation Space and Environmental Medicine, 60 (2), 136-140.

Levison, W. H. (1981). Effects of whole-body motion simulation on flight skill development (Final Report AFOSR-TR-82-006). Washington, DC: Bolling Air Force Base, Office of Scientific Research.

Levison, W. H., & Junker, A. M. (1977). A model for the pilot's use of motion cues in roll-axis tracking tasks (Interim Scientific Report AMRL-TR-77-40). Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory, Aerospace Medical Division, Air Force Systems Command.

Longridge, T., Ray, P., Boothe, E., & Bürki-Cohen, J. (1996, May). Initiative towards more affordable flight simulators for U.S. commuter airline training. Proceedings of the Royal

Aeronautical Society Conference on Training—Lowering the Cost, Maintaining the Fidelity, London, 2.1-2.17.

Martin, E. A. (1985, April). An investigation regarding the use of a dynamic seat-pan display for training and as a device for communicating roll-axis motion information. Paper presented at the Third Symposium on Aviation Psychology, Columbus: Ohio State University, Aviation Psychology Laboratory.

Martin, E. L. (1981). Training effectiveness of platform motion: Review of motion research involving the advanced simulator for pilot training and the simulator for air-to-air combat (Final Report AFHRL-TR-79-51) Brooks Air Force Base, TX: Air Forces Human Resources Laboratory, .

McCauley, M. E. (Ed.). (1984). Research issues in simulator sickness: Proceedings of a workshop. Washington, DC: National Research Council.

McCauley, M. E., Hettinger, L. J., Sharkey, T. J., & Sinacori, J. B. (1990). The effects of simulator visual-motion asynchrony on simulator induced sickness. American Institute of Aeronautics and Astronautics.

McGuiness, J., Bouwman, J.H., & Forbes, J.M. (1981). Simulator sickness occurrences in the 2E6 Air Combat Maneuvering Simulator (NAVTRAEQUIPCEN 80-C-0135-4500-1). Orlando, FL: Naval Training Equipment Center.

Oman, C. (1991). Sensory conflict in motion sickness: An observer theory approach. In S. R. Ellis, M. K. Kaiser, & A. C. Grunwald (Eds.), Pictorial communication in virtual and real environments (pp. 362-376). New York: Taylor and Francis.

Peters, R. A. (1969). Dynamics of the Vestibular System and Their Relation to Motion Perception, Spatial Disorientation, and Illusions (NASA CR-1309). National Aeronautics and Space Administration, Washington, D.C.

Reid, L. D., & Nahon, M. A. (1988). Response of airline pilots to variations in flight simulator motion algorithms. Journal of Aircraft, 25 (7), 639-646.

Roscoe, S.N. (1980). Aviation Psychology. Ames, IA: The Iowa State University Press.

Ryan, L. E., Scott, P. G., & Browning, R. F. (1978). The effects of simulator landing practice and the contribution of motion simulation to P-3 training (TAEG Report No. 63). Orlando, FL: Navy Training Analysis and Evaluation Group.

Salas, E., Bowers, C.A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. The International Journal of Aviation Psychology, 8(3), 197-208.

Sedgwick, H. A. (1986). Section IV: Space and motion perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of perception and human performance: Sensory processes and perception (Vol. I). New York: John Wiley and Sons.

Siegel, S. (1956). Nonparametric Statistics for the Behavioral Sciences. New York, NY: McGraw-Hill Book Company Inc.

White, A.D. & Rodchenko, V.V. (1999). Motion Fidelity Criteria Based on Human Perception and Performance (AIAA 99-4330). American Institute of Aeronautics and Astronautics.

Young, L. R. (1967). Some effects of motion cues on manual tracking. Journal of Spacecraft and Rockets, 4 (10), 1300-1303.

Zaichik, L. E., Rodchenko, V. V., Rufov, I. V., Yashin, Y. P., and White, A. D. (1999). Acceleration Perception (AIAA 99-4334-CP). American Institute of Aeronautics and Astronautics.

APPENDIX A – AIRPORT, WEATHER, AND AIRCRAFT INFORMATION

Pilots are in position for take-off	
University Park Airport	**State College, PA**
Runway	24
Runway Length	5000 feet
Elevation	1205 feet MSL
Time of Day	dusk
Weather	
Ceiling	indefinite
Visibility	one quarter mile
Temperature	55^{o} F
Wind	150^{o} at 10 knots
Altimeter Setting	30.08 in Hg
Airplane	
Weight	20,500 pounds
Fuel	2,500 pounds
Center of Gravity	22% MAC
Flap Setting	9
Trim Setting	22
V_1/V_R	105 KIAS
V_2	111 KIAS
V_{yse}	126 KIAS
Simulator Sound Level	65% (~100% airplane sound level)

APPENDIX B – INSTRUCTOR BRIEFING

What is this all about?

This is a test of the simulator. Simulator training has become indispensable for safe pilot training and airplane operation. You will be helping us to determine what kind of simulator best serves that purpose.

The procedures have been jointly determined with ALPA (Air Line Pilots Association), participating air carrier, training center, FAA, and Human Factors Scientists. We have made every effort to design an experiment that will lead to informative results with the end goal of improving training for all pilots.

What are we asking you to do?

<u>Maneuvers</u>
During the first 40 minutes before the pilots' maneuver validation they will practice different take-off maneuvers. Only the captain will be flying (the pilots should not switch seats). You will be evaluating their performance just as you normally do. We ask you not to provide any coaching or feedback to the crew.

The experiment will be controlled by a laptop computer. It will (1) determine the maneuvers and the environment; (2) set up the simulator and ask you to verify; (3) ask you for a grade; and (4) give you information. Some of the information is for the pilots, other information is for you only. Please follow the directions carefully and make sure that the crew does not take off before you have cleared them (as directed by the laptop).

Please call the technician only when a problem makes it impossible to continue. Do not try to manipulate the laptop. All other comments should be reserved for the questionnaires. Do not discuss them with the pilots or call the technician.

<u>Questionnaires</u>
After the pilots have flown a few maneuvers, we will ask all of you questions about the simulator. At the end you will have one more chance to tell us what you thought.
If the laptop or simulator fails and the experiment has to be terminated, please hand out the questionnaires to the crew and yourself and fill them out.

And what if the crew doesn't do well?
Participation is anonymous. This is an ALPA approved non-jeopardy event. Their participation will have no effect on their record or careers. Although pilots may refuse participation, any such withdrawal is likely to render the results less representative.

Need for Discretion
The value of the experiment depends critically on preventing pilot expectation. Thus, we critically depend on your discretion. We count on you not disclosing information that is intended for you only and also not providing any coaching or feedback. We also ask you not to discuss the experiment after this session. We appreciate your cooperation, and we will provide a full briefing after the conclusion of the experiment.

Airport, Weather, and Aircraft Information

Pilots are in position for take-off	
University Park Airport	**State College, PA**
Runway	24
Runway Length	5000 feet
Elevation	1205 feet MSL
Time of Day	Dusk
Weather	
Ceiling	Indefinite
Visibility	one quarter mile
Temperature	55^{o} F
Wind	150^{o} at 10 knots
Altimeter Setting	30.08
Airplane	
Weight	20,500 pounds
Fuel	2,500 pounds
Center of Gravity	22% Mac
Flap Setting	9
Trim Setting	22
V_1/V_R	105 KIAS
V_2	111 KIAS
V_{yse}	126 KIAS
Sound Level	65%

APPENDIX C – PILOT BRIEFING

NOTE: If you participated in a similar experiment during your 6-month LOFT, inform the simulator technician and proceed directly to the maneuver validation.

PILOT BRIEFING

What is this all about?

This is a test of the simulator. Simulator training has become indispensable for safe pilot training and safe airplane operation. You will be helping us to determine what kind of simulator best serves that purpose.

The procedures have been jointly determined by a team consisting of representatives from the ALPA (Air Line Pilots Association), participating air carrier, training center, FAA, and Human Factors Scientists. We have made every effort to design an experiment that will lead to informative results with the end goal of improving training for all pilots.

What are we asking you to do?

<u>Maneuvers</u>

For the first 30-40 minutes before your maneuver validation you will fly some maneuvers, all of which you have practiced before. Only the captain will be flying (you should not switch seats).The instructor may have you repeat some of the maneuvers, but will not provide any coaching or feedback. Remember, we are testing the simulator, not you. Of course, we still want you to do your best!

<u>Questionnaire</u>

After you have flown a few maneuvers, we will ask you questions about the simulator.
At the end you will have one more chance to tell us what you thought.
If you notice anything unusual, please reserve comments for the questionnaires. Do not discuss them with each other or call the technician.

And what if you don't do well?

Pilot performance will tell us something about the simulator, not about pilots. Moreover, your name will never be associated with your performance. This is a non- jeopardy event that has been cleared with ALPA and your airline. Your participation will have no effect on your record or your career.

Need for Discretion

The conclusions for this experiment will be based on the responses gathered from many pilots. It is **critical** that all are fresh to the experiment without expectations or preconceived opinions. Thus, please don't discuss your experience with your colleagues. We will provide a full briefing after the conclusion of the experiment.

Right to Withdraw

Your data is a reflection of the simulator. To draw valid conclusions on the simulator we need data from a representative sample of pilots—that includes you! However, you do have the right to refuse participation by filling out a withdrawal form.

Withdrawal Form

Before making your final decision, make sure you understand the following:

1. In this experiment, pilot performance is used to evaluate the effectiveness of the simulator, not the proficiency of the pilot.
2. Your participation is anonymous and will not impact your record or career.
3. To draw valid conclusions on simulator effectiveness, it is crucial to have data from a representative set of pilots—including you.
4. If you are withdrawing because you think you have done poorly, it is even more important for you to continue because it may be exactly your data that will point to a potential flaw in the simulator. That is, if all pilots who think they've done poorly withdraw, we would falsely conclude that the simulator characteristics tested are beneficial to pilot training. Therefore, withdrawal of any pilot may distort the results of this experiment and ultimately lead to the development of ineffective training equipment.

If You Still Want To Withdraw:

Please fill out the form below and your data will be destroyed.

Date	
Time	
Instructor's Name	
Reason For Withdrawal	

APPENDIX E – QUESTIONNAIRES

Questionnaire for
Pilot Flying

Date: _____________

Time: _____________

Number of flight hours: _____________

Instructor: _____________

Experiment Crew Number (ask instructor for this number): _____________

---

1. In this questionnaire you are asked to evaluate the simulator. You are asked to make one of two comparisons, as indicated on each page:
    a) the **simulator today** to the **simulator the last time** you flew it
    b) the **simulator** (as flown today) to the **airplane**
2. Please base all of your judgments on the **maneuvers that you have flown so far today**. For comparisons with the airplane, you may have to base your judgments on how you would **expect** the airplane to behave during these maneuvers.
3. Please indicate each judgment by placing an **X** in the appropriate box.

Control Precision

Compare simulator to **airplane**

My control precision in the simulator today was ...

Controlled Variables	1	2	3	4	5
	much worse than expected in airplane	somewhat worse	the same as expected in airplane	somewhat better	much better than expected in airplane
altitude control					
heading control					
airspeed control					
roll control					
pitch control					
overall control precision					

Control Strategy and Technique

Compare simulator to **airplane**

Operating the controls of the simulator today was ...

Controls	1	2	3	4	5
	much harder than expected in airplane	somewhat harder	the same as expected in airplane	somewhat easier	much easier than expected in airplane
rudder inputs					
aileron inputs					
elevator inputs					
power inputs					
rudder trim inputs					
aileron trim inputs					
elevator trim inputs					
overall control strategy and technique					

Read This First:

- If you flew **exactly 2** take-off runs, please move on to the next page (page 5).
- If you flew **3 or more** take-off runs, please continue with this page.

Gaining Proficiency

Compare simulator **today** to simulator **last time**

Gaining proficiency in the simulator today was ...

	1	2	3	4	5
	much harder than last time	somewhat harder	the same as last time	somewhat easier	much easier than last time
Overall					

Physical and Mental Workload

Compare simulator to **airplane**

Workload in the simulator today was …

Type of Workload	1	2	3	4	5
	much higher than expected in airplane	somewhat higher	the same as expected in airplane	somewhat lower	much lower than expected in airplane
Physical					
Mental					

Comfort

(absence of nausea and simulator-induced disorientation)

Compare simulator **today** to simulator **last time**

My comfort in the simulator today was ...

	1	2	3	4	5
	much worse than last time	somewhat worse	the same as last time	somewhat better	much better than last time
overall comfort					

Acceptability

(absence of deficiencies that would make it harder to fly)

Compare simulator **today** to simulator **last time**

Acceptability of the simulator today was ...

	1	2	3	4	5
	much worse than last time	somewhat worse	the same as last time	somewhat better	much better than last time
Overall Acceptability					

Stop here!

Return questionnaire
to instructor
before flying some more.

Thanks!

Final Question

We are interested in how your opinion may have changed now that you have flown some additional maneuvers. Please circle your answer to **each** of the questions below (leaving your previous answers as they are).

1. My control precision in the simulator today was ...

a) much worse than expected in **airplane**
b) somewhat worse
c) the same as expected in **airplane**
d) somewhat better
e) much better than expected in **airplane**

2. Operating the controls of the simulator today was ...

a) much harder than expected in **airplane**
b) somewhat harder
c) the same as expected in **airplane**
d) somewhat easier
e) much easier than expected in **airplane**

3. Gaining proficiency in the simulator today was ...

a) much harder than last time in **simulator**
b) somewhat harder
c) the same as last time in **simulator**
d) somewhat easier
e) much easier than last time in **simulator**

4. Workload in the simulator today was …

a) much higher than expected in **airplane**
b) somewhat higher
c) the same as expected in **airplane**
d) somewhat lower
e) much lower than expected in **airplane**

5. My comfort in the simulator today was ...

a) much worse than last time in **simulator**
b) somewhat worse
c) the same as last time in **simulator**
d) somewhat better
e) much better than last time in **simulator**

6. Acceptability of the simulator today was ...

a) much worse than last time in **simulator**
b) somewhat worse
c) the same as last time in **simulator**
d) somewhat better
e) much better than last time in **simulator**

7. Do you have any additional comments?

____

____

____

____

____

____

____

Thank you for your participation! Please return this questionnaire to the instructor.

Questionnaire for
Pilot Not Flying

Date: _____________

Time: _____________

Number of flight hours: _____________

Instructor: _____________

Experiment Crew Number (ask instructor for this number): _____________

---

1. In this questionnaire you are asked to evaluate the flight simulator.
2. Each page will indicate whether to base your judgment on
   a) your **observation of the pilot flying** or
   B. **your own experience**.

III. Each page also indicates which of two comparisons you should make:

   A. the **simulator today** to the **simulator the last time** you were in it as PNF

   B. the **simulator** (as experienced today) to the **airplane**, as you experience it as a PNF

IV. Please base all of your judgments on the **maneuvers that you have experienced so far today**. For comparisons with the airplane, you may have to base your judgments on how you would **expect** the airplane to behave during these maneuvers.

V. Please indicate each judgment by placing an **X** in the appropriate box.

Control Precision of Pilot Flying

Compare simulator to **airplane**

The PF's control precision in the simulator today was ...

Controlled Variables	1	2	3	4	5
	much worse than expected in airplane	somewhat worse	the same as expected in airplane	somewhat better	much better than expected in airplane
altitude control					
heading control					
airspeed control					
roll control					
pitch control					
overall control precision					

Control Strategy and Technique of Pilot Flying

Compare simulator to **airplane**

For the PF, operating the controls of the simulator today appeared ...

Controls	1	2	3	4	5
	much harder than expected in airplane	somewhat harder	the same as expected in airplane	somewhat easier	much easier than expected in airplane
rudder inputs					
aileron inputs					
elevator inputs					
power inputs					
rudder trim inputs					
aileron trim inputs					
elevator trim inputs					
overall control strategy and technique					

Read This First:

- If the pilot flying flew **exactly 2** take-off runs, please move on to the next page (page 5).
- If the pilot flying flew **3 or more** take-off runs, please continue with this page.

Gaining Proficiency for Pilot Flying

Compare simulator **today** to simulator **last time**

For the PF, gaining proficiency in the simulator today was …

	1	2	3	4	5
	much harder than last time	somewhat harder	the same as last time	somewhat easier	much easier than last time
Overall					

Physical and Mental Workload of Pilot Flying

Compare simulator to **airplane**

The PF's Workload in the simulator today was …

Type of Workload	1	2	3	4	5
	much higher than expected in airplane	somewhat higher	the same as expected in airplane	somewhat lower	much lower than expected in airplane
Physical					
Mental					

Comfort of Pilot Not Flying
(absence of nausea and simulator-induced disorientation)

Compare simulator **today** to simulator **last time**

My comfort in the simulator today was ...

	1	2	3	4	5
	much worse than last time	somewhat worse	the same as last time	somewhat better	much better than last time
overall comfort					

Acceptability to Pilot Not Flying

(absence of deficiencies that would make it harder to perform crew duties)

Compare simulator **today** to simulator **last time**

Acceptability of the simulator today was ...

	1	2	3	4	5
	much worse than last time	somewhat worse	the same as last time	somewhat better	much better than last time
Overall Acceptability					

Stop here!

Return questionnaire
to instructor
before continuing
with more maneuvers.

Thanks!

Final Question

We are interested in how your opinion may have changed now that you have experienced some additional maneuvers. Please circle your answer to **each** of the questions below (leaving your previous answers as they are).

I. The PF's control precision in the simulator today was…

A. much worse than expected in **airplane**
B. somewhat worse
C. the same as expected in **airplane**
D. somewhat better
E. much better than expected in **airplane**

II. For the PF, operating the controls of the simulator today appeared ...

A. much harder than expected in **airplane**
B. somewhat harder
C. the same as expected in **airplane**
D. somewhat easier
E. much easier than expected in **airplane**

III. For the PF, gaining proficiency in the simulator today was …

A. much harder than last time in **simulator**
B. somewhat harder
C. the same as last time in **simulator**
D. somewhat easier
E. much easier than last time in **simulator**

IV. The PF's Workload in the simulator today was …

A. much higher than expected in **airplane**
B. somewhat higher
C. the same as expected in **airplane**
D. somewhat lower
E. much lower than expected in **airplane**

V. My comfort in the simulator today was ...

A. much worse than last time in **simulator**
B. somewhat worse
C. the same as last time in **simulator**
D. somewhat better
E. much better than last time in **simulator**

VI. Acceptability of the simulator today was ...

A. much worse than last time in **simulator**
B. somewhat worse
C. the same as last time in **simulator**
D. somewhat better
E. much better than last time in **simulator**

VII. Do you have any additional comments?

____
____
____
____
____
____
____
____

Thank you for your participation! Please return this questionnaire to the instructor.

Questionnaire for
Instructor

Name: (First)______________________ (Last) ______________________

Date: _____________

Time: _____________

Years as Instructor: _____________

Experiment Crew Number (as indicated on laptop): _____________

---

1. In this questionnaire you are asked to evaluate the performance of the **pilot flying**. You will either:

   a) use company standard ratings

   b) compare the performance of the pilot flying to the performance of an average pilot flying coming in for his/her 12 months maneuver validation in the simulator, hereafter referred to as the "Average Pilot Flying". Please do not base "average" on any experiences you have had previously participating in this experiment, but **only** on your experiences with normal 12 months maneuver validations.

II. Please base all of your judgments on the **maneuvers that have been flown so far today**.

III. Please indicate each judgment by placing an **X** in the appropriate box.

Performance of Pilot Flying

Rate overall performance of PF by choosing one of the following grades.

The overall performance of the Pilot Flying was ...

Controlled Variables	1	2	3	4
	unsatisfactory	practical test standards	company standards	excellent
altitude control				
heading control				
airspeed control				
roll control				
pitch control				
overall performance				

Control Strategy and Technique of Pilot Flying

Compare overall control behavior of PF to **Average Pilot Flying**

The overall control strategy and technique used by the Pilot Flying was ...

Controls	1	2	3	4	5
	much worse than the Averge Pilot Flying	somewhat worse than the Average Pilot Flying	the same as the Average Pilot Flying	somewhat better than the average Pilot Flying	much better than the Average Pilot Flying
rudder inputs					
aileron inputs					
elevator inputs					
power inputs					
rudder trim inputs					
aileron trim inputs					
elevator trim inputs					
overall control strategy and technique					

Read This First:

- If the pilots flew **exactly 2** take-off runs, please move on to the next page (page 5).
- If they flew **3 or more** take-off runs, please continue with this page.

Gaining Proficiency (for Pilot Flying)

Compare PF to **Average Pilot Flying**

Gaining proficiency for the Pilot Flying was …

	1	2	3	4	5
	much harder than for the Average Pilot Flying	somewhat harder	the same as for the Average Pilot Flying	somewhat easier	much easier than for the Average Pilot Flying
Overall					

Physical and Mental Workload of Pilot Flying

Compare PF to **Average Pilot Flying**

The workload of the Pilot Flying appeared to be …

Type of Workload	1	2	3	4	5
	much higher than of the Average Pilot Flying	somewhat higher	the same as of the Average Pilot Flying	somewhat lower	much lower than of the Average Pilot Flying
Physical					
Mental					

Comfort

(absence of nausea and simulator-induced disorientation)

Here we would like you to **compare** your comfort **today** to the **usual** comfort you experience during equivalent maneuvers in a normal 12 months maneuver validation, but **not** on your experiences participating in this experiment.

My comfort in the simulator today was ...

	1	2	3	4	5
	much worse than usual	somewhat worse	the same as usual	somewhat better	much better than usual
overall comfort					

Stop here!

Collect questionnaires
from both pilots
when they are done
& before continuing
with more maneuvers.

Thanks!

Final Question

We are interested in how your opinion may have changed now that you have experienced some additional maneuvers. Please circle your answer to **each** of the questions below (leaving your previous answers as they are).

I. The PF's overall performance in the simulator today …

A. was unsatisfactory (1)
B. met practical test standards (2)
C. met company standards (3)
D. was excellent (4)

II. The overall control strategy and technique used by the Pilot Flying was ...

A. much worse than the Average Pilot Flying
B. somewhat worse
C. the same as the Average Pilot Flying
D. somewhat better
E. much better than the Average Pilot Flying

III. Gaining proficiency for the Pilot Flying was …

A. much harder than for the Average Pilot Flying
B. somewhat harder
C. the same as for the Average Pilot Flying
D. somewhat easier
E. much easier than for the Average Pilot Flying

IV. The workload of the Pilot Flying appeared to be …

A. much higher than of the Average Pilot Flying
B. somewhat higher
C. the same as of the Average Pilot Flying
D. somewhat lower
E. much lower than of the Average Pilot Flying

V. My comfort in the simulator today was ...

A. much worse than usual
B. somewhat worse
C. the same as usual
D. somewhat better
E. much better than usual

VI. Do you have any additional comments?

____

____

____

____

____

____

____

____

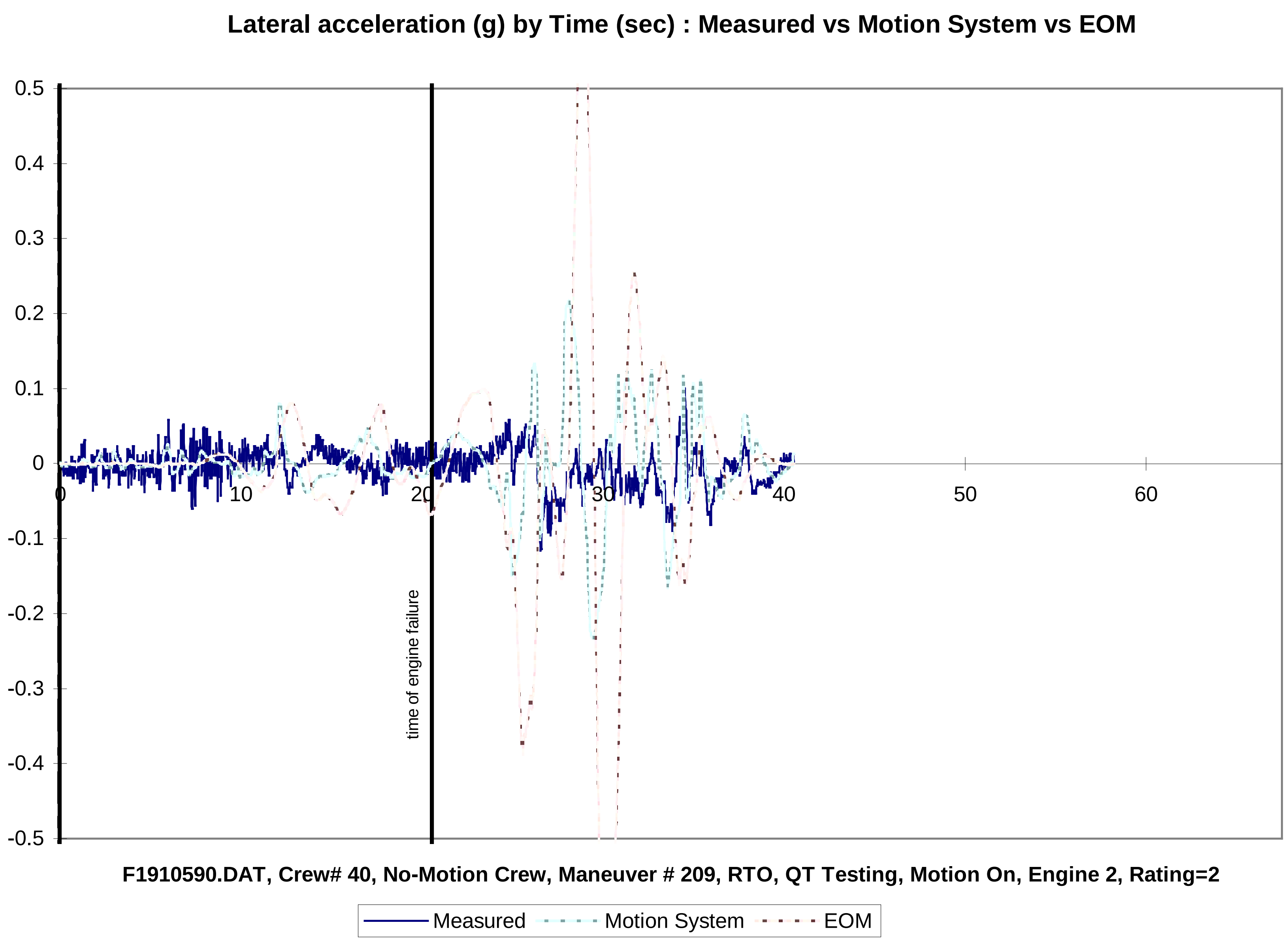
____

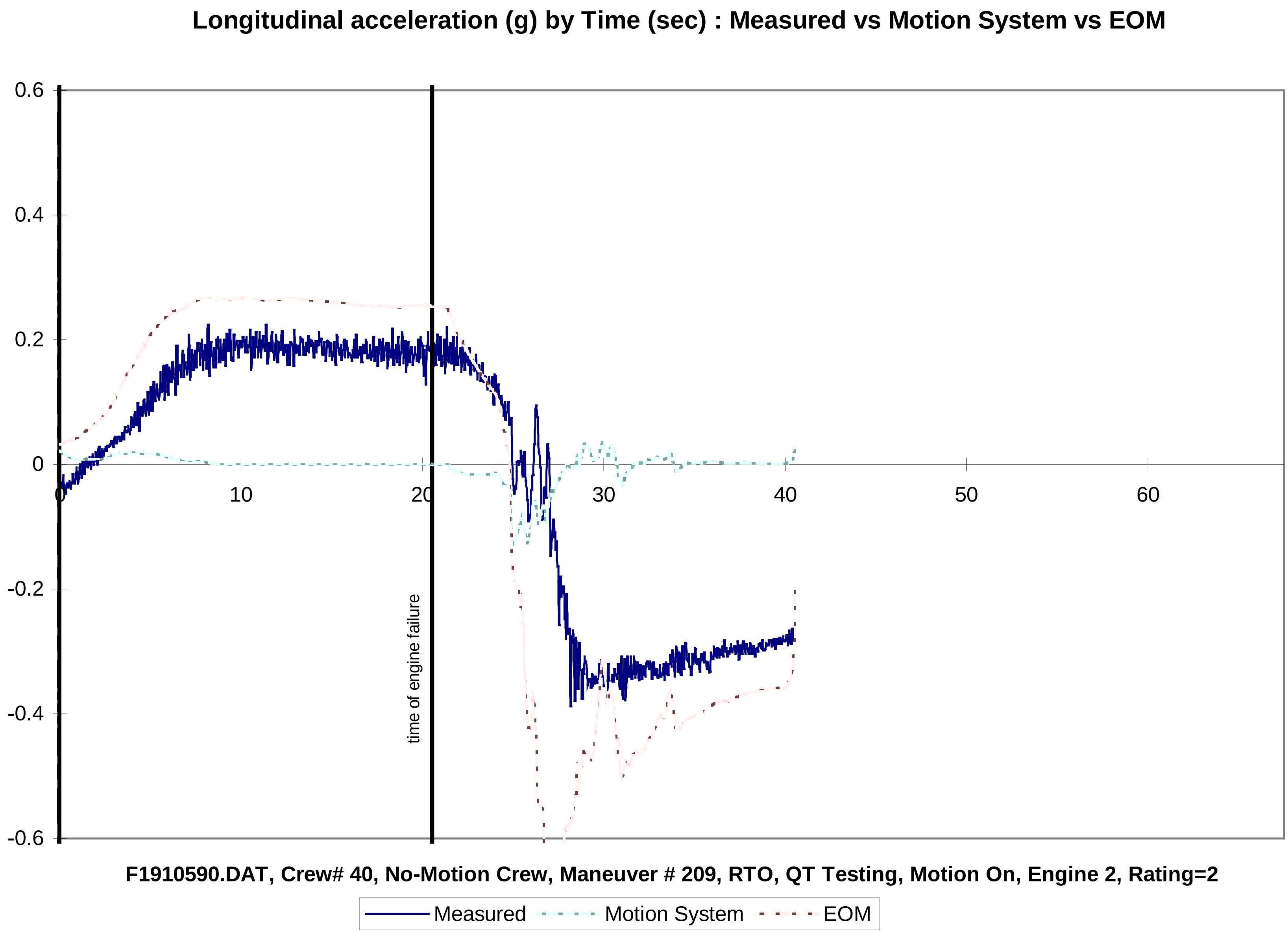
Thank you for your participation! Please collect the questionnaires from both pilots when they are done.

APPENDIX F – LIST OF RECORDED VARIABLES

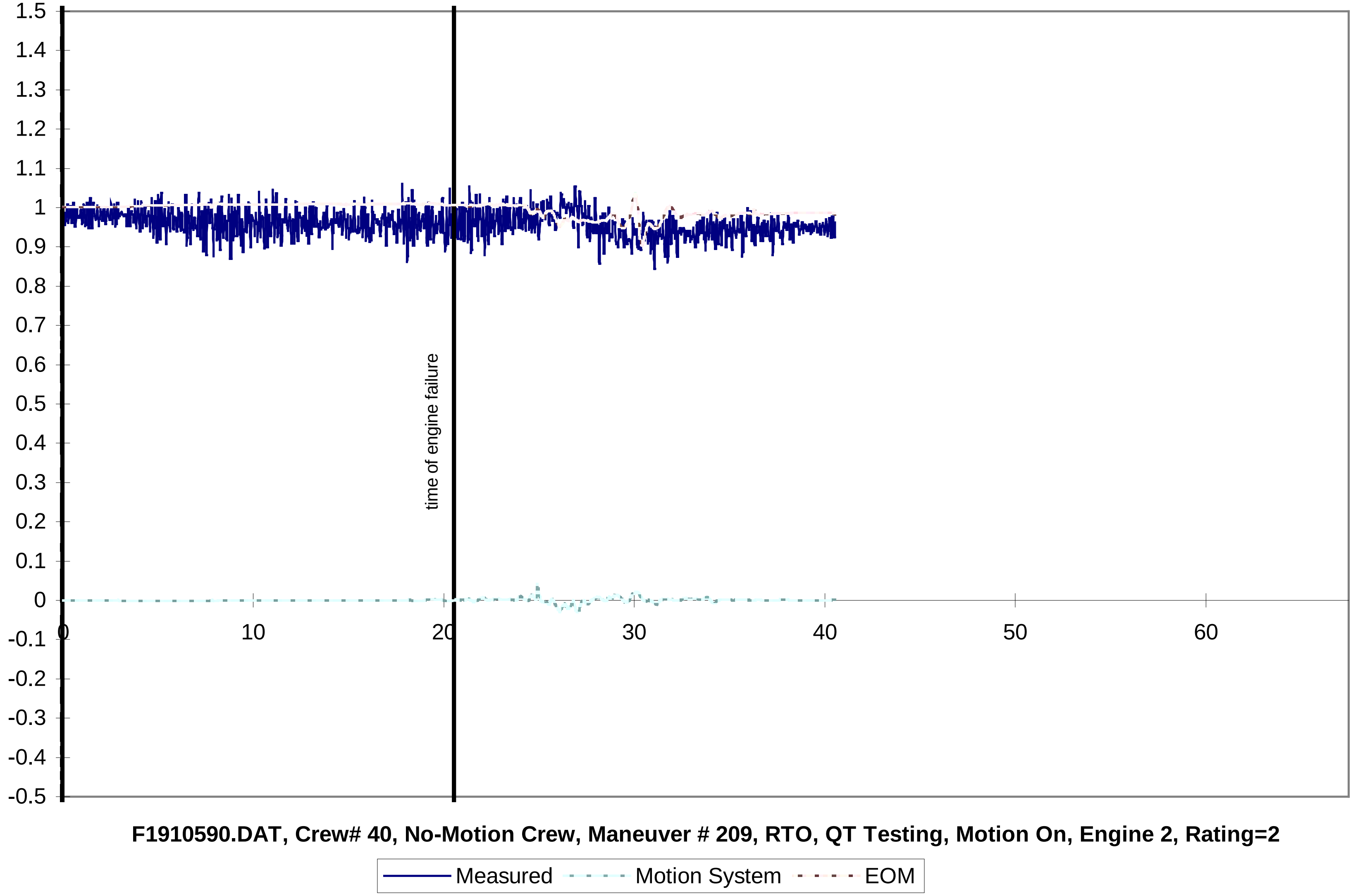
No.	***Variables***	***Unit***
1	Indicated Airspeed	knots
2	Ground Speed	knots
3	Altitude	ft, MSL
4	Radar Altitude	ft, AGL
5	Rate Of Climb	FPM
6	Magnetic Heading	degrees
7	Flap Position	degrees
8	Pitch Attitude	degrees
9	Pitch Rate	degrees/sec
10	Pitch Acceleration	degrees/sec^2
11	Roll Attitude	degrees
12	Roll Rate	degrees/sec
13	Roll Acceleration	degrees/sec^2
14	Sideslip	degrees
15	Yaw Rate	degrees/sec
16	Yaw Acceleration	degrees/sec^2
17	Engine 1 Speed	% RPM
18	Engine 2 Speed	% RPM
19	Engine 1 Torque	%
20	Engine 2 Torque	%
21	Engine 1 PLA	degrees
22	Engine 2 PLA	degrees
23	Engine 1 Condition Lever Angle	degrees
24	Engine 2 Condition Lever Angle	degrees
25	Longitudinal Acceleration, from EOM	g
26	Lateral acceleration, from EOM	g
27	Vertical Acceleration, from EOM	g
28	Longitudinal Acceleration, Measured	g
29	Lateral Acceleration, Measured	g
30	Vertical Acceleration, Measured	g
31	Lateral Deviation (from initial position of aircraft)	ft
32	Ground Distance	ft
33	Rudder Trim Position	units
34	Aileron Trim Position	units
35	Elevator Trim Position	units
36	Roll Rate, measured	degrees/sec
37	Column Position	inches
38	Wheel Position	degrees
39	Pedal Position	inches
40	Column Force	lbs
41	Wheel Force	lbs

42	Pedal Force	lbs
43	Applied Brake Pressure, Pilot Left	psi
44	Applied Brake Pressure, Pilot Right	psi
45	Applied Brake Pressure, Copilot Left	psi
46	Applied Brake Pressure, Copilot Right	psi
47	Brake Pedal Force, Pilot Left	lbs
48	Brake Pedal Force, Pilot Right	lbs
49	Brake Pedal Force, Copilot Left	lbs
50	Brake Pedal Force, Copilot Right	lbs
51	Brake Pedal Position, Pilot Left	inches
52	Brake Pedal Position, Pilot Right	inches
53	Brake Pedal Position, Copilot Left	inches
54	Brake Pedal Position, Copilot Right	inches
55	Left Engine Failed Flag	T=failed
56	Right Engine Failed Flag	T=failed
57	Weight On Wheel, Nose Flag	T=weight on
58	Weight On Wheel, Left Main Flag	T=weight on
59	Weight On Wheel, Right Main Flag	T=weight on
60	Engine 1 Low Pressure Shut Off Valve Flag	T=shut
61	Engine 2 Low Pressure Shut Off Valve Flag	T=shut
62	Longitudinal Acceleration, From Motion System	g
63	Lateral Acceleration, From Motion System	g
64	Vertical Acceleration, From Motion System	g
65	Pitch Acceleration, From Motion System	degrees/sec^2
66	Roll Acceleration, From Motion System	degrees/sec^2
67	Yaw Acceleration, From Motion System	degrees/sec^2
68	Leg 1 Position	volts
69	Leg 2 Position	volts
70	Leg 3 Position	volts
71	Leg 4 Position	volts
72	Leg 5 Position	volts
73	Leg 6 Position	volts
74	Aural Cue Volume	%
75	Elevator Position	degrees
76	Aileron Position	degrees
77	Rudder Position	degrees
78	Landing Gear Selector Handle Position Flag	T=up

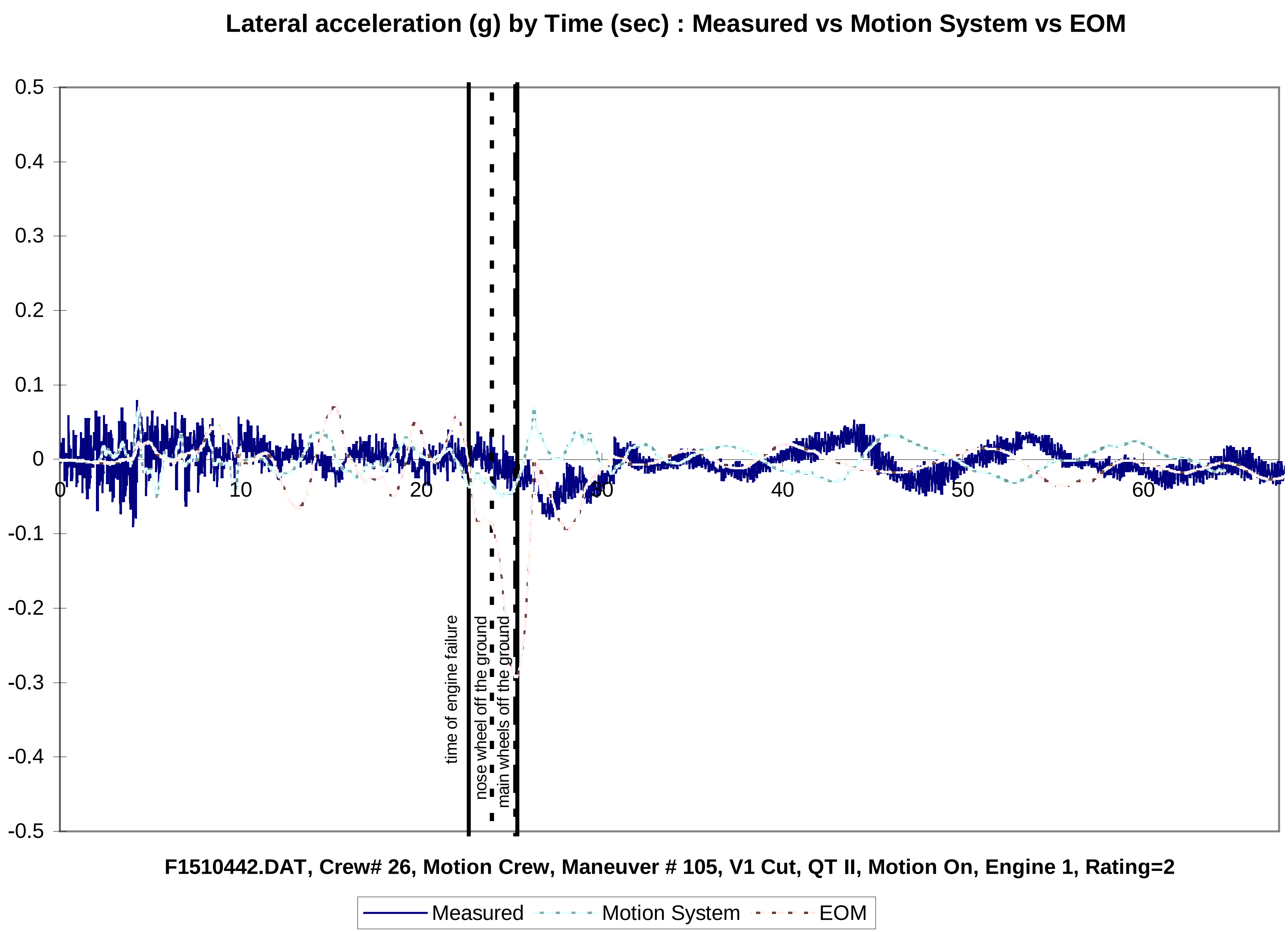

Lateral acceleration (g) by Time (sec) : Measured vs Motion System vs EOM
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
0
10
20
30
40
50
60
time of engine failure
F1910590.DAT, Crew# 40, No-Motion Crew, Maneuver # 209, RTO, QT Testing, Motion On, Engine 2, Rating=2
Measured
Motion System
EOM


Longitudinal acceleration (g) by Time (sec) : Measured vs Motion System vs EOM
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
0
10
20
30
40
50
60
time of engine failure
F1910590.DAT, Crew# 40, No-Motion Crew, Maneuver # 209, RTO, QT Testing, Motion On, Engine 2, Rating=2
Measured
Motion System
EOM

Vertical acceleration (g) by Time (sec) : Measured vs Motion System vs EOM

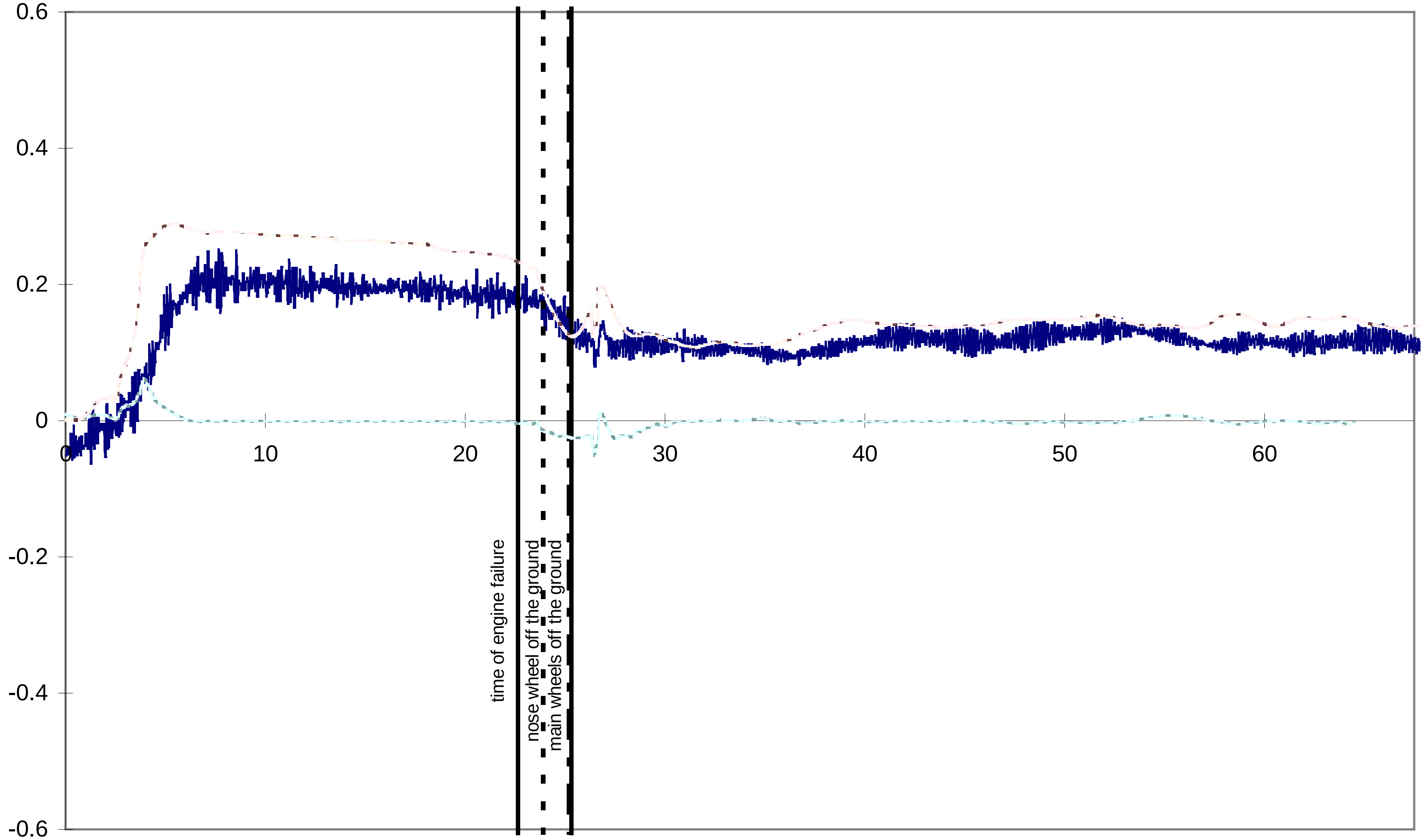


F1910590.DAT, Crew# 40, No-Motion Crew, Maneuver # 209, RTO, QT Testing, Motion On, Engine 2, Rating=2

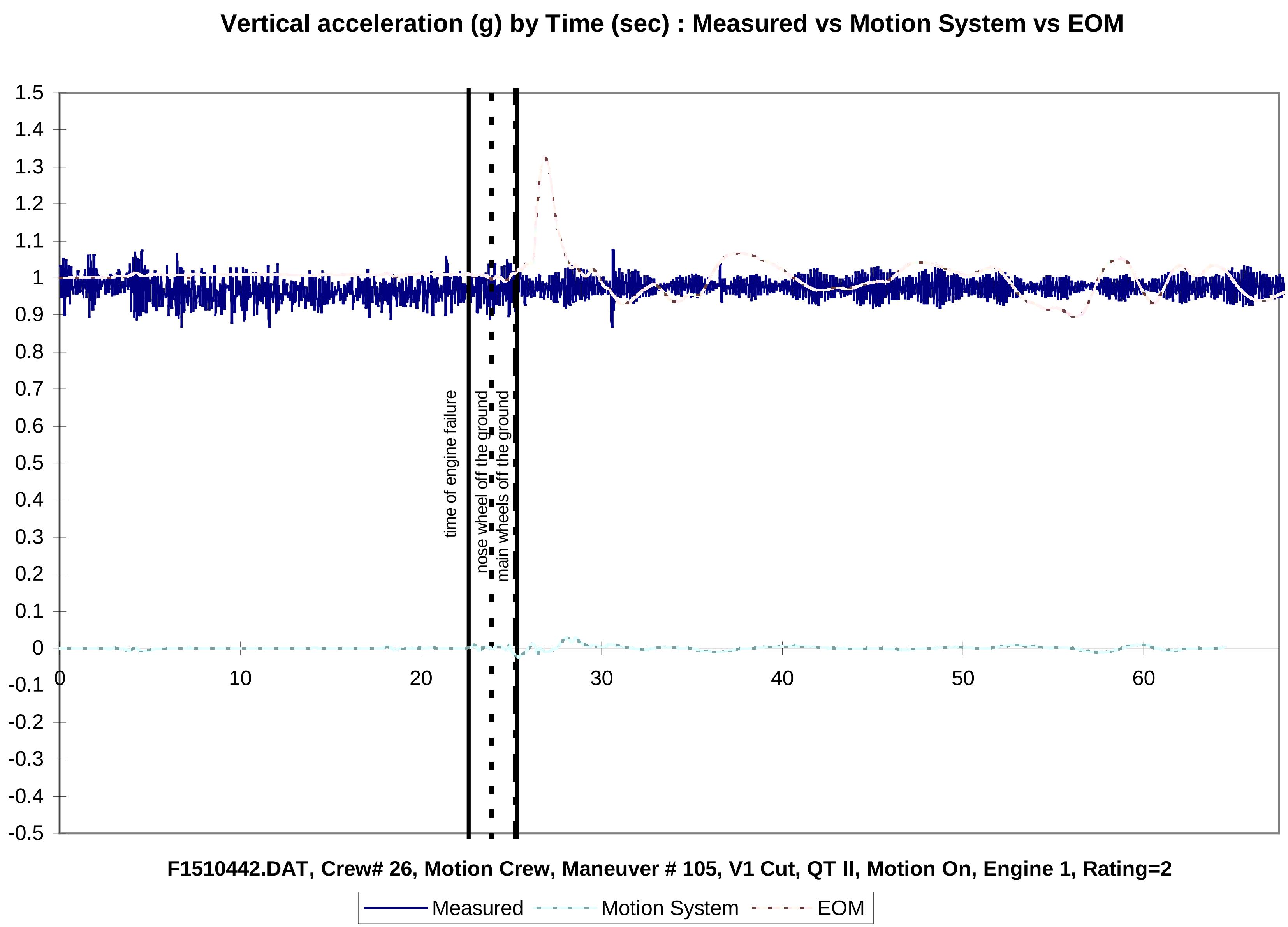


Lateral acceleration (g) by Time (sec) : Measured vs Motion System vs EOM
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
0
10
20
30
40
50
60
time of engine failure
nose wheel off the ground
main wheels off the ground
F1510442.DAT, Crew# 26, Motion Crew, Maneuver # 105, V1 Cut, QT II, Motion On, Engine 1, Rating=2
Measured
Motion System
EOM

Longitudinal acceleration (g) by Time (sec) : Measured vs Motion System vs EOM



F1510442.DAT, Crew# 26, Motion Crew, Maneuver # 105, V1 Cut, QT II, Motion On, Engine 1, Rating=2



Vertical acceleration (g) by Time (sec) : Measured vs Motion System vs EOM
1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
0
10
20
30
40
50
60
time of engine failure
nose wheel off the ground
main wheels off the ground
F1510442.DAT, Crew# 26, Motion Crew, Maneuver # 105, V1 Cut, QT II, Motion On, Engine 1, Rating=2
Measured
Motion System
EOM

APPENDIX H - PRACTICAL TEST STANDARDS/COMPANY STANDARDS (PERFORMANCE)

MANEUVER TYPE	VARIABLE	PTS (FAA)	COMPANY STANDARDS
Normal Take-Off	heading	+/- 5 degrees	+/- 5 degrees
	Airspeed	+/- 5 knots	+/- 5 knots (V2/Vyse)
	initial attitude		10 degrees
Rejected Take-Off (95 KIAS)	None	no numerical standards; • "aligns the airplane on the runway centerline" • "reduces power smoothly and promptly"	no numerical standards; "directional control should be maintained at all times and the a/c should be brought to a stop on the runway centerline"
V_1 Cut, Take-Off continued (110 KIAS)	Heading	After the climb has been established, desired heading +/- 5 degrees	+/- 5 degrees
	Airspeed	After the climb has been established, desired airspeed +/- 5 knots	+5/0 knots (V2/Vyse). If can not pitch up enough to maintain V2 before engine secured, pitch to10 degrees up until engine secured, then maintain V2 until 400 feet AGL, then accelerate to Vyse, retract wingflaps, and thereafter observe Vyse +5/0 knots, with flaps retracted
	bank angle toward good engine	up to 5 degrees	5 degrees toward good engine
	alignment	maintain airplane aligned with the runway heading	same as PTS

APPENDIX I - LIST OF RTO MEASURES

PERFORMANCE			
Short Name	Name	Type	Description
MABS31	Mean Absolute Lateral Deviation	Directional	Mean of absolute lateral deviation of the airplane during 15 sec after engine failure.
RMS31	RMS Lateral Deviation	Directional	Root mean square of the lateral deviation of the center of gravity of the airplane from the runway centerline during 15 sec after engine failure.
S_ABS15	Integrated Yaw Activity	Directional	Integral of absolute yaw rate during 15 sec after engine failure: $\int_0^{15} \lvert r \rvert \, dt$, where r is yaw rate.
RMS6C	RMS Heading Deviation	Directional	Root mean square of the absolute heading deviation from 242^o (runway heading) for 15 sec after engine failure.
MX6C	Maximum Heading Deviation	Directional	Maximum absolute heading deviation from 242^o (runway heading) during 15 sec after engine failure.
RT	Power Lever Reaction Time		The duration from the time when the failed engine torque reduces to 80% of its full power to the time when the torque of the good engine has been reduced to 80%.
RTP	Pedal Reaction Time	Directional	Time for the pedal position to exit 0.5-inch band about its position at engine failure, taking into account the engine failure side.
WORKLOAD			
Short Name	Name	Type	Description
SD39	STD Pedal Position	Directional	Standard deviation of pedal positions during 15 sec after engine failure.
MXC39	Number of Pedal Reversals	Directional	The number of times the pedal exits a 1-inch band centered at its neutral position during 15 sec after engine failure.
TMED39	Sum of Pedal Spectrum f>0.5 Hz	Directional	Sum of pedal spectrum for frequency > 0.5 Hz during 15 sec after engine failure.

APPENDIX J - LIST OF V_1 CUT MEASURES

PERFORMANCE			
Short Name	Name	Type	Description
MX11	Maximum Absolute Bank Angle	Lateral	Maximum absolute bank angle during 15 sec after engine failure.
TOB11	Integrated Bank Angle Exceedance	Lateral	Integral of the absolute bank angle outside of a 5^o band around wing level position during 15 sec after engine failure : $\int_0^{15} g(t)dt$, where $g(t) = \begin{cases} \|\Phi(t)\| - 5 & \text{for } \|\Phi(t)\| > 5 \\ 0 & \text{for } \|\Phi(t)\| < 5 \end{cases}$ with Φ is the bank angle.
TPOS11	Integrated Failure Induced Bank Angle	Lateral	Integral of the absolute bank angle in the direction of the failed engine during 15 sec after engine failure: $\int_0^{15} h(t)dt$, where $h(t) = \begin{cases} -\Phi & \text{for } \Phi < 0 \\ 0 & \text{for } \Phi \geq 0 \end{cases}$ if the left engine fails and $h(t) = \begin{cases} \Phi & \text{for } \Phi > 0 \\ 0 & \text{for } \Phi \leq 0 \end{cases}$ if the right engine fails.
S_ABS12	Integrated Roll Activity	Lateral	Integral of absolute roll rate during 15 sec after engine failure: $\int_0^{15} \|p\|dt$, where p is roll rate.
S_ABS15	Integrated Yaw Activity	Directional	Integral of absolute yaw rate during 15 sec after engine failure: $\int_0^{15} \|r\|dt$, where r is yaw rate.
RMS6C	RMS Heading Deviation	Directional	Root mean square of the absolute heading deviation from 242^o for 15 sec after engine failure.
MX6C	Maximum Heading Deviation	Directional	Maximum absolute heading deviation from 242^o during 15 sec after engine failure.

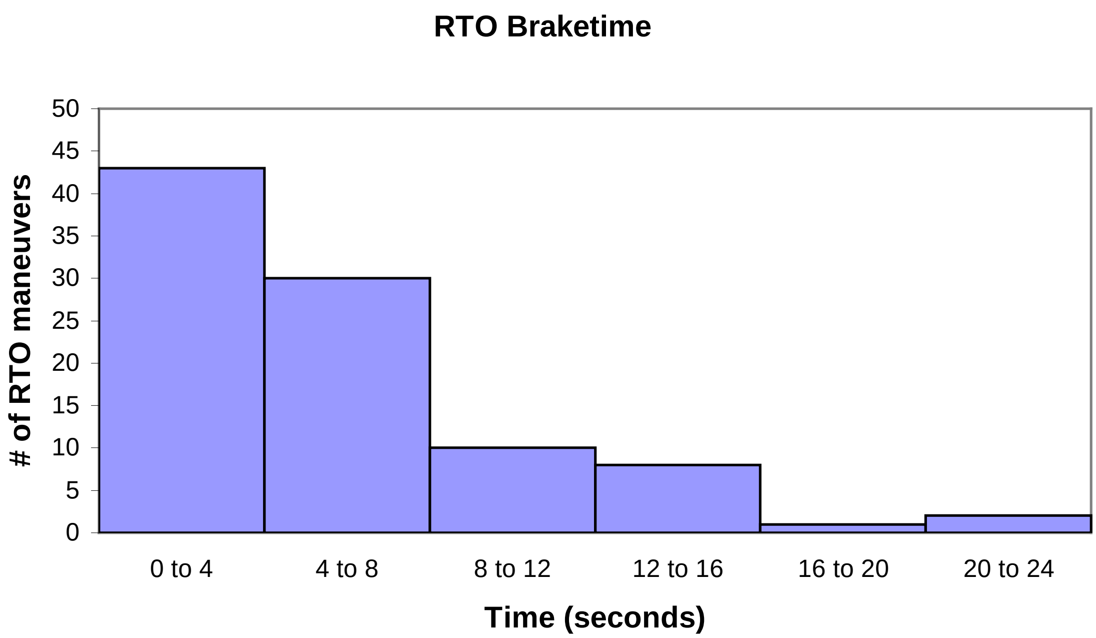
PERFORMANCE (cont'd)			
Short Name	Name	Type	Description
TOB6C	Integrated Heading Exceedance	Directional	Integral of the absolute heading deviation exceeding $\pm 5^{o}$ around the nominal direction of 242^{o} during 15 sec after engine failure: $\int_{0}^{15} f(t)dt$, where $f(t)=\begin{cases} \lvert\Psi(t)\rvert-5 & \text{for } \lvert\Psi(t)\rvert>5 \\ 0 & \text{for } \lvert\Psi(t)\rvert\le 5 \end{cases}$ with Ψ the corrected heading (actual heading minus 242^{o}).
TOB1	Integrated Airspeed Exceedance	Longitudinal	Integral of absolute IAS deviation outside (0, +5 knots) band from V_2 (=111 knots) during 15 sec after engine failure: $\int_{0}^{15} v(t)dt$, where $v(t)=\begin{cases} \lvert V(t)\rvert-5 & \text{for} \quad V(t)>5 \\ \lvert V(t)\rvert & \text{for} \quad V(t)<0 \\ 0 & \text{for} \quad 0\le V(t)\le 5 \end{cases}$ with $V(t)$ the deviation of the IAS from V_2.
TALT	Time to Reach 400 ft. Altitude	Longitudinal	The lapse time from the time of engine failure to the time when altitude reaches 400 feet AGL.
SD8	STD Pitch Angle	Longitudinal	Standard deviation of pitch angle during 15 sec after engine failure.
RT38	Wheel Reaction Time	Lateral	Time for the wheel position to exit 5-degree band about its initial position before engine failure, taking into account the engine failure side.
RT39	Pedal Reaction Time	Directional	Time for the pedal position to exit 0.75-inch band about its initial position, taking into account the effect of sidewind and engine failure side.
WORKLOAD			
Short Name	Name	Type	Description
SD37	STD Column Position	Longitudinal	Standard deviation of column positions during 15 sec after engine failure.
SD38	STD Wheel Position	Lateral	Standard deviation of wheel positions during 15 sec after engine failure.
SD39	STD Pedal Position	Directional	Standard deviation of pedal positions during 15 sec after engine failure.

WORKLOAD (cont'd)			
Short Name	Name	Type	Description
MXC37	Column Reversals	Longitudinal	The number of times the column exits a 4-inch band centered at its neutral position during 15 sec after engine failure.
MXC38	Wheel Reversals	Lateral	The number of times the wheel exits a 10^o band centered at its neutral position during 15 sec after engine failure.
MXC39	Pedal Reversals	Directional	The number of times the pedal exits a 1-inch band centered at its neutral position during 15 sec after engine failure.
TMED37	Sum of Column Spectrum f>0.5 Hz	Longitudinal	Sum of column spectrum for frequency > 0.5 Hz during 15 sec after engine failure.
TMED38	Sum of Wheel Spectrum f>0.5 Hz	Lateral	Sum of wheel spectrum for frequency > 0.5 Hz during 15 sec after engine failure.
TMED39	Sum of Pedal Spectrum f>0.5 Hz	Directional	Sum of pedal spectrum for frequency > 0.5 Hz during 15 sec after engine failure.

APPENDIX K - BRAKE DATA

This summary describes the brake data from the experiment and explains why they cannot be used for measuring reaction time in Rejected Take-Off maneuvers.

It was initially thought that the brake data could be utilized as a way to infer pilots' reaction times in Rejected Take-Off (RTO) maneuvers. This thought was based on the presumption that pressing the brake would be the first thing the pilot does once he detects the engine failure during take-off at the speed below V_1. Examination of the brake data, however, revealed that the presumption is false. The chart below shows the distribution of the brake time for the RTO maneuvers from the experiment, where brake time is defined as the time from the engine failure to the time when brake pedal action is observed. Brake time longer than 20 seconds is observed for some maneuvers. However, even in these extreme cases, the pilots were able to stop the airplane safely.



There are several possible explanations as to why PFs did not use the brake immediately following an engine failure:

- The airplane simulated was propeller-driven (turboprop). For this type of airplane, the drag due to the wind milling propellers is quite high and provides enough braking force to the airplane. Hence, once the PF lowered the power level of the good engine following the engine failure and felt that the drag experienced by the airplane was quite high, he did not feel the urgency to hit the brake pedals.
- The airplane also has auto spoilers. The spoilers are auto-armed (in ready position) when the engines are at full throttle. They are deployed automatically when a throttle is retarded and the aircraft is still on the ground. The spoilers generate additional drag, which may delay PF use of the brake pedals.

- The runway length for the experiment was about 1000 ft longer than the critical length for such an airplane in the take-off configuration used. The extra runway length above the critical value might have influenced PFs' decision on when to start pressing the brake pedals.
- The FAA standard says that in case the engine failure occurs before V_1, the PF should reduce the power (of the good engine) smoothly and promptly. This standard is subject to personal interpretation. There is no requirement on using the brake pedals immediately after the engine failure.
- In some maneuvers, the PNFs interfered with the braking process. The table below displays some maneuvers where PNF interference was observed. The table shows that on some maneuvers, the PNF applied the brake before the PF.

			1st Time Pilot Hit the Brake After Engine Failure		Max Brake Force	
Maneuver	Crew	Brake Time (sec)	Pilot (sec)	Co-Pilot (sec)	Pilot Max (lbs)	Co-Pilot Max (lbs)
31	6	4.76	4.36	26.68	94.836	67.425
32	6	3.62	3.48	17.70	89.723	81.860
37	6	2.92	2.70	18.84	73.295	11.943
144	31	13.10	11.98	7.02	62.242	49.459
145	31	16.50	16.50	4.82	87.060	43.964
151	31	14.62	14.30	5.44	66.903	43.964
170	35	14.54	13.80	0.52	63.173	122.030
176	35	21.32	21.10	2.30	45.918	131.840
187	37	7.72	7.56	0.06	52.607	0.363
308	52	2.06	1.90	29.42	77.471	0.545
333	59	3.86	3.62	18.90	41.105	11.626

APPENDIX L - RTO LOGISTIC REGRESSION RESULTS

Notes: sle=significant level of entry
sls=significant level of stay
DF=degrees-of-freedom
See Appendix I for the complete description of the abbreviated RTO measures

Motion On

1. All measures offered, Stepwise (sle=sls=.1): Selected Mean Abs Lateral Deviation and RMS Heading Deviation

-2 log likelihood = 107.399
Chi-Square (df=2) = 16.286
Pr > Chi-Square = 0.0003

Measure	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCP1	1	-4.6186	0.9126	25.6113	0.0001		
INTERCP2	1	-2.3353	0.6939	11.3255	0.0008		
INTERCP3	1	2.3459	1.1120	4.4503	0.0349		
M_ABS31	1	0.0719	0.0379	3.6084	0.0575	0.3477	1.075
RMS6C	1	0.3559	0.2110	2.8455	0.0916	0.4018	1.428
Residual Chi-Square (df=9) = 4.4445 Pr > Chi-Square = 0.8798							

2. Stepwise, forcing (one by one) Max Heading Deviation, Power Lever Reaction Time, Pedal Reaction Time, STD Pedal Position, and Integrated Yaw Activity: No significance

Motion Off

1. All measures offered, Stepwise (sle=sls=.1): Selected Mean Abs Lateral Deviation

-2 log likelihood = 47.074
Chi-Square (df=1) = 8.686
Pr > Chi-Square = 0.0032

Measure	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCP1	1	-5.3289	1.5534	11.7682	0.0006		
INTERCP2	1	-1.6247	0.7757	4.3862	0.0362		
INTERCP3	1	1.7629	1.1350	2.4126	0.1204		
M_ABS31	1	0.1307	0.0500	6.8332	0.0089	0.7284	1.140
Residual Chi-Square (df=10) = 8.2182 Pr > Chi-Square = 0.6075							

2. Stepwise, forcing STD Pedal Position: Selected STD Pedal Position and RMS Heading Deviation

-2 log likelihood = 44.965
Chi-Square (df=2) = 10.796
Pr > Chi-Square = 0.0045

Measure	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCP1	1	-7.3050	2.1254	11.8133	0.0006		
INTERCP2	1	-3.4289	1.3429	6.5199	0.0107		
INTERCP3	1	0.1207	1.4895	0.0066	0.9354		
SD39	1	1.7695	1.0241	2.9857	0.0840	0.4738	5.868
RMS6C	1	0.3984	0.2051	3.7746	0.0520	0.5498	1.489
Residual Chi-Square (df=9) = 6.4827 Pr > Chi-Square = 0.6908							

3. Stepwise, forcing (one by one) Power Lever Reaction Time, Pedal Reaction Time, Max Heading Deviation, STD Wheel Position, and Sum of Pedal Spectrum f>.5 Hz: No significance

APPENDIX M - V_1 CUT LOGISTIC REGRESSION RESULTS

Notes: sle=significant level of entry
sls=significant level of stay
DF=degrees-of-freedom
See Appendix J for the complete description of the abbreviated V_1 Cut measure

<u>Motion On</u>

1. All measures offered, Stepwise (sle=sls=.1): Selected Time to Reach 400 ft Altitude, STD Column Position, STD Pitch Angle, and Integrated Airspeed Exceedance

-2 log likelihood = 129.913
Chi-Square (df=4) = 26.444
Pr > Chi-Square = 0.0001

Measure	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCP1	1	-15.6179	3.4012	21.0851	0.0001		
INTERCP2	1	-12.9880	3.2052	16.4197	0.0001		
TALT	1	0.1518	0.0477	10.1515	0.0014	0.6570	1.164
SD37	1	2.9005	1.2853	5.0923	0.0240	0.3082	18.183
SD8	1	0.8889	0.3310	7.2132	0.0072	0.4292	2.432
TOB1	1	0.0311	0.0108	8.2821	0.0040	0.4551	1.032
Residual Chi-Square (df=17) = 21.9791 Pr > Chi-Square = 0.1855							
Note: *All are longitudinal measures*							

2. Stepwise, forcing RMS Heading Deviation: Selected RMS Heading Deviation, Time to Reach 400 ft Altitude, STD Column Position, STD Pitch Angle, STD Wheel Position, Integrated Airspeed Exceedance, Integrated Roll Activity, and Sum of Wheel Spectrum f>.5 Hz

-2 log likelihood = 116.722
Chi-Square (df=8) = 39.635
Pr > Chi-Square = 0.0001

Measure	DF	Parameter	Standard	Wald	Pr >	Standardized	Odds

		Estimate	Error	Chi-Square	Chi-Square	Estimate	Ratio
INTERCP1	1	-20.2696	3.9958	25.7324	0.0001		
INTERCP2	1	-17.2232	3.7226	21.4057	0.0001		
RMS6C	1	0.3710	0.1299	8.1595	0.0043	0.5896	1.449
TALT	1	0.1696	0.0522	10.5614	0.0012	0.7337	1.185
SD37	1	4.8588	1.5328	10.0482	0.0015	0.5162	128.873
SD8	1	1.6559	0.4239	15.2556	0.0001	0.7994	5.238
SD38	1	-1.0001	0.3291	9.2368	0.0024	-1.4964	0.368
TOB1	1	0.0352	0.0120	8.6690	0.0032	0.5160	1.036
S_ABS12	1	0.0821	0.0365	5.0660	0.0244	0.8790	1.086
TMED38	1	0.0282	0.0140	4.0754	0.0435	0.4515	1.029
Residual Chi-Square (df=13) = 13.6269 Pr > Chi-Square = 0.4006							
Note: *SD38 is positively associated*							

3. Stepwise, forcing Integrated Bank Angle Exceedance: Selected Integrated Bank Angle Exceedance, Time to Reach 400 ft Altitude, Max Abs Bank Angle, STD Column Position, STD Pitch Angle, and Integrated Airspeed Exceedance

-2 log likelihood = 123.802
Chi-Square (df=6) = 32.555
Pr > Chi-Square = 0.0001

Measure	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCP1	1	-14.9442	3.5763	17.4613	0.0001		
INTERCP2	1	-12.0961	3.3776	12.8257	0.0003		
TOB11	1	0.0732	0.0345	4.5066	0.0338	0.5764	1.076
TALT	1	0.1540	0.0514	8.9910	0.0027	0.6665	1.167
MX11	1	-0.2906	0.1183	6.0397	0.0140	-0.6797	0.748
SD37	1	3.6347	1.3686	7.0526	0.0079	0.3862	37.890
SD8	1	0.9767	0.3512	7.7356	0.0054	0.4716	2.656
TOB1	1	0.0336	0.0111	9.1089	0.0025	0.4919	1.034
Residual Chi-Square (df=15) = 19.8901 Pr > Chi-Square = 0.1762							
Note: *MX11 is positively associated*							

4. Stepwise, forcing (one by one) Wheel Reaction Time, Pedal Reaction Time, RMS Heading Deviation, STD Pedal Position, Integrated Failure Induced Bank Angle, Integrated Roll Activity, Integrated Yaw Activity, Column Reversals, Wheel Reversals, Pedal Reservals, and Sum of Pedal Spectrum f>.5 Hz: No significance

<u>Motion Off</u>

1. All measures offered, Stepwise (sle=sls=.1): None selected

2. Stepwise, forcing Pedal Reaction Time: Selected Pedal Reaction Time and Time to Reach 400 ft Altitude

-2 log likelihood = 83.567
Chi-Square (df=2) = 6.272
Pr > Chi-Square = 0.0435

Measure	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCP1	1	-6.1429	2.3339	6.9278	0.0085		
INTERCP2	1	-3.6461	2.2005	2.7454	0.0975		
RT39	1	0.5480	0.3112	3.1006	0.0783	0.3178	1.730
TALT	1	0.1035	0.0491	4.4488	0.0349	0.4570	1.109
Residual Chi-Square = 15.0387 with 19 DF Pr > Chi-Square = 0.7201							

3. Stepwise, forcing Max Abs Bank Angle: Selected Max Abs Bank Angle, STD Wheel Position, and Integrated Bank Angle Exceedance

-2 log likelihood = 80.463
Chi-Square (df=3) = 9.376
Pr > Chi-Square = 0.0247

Measure	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCP1	1	-1.5848	1.4023	1.2773	0.2584		
INTERCP2	1	1.0902	1.3908	0.6145	0.4331		
MX11	1	-0.4111	0.1730	5.6495	0.0175	-0.9282	0.663
SD38	1	0.4034	0.1838	4.8152	0.0282	0.6076	1.497
TOB11	1	0.0836	0.0466	3.2200	0.0727	0.5990	1.087
Residual Chi-Square (df=18) = 13.8749 Pr > Chi-Square = 0.7372							
Note: *MX11 is positively associated*							

4. Stepwise, forcing (one by one) Wheel Reaction Time, RMS Heading Deviation, Integrated Bank Angle Exceedance, Integrated Failure Induced Bank Angle, Integrated Roll Activity, and Integrated Yaw Activity: No significance

APPENDIX N - RTO FIRST LOOK EVALUATION DATA SUMMARY

Motion Group		
Measures	**Mean**	**SE**
Mean Abs Lateral Deviation (ft)	16.898	2.425
RMS Lateral Deviation (ft)	20.399	3.270
Integrated Yaw Activity (deg)	48.408	5.200
RMS Heading Deviation (deg)	3.658	0.299
Max Heading Deviation (deg)	8.249	0.789
Power Lever Reaction Time (sec)	2.033	0.165
Pedal Reaction Time (sec)	1.511	0.172
STD Pedal Position (in)	1.321	0.136
Number of Pedal Reversals	2.500	0.365
Sum of Pedal Spectrum f>0.5 Hz (in.sec)	0.402	0.053

Table N-1. Motion Group: Means and Standard Errors

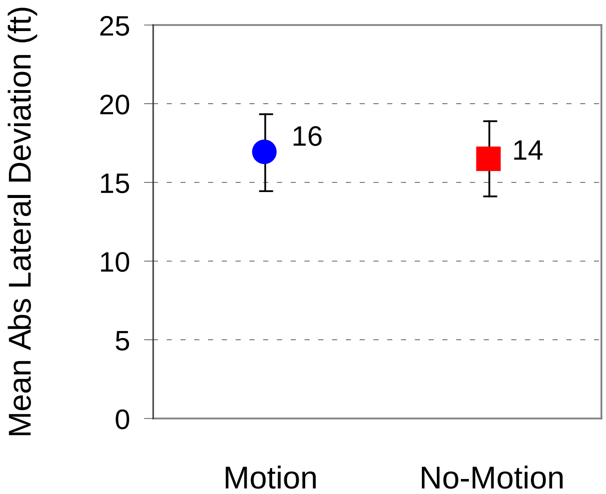
No-Motion Group		
Measures	**Mean**	**SE**
Mean Abs Lateral Deviation (ft)	16.488	2.416
RMS Lateral Deviation (ft)	19.501	2.969
Integrated Yaw Activity (deg)	68.495	7.519
RMS Heading Deviation (deg)	4.492	0.450
Max Heading Deviation (deg)	10.556	1.220
Power Lever Reaction Time (sec)	2.287	0.233
Pedal Reaction Time (sec)	1.330	0.270
STD Pedal Position (in)	1.219	0.148
Number of Pedal Reversals	2.286	0.474
Sum of Pedal Spectrum f>0.5 Hz (in.sec)	0.394	0.103

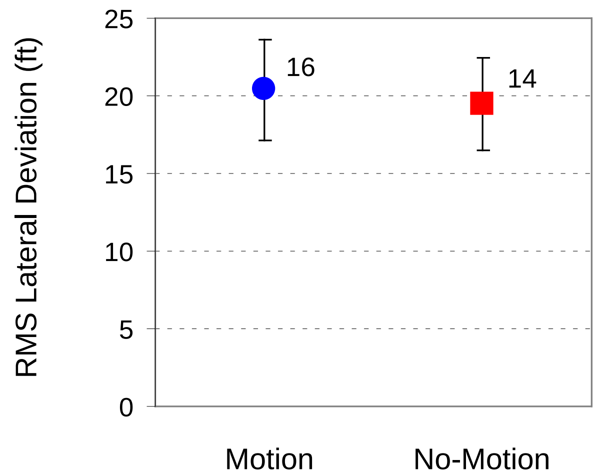
Table N-2. No-Motion Group: Means and Standard Errors

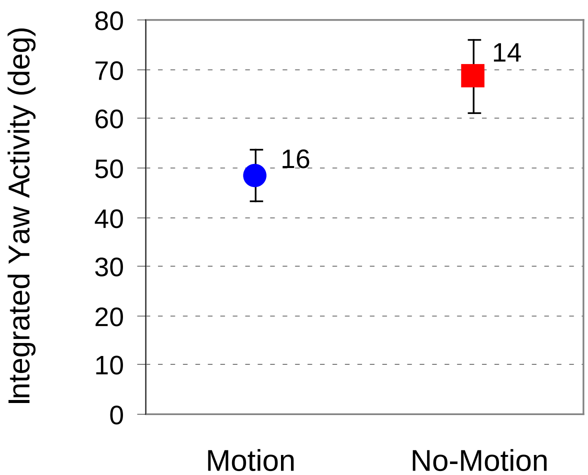
Measures	**df_{within}**	**ss_{within}**	**$df_{between}$**	**$ss_{between}$**	**F**	***p***
Mean Abs Lateral Deviation	28	2472.914	1	1.254	0.014	0.906
RMS Lateral Deviation	28	4170.330	1	6.033	0.041	0.842
Integrated Yaw Activity	28	16777.668	1	3012.796	5.028	0.033
RMS Heading Deviation	28	58.305	1	5.187	2.491	0.126
Max Heading Deviation	28	420.307	1	39.739	2.647	0.115
Power Lever Reaction Time	28	16.351	1	0.484	0.829	0.370
Pedal Reaction Time	28	20.384	1	0.245	0.337	0.566
STD Pedal Position	28	8.413	1	0.078	0.260	0.614
Pedal Reversals	28	72.857	1	0.343	0.132	0.719
Sum of Pedal Spectrum f>0.5 Hz	28	2.623	1	0.000	0.005	0.944

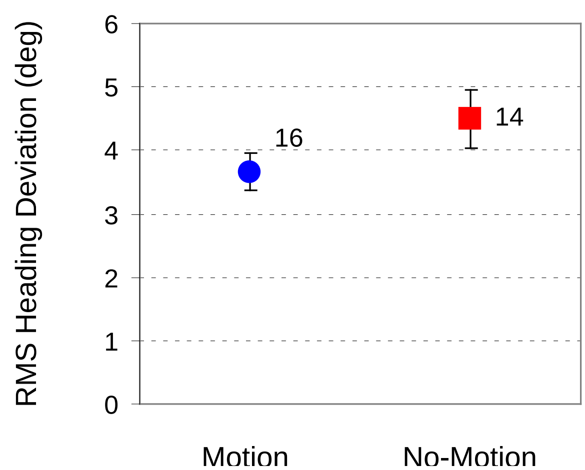
Table N-3. Group Differences: ANOVA Results

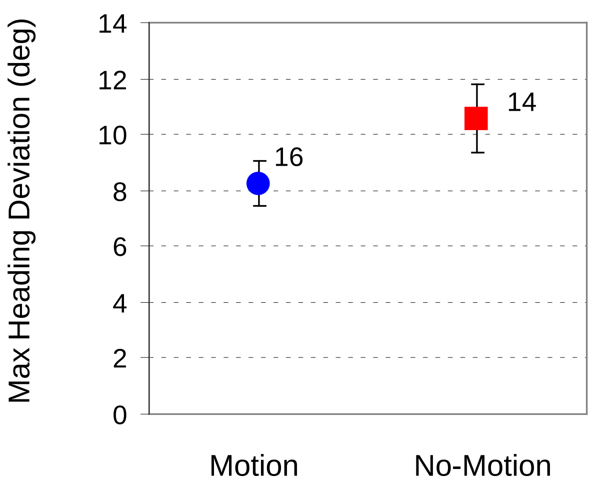
RTO First Look
(Note: *Sample sizes are shown next to each data point*)

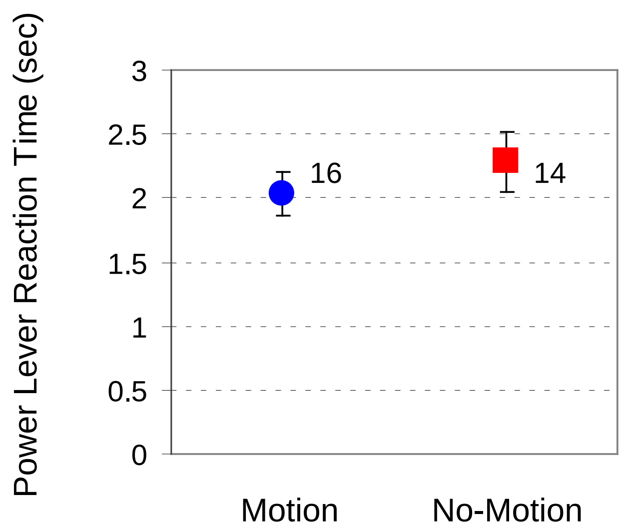




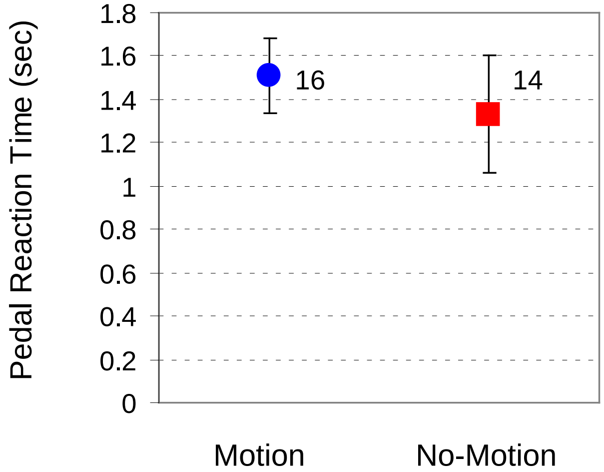


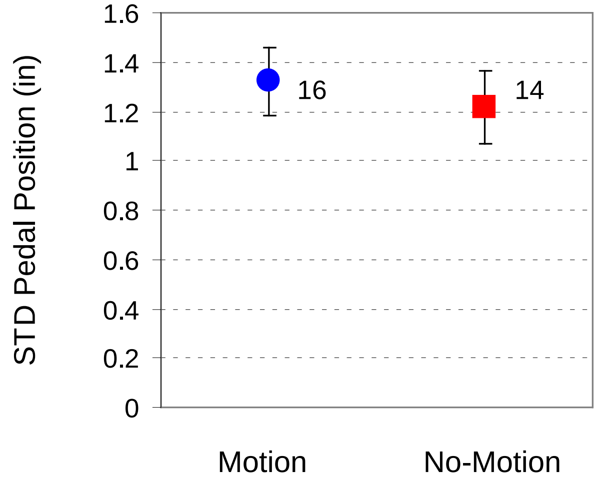


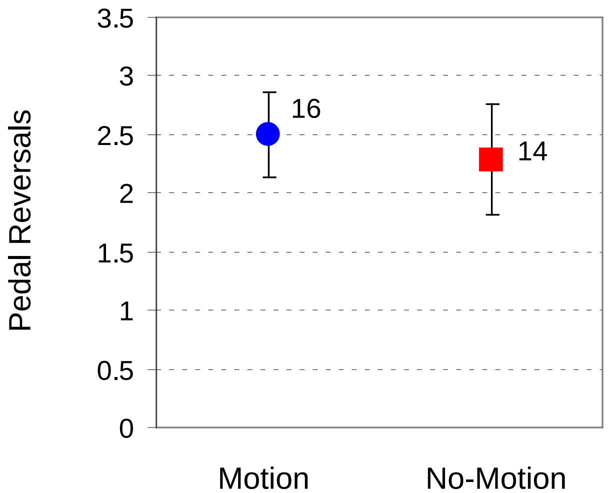


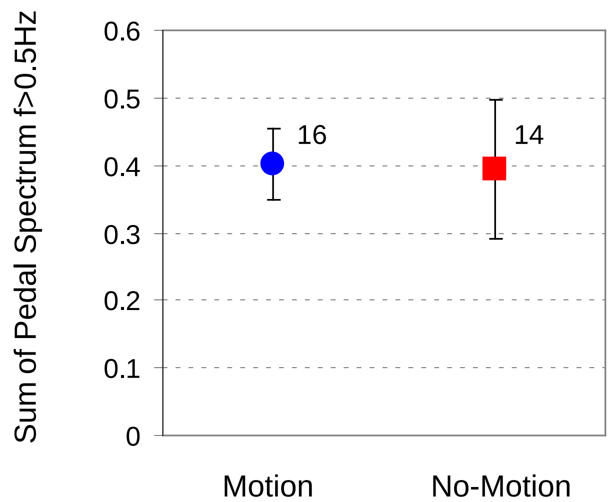


RTO First Look
(Note: *Sample sizes are shown next to each data point*)









APPENDIX O - V_1 CUT FIRST LOOK EVALUATION DATA SUMMARY

Motion Group		
Measures	**Mean**	**SE**
Max Abs Bank Angle (deg)	12.476	0.989
Integrated Bank Angle Exceedance (deg.sec)	21.230	3.768
Integrated Failure Induced Bank Angle (deg.sec)	54.275	7.027
Integrated Roll Activity (deg)	60.926	5.503
Integrated Yaw Activity (deg)	21.283	1.857
RMS Heading Deviation (deg)	4.759	0.708
Max Heading Deviation (deg)	8.160	0.989
Integrated Heading Exceedance (deg.sec)	15.240	7.511
Integrated Airspeed Exceedance (kts.sec)	28.605	8.004
Time to Reach 400 ft. Altitude (sec)	40.171	2.076
STD Pitch Angle (deg)	4.241	0.195
Wheel Reaction Time (sec)	2.286	0.285
Pedal Reaction Time (sec)	1.919	0.270
STD Column Position (in)	1.297	0.039
STD Wheel Position (deg)	9.856	0.725
STD Pedal Position (in)	0.985	0.079
Number of Column Reversals	4.444	0.315
Number of Wheel Reversals	2.778	0.375
Number of Pedal Reversals	1.556	0.271
Sum of Column Spectrum f>0.5 Hz (in.sec)	0.958	0.121
Sum of Wheel Spectrum f>0.5 Hz (deg.sec)	0.012	0.002
Sum of Pedal Spectrum f>0.5 Hz (in.sec)	0.318	0.035

Table O-1. Motion Group: Means and Standard Errors

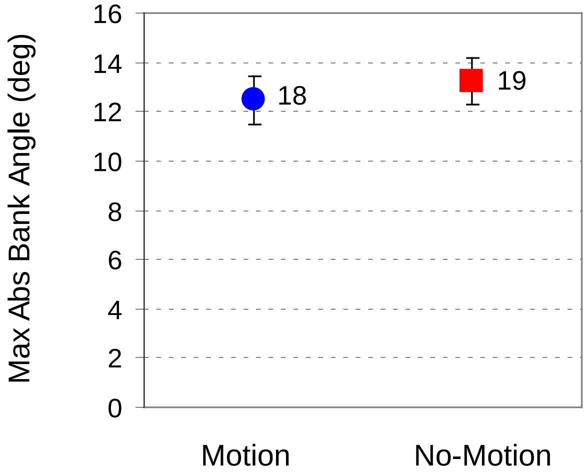
No-Motion Group		
Measures	**Mean**	**SE**
Max Abs Bank Angle (deg)	13.218	0.961
Integrated Bank Angle Exceedance (deg.sec)	21.752	3.577
Integrated Failure Induced Bank Angle (deg.sec)	58.743	5.959
Integrated Roll Activity (deg)	64.890	5.047
Integrated Yaw Activity (deg)	24.089	1.738
RMS Heading Deviation (deg)	5.832	0.664
Max Heading Deviation (deg)	10.092	1.009
Integrated Heading Exceedance (deg.sec)	25.372	6.296
Integrated Airspeed Exceedance (kts.sec)	28.767	12.368
Time to Reach 400 ft. Altitude (sec)	36.945	3.550
STD Pitch Angle (deg)	4.928	0.344
Wheel Reaction Time (sec)	2.754	0.266
Pedal Reaction Time (sec)	2.175	0.206
STD Column Position (in)	1.215	0.058
STD Wheel Position (deg)	10.647	0.682
STD Pedal Position (in)	1.009	0.047
Number of Column Reversals	3.789	0.311
Number of Wheel Reversals	3.000	0.315
Number of Pedal Reversals	1.105	0.105
Sum of Column Spectrum f>0.5 Hz (in.sec)	0.702	0.158
Sum of Wheel Spectrum f>0.5 Hz (deg.sec)	0.014	0.002
Sum of Pedal Spectrum f>0.5 Hz (in.sec)	0.324	0.039

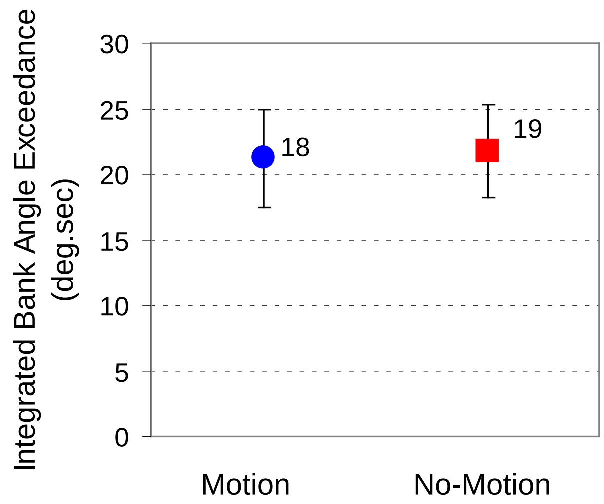
Table O-2. No-Motion Group: Means and Standard Errors

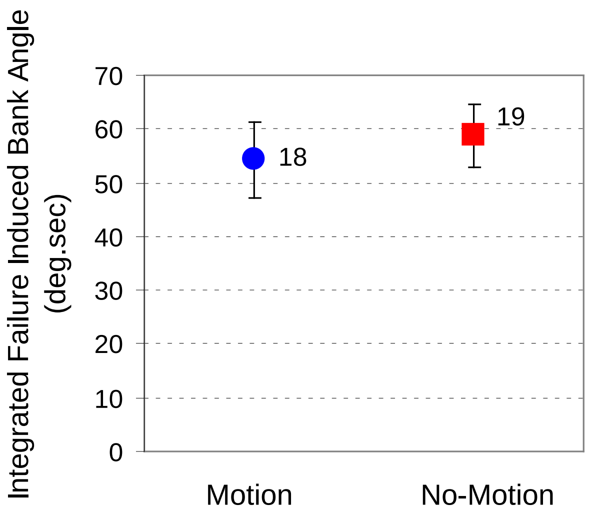
Measures	df $_{within}$	ss $_{within}$	df $_{between}$	ss $_{between}$	F	*p*
Max Abs Bank Angle	35	614.582	1	5.082	0.289	0.594
Integrated Bank Angle Exceedance	35	8720.214	1	2.519	0.010	0.920
Integrated Failure Induced Bank Angle	35	27254.715	1	184.577	0.237	0.629
Integrated Roll Activity	35	17976.296	1	145.248	0.283	0.598
Integrated Yaw Activity	35	2088.429	1	72.782	1.220	0.277
RMS Heading Deviation	35	304.055	1	10.645	1.225	0.276
Max Heading Deviation	35	647.590	1	34.497	1.864	0.181
Integrated Heading Exceedance	35	30820.008	1	948.917	1.078	0.306
Integrated Airspeed Exceedance	35	71920.472	1	0.241	0.000	0.991
Time to Reach 400 ft. Altitude	35	1905.718	1	96.186	1.767	0.192
STD Pitch Angle	35	52.268	1	4.366	2.923	0.096
Wheel Reaction Time	35	49.005	1	2.026	1.447	0.237
Pedal Reaction Time	35	36.799	1	0.605	0.575	0.453
STD Column Position	35	1.624	1	0.062	1.338	0.255
STD Wheel Position	35	320.197	1	5.774	0.631	0.432
STD Pedal Position	35	2.649	1	0.005	0.072	0.790
Column Reversals	35	63.602	1	3.965	2.182	0.149
Wheel Reversals	35	77.111	1	0.456	0.207	0.652
Pedal Reversals	35	26.234	1	1.874	2.500	0.123
Sum of Column Spectrum f>0.5 Hz	35	12.960	1	0.607	1.639	0.209
Sum of Wheel Spectrum f>0.5 Hz	35	28431.685	1	210.557	0.259	0.614
Sum of Pedal Spectrum f>0.5 Hz	35	0.882	1	0.000	0.016	0.899

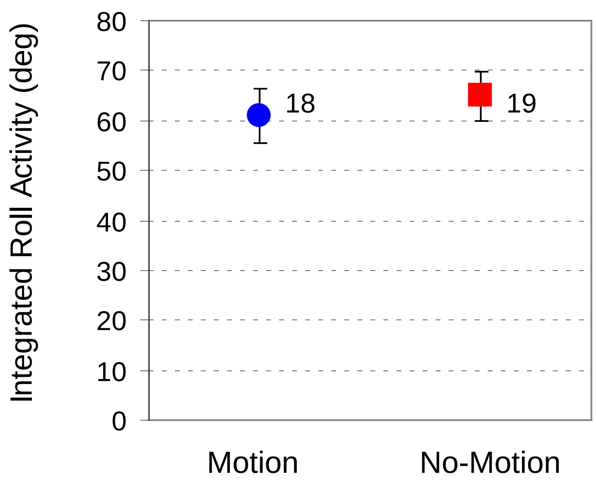
Table O-3. Group Differences: ANOVA Results

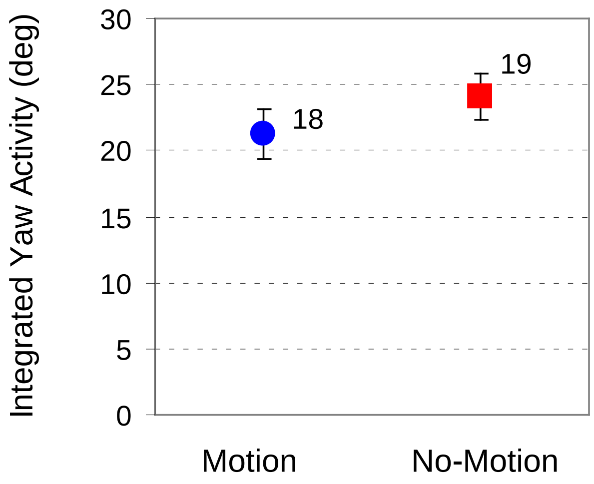
V1 Cut First Look
(Note: *Sample sizes are shown next to each data point*)

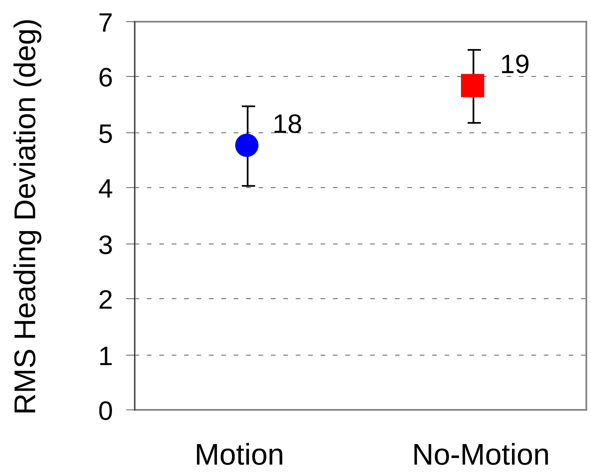




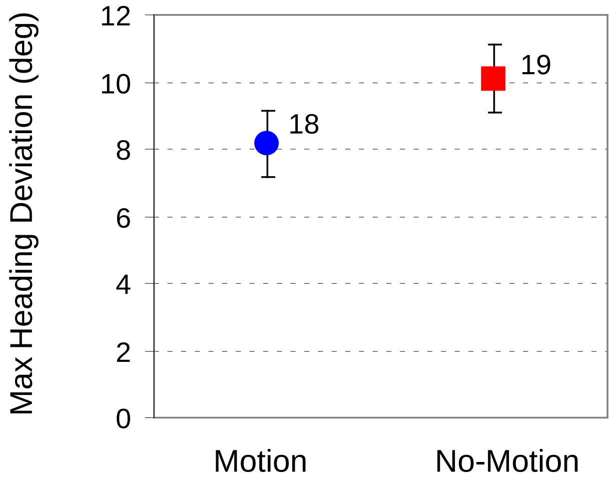


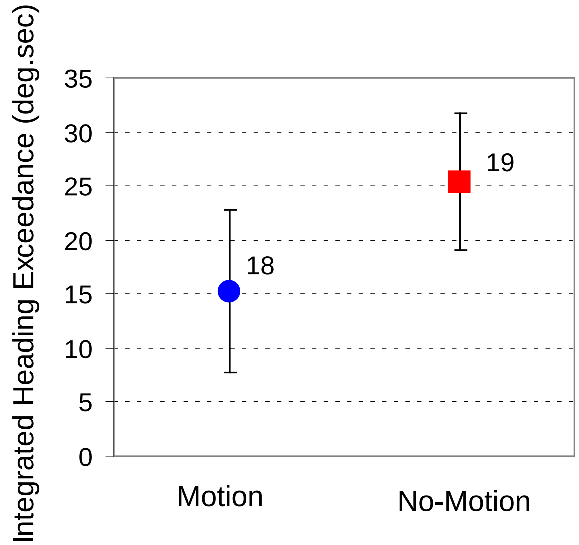


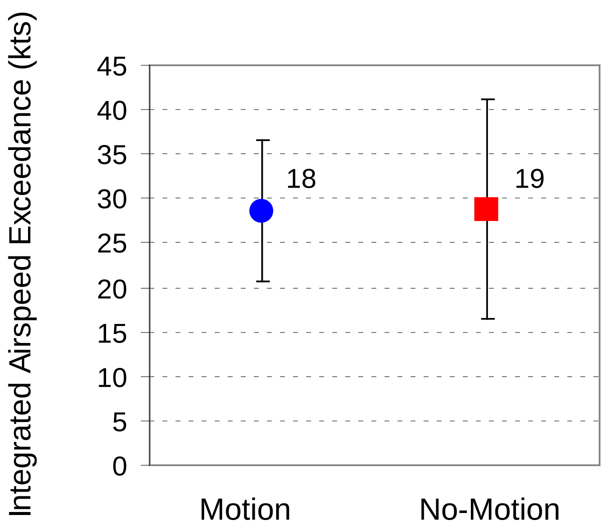


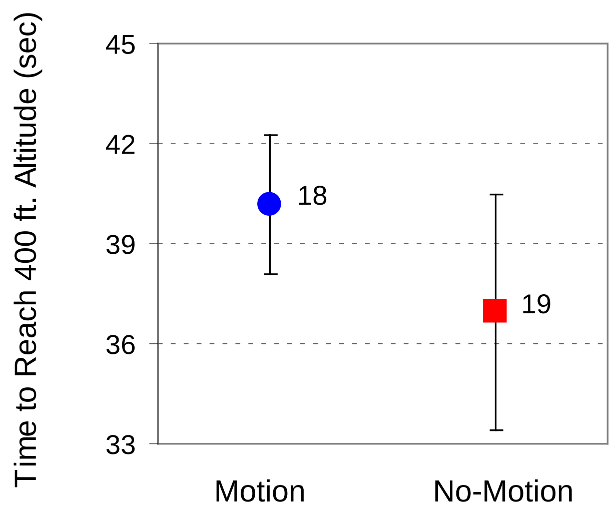


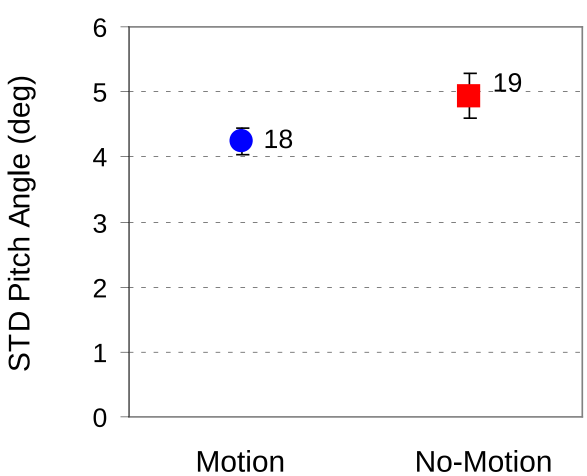
V1 Cut First Look
(Note: *Sample sizes are shown next to each data point*)

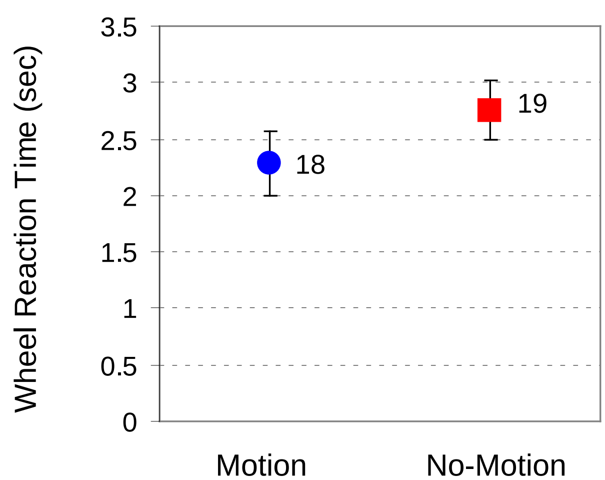




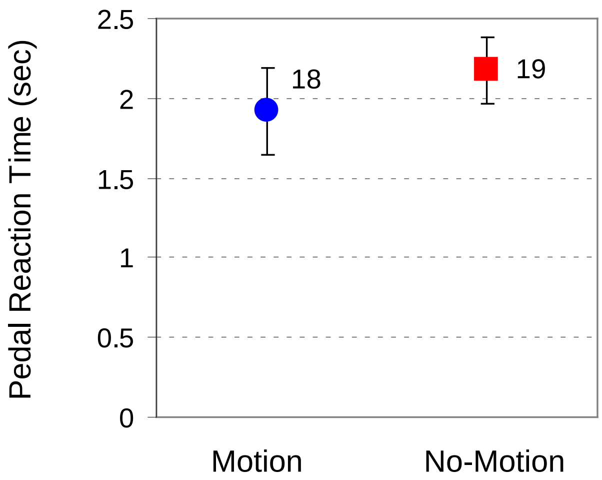


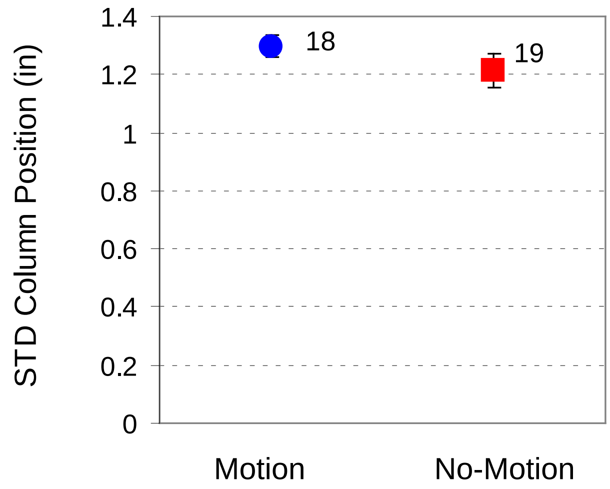


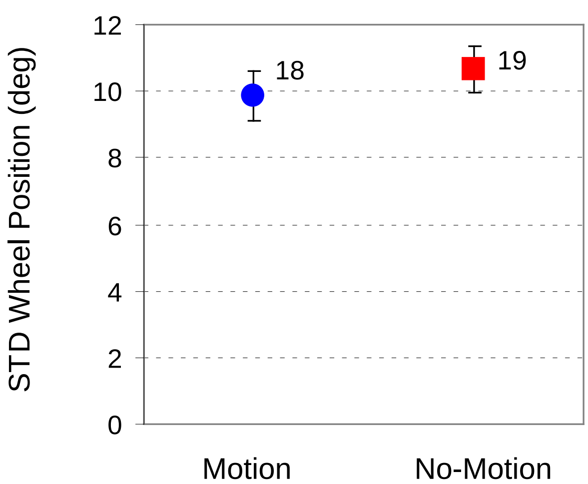


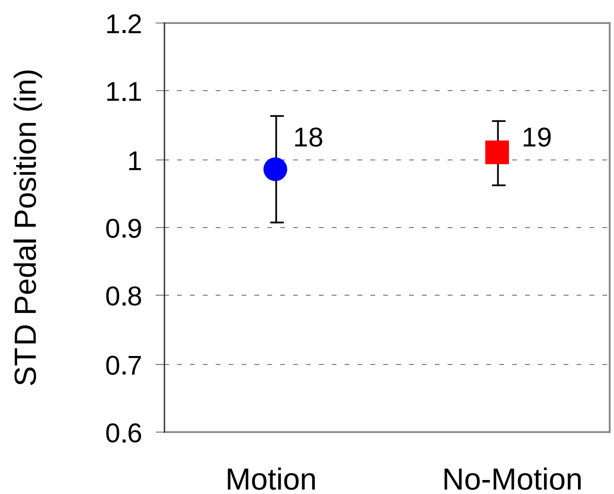


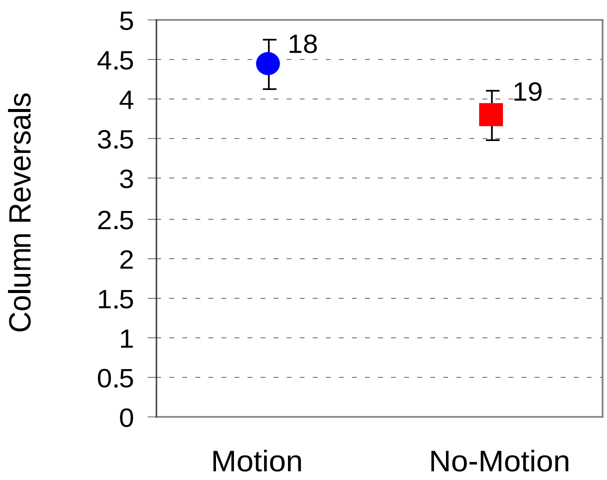
V1 Cut First Look
(Note: *Sample sizes are shown next to each data point*)

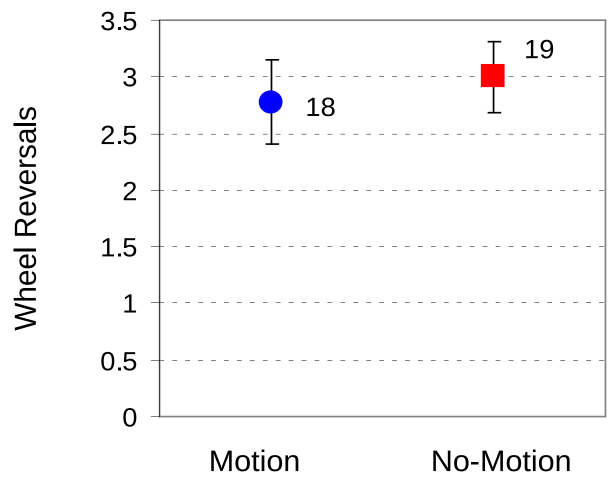




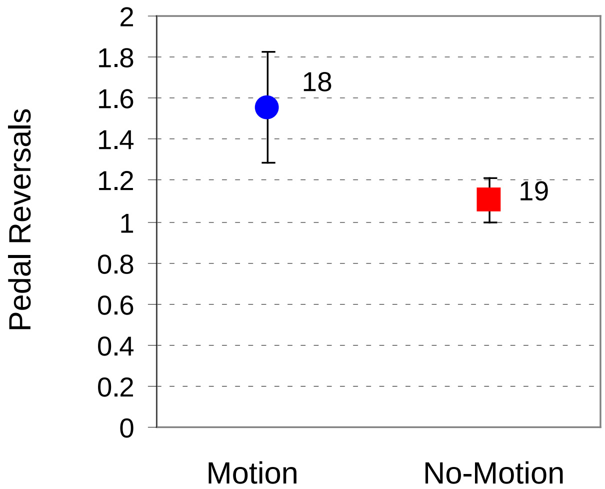


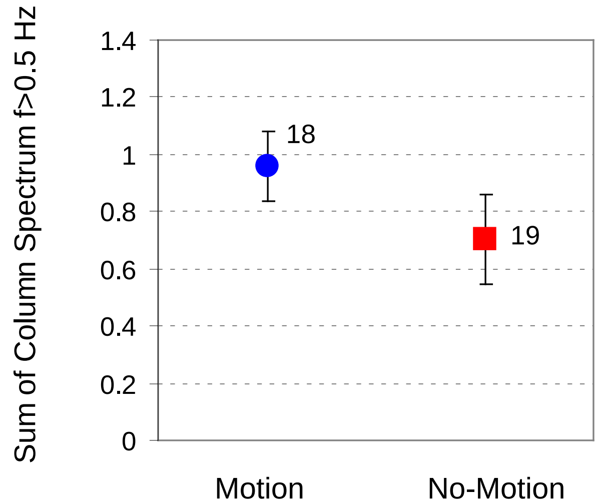


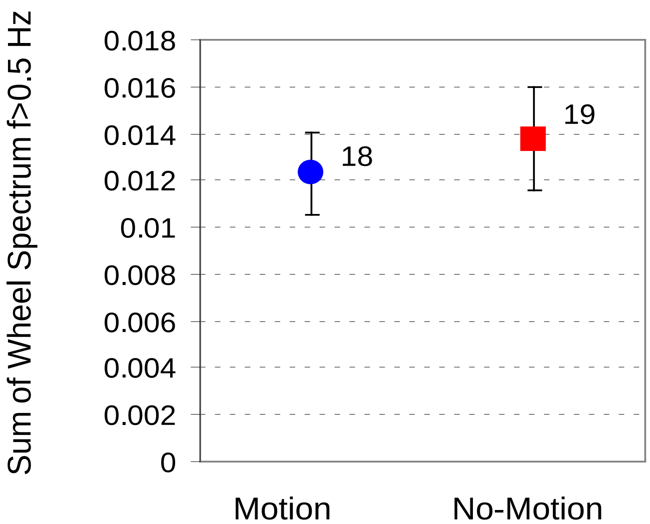


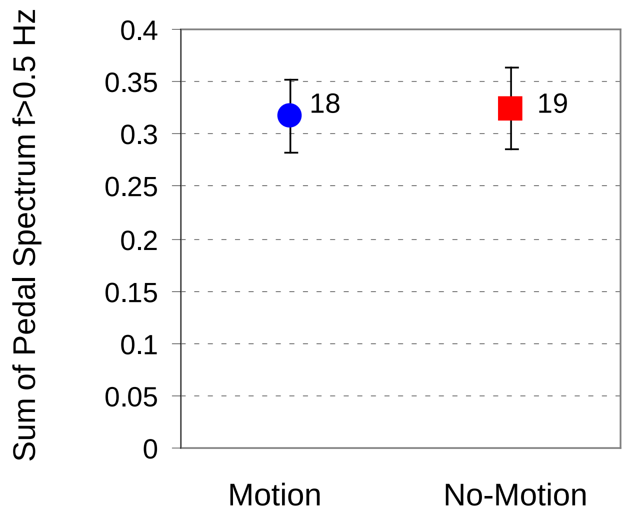


V1 Cut First Look
(Note: *Sample sizes are shown next to each data point*)









APPENDIX P – RTO TRANSFER TESTING DATA SUMMARY

Motion Group		
Measures	**Mean**	**SE**
Mean Abs Lateral Deviation (ft)	9.763	1.923
RMS Lateral Deviation (ft)	12.042	2.425
Integrated Yaw Activity (deg)	51.971	5.334
RMS Heading Deviation (deg)	2.971	0.257
Max Heading Deviation (deg)	5.984	0.485
Power Lever Reaction Time (sec)	1.948	0.246
Pedal Reaction Time (sec)	0.976	0.129
STD Pedal Position (in)	1.011	0.125
Number of Pedal Reversals	3.250	0.487
Sum of Pedal Spectrum f>0.5 Hz (in.sec)	0.370	0.138

Table P-1. Motion Group: Means and Standard Errors

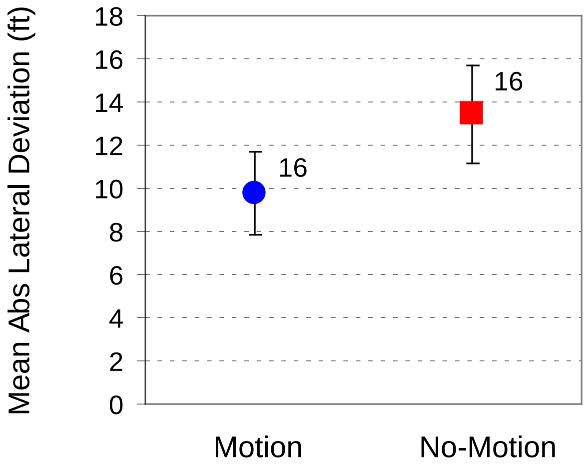
No-Motion Group		
Measures	**Mean**	**SE**
Mean Abs Lateral Deviation (ft)	13.429	2.273
RMS Lateral Deviation (ft)	16.794	3.166
Integrated Yaw Activity (deg)	49.380	4.527
RMS Heading Deviation (deg)	3.926	0.823
Max Heading Deviation (deg)	6.346	0.885
Power Lever Reaction Time (sec)	2.040	0.171
Pedal Reaction Time (sec)	0.993	0.091
STD Pedal Position (in)	0.840	0.136
Number of Pedal Reversals	3.000	0.570
Sum of Pedal Spectrum f>0.5 Hz (in.sec)	0.181	0.040

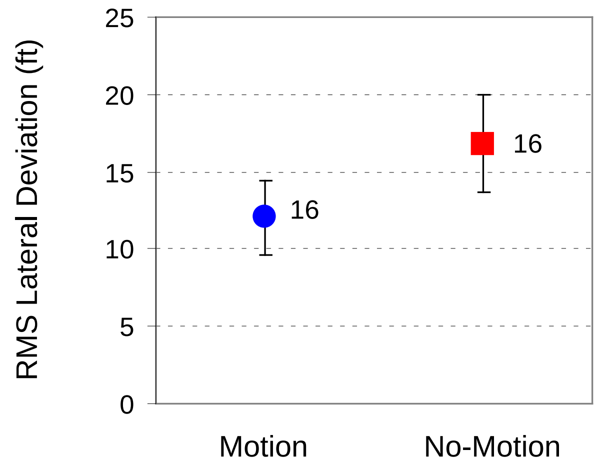
Table P-2. No-Motion Group: Means and Standard Errors

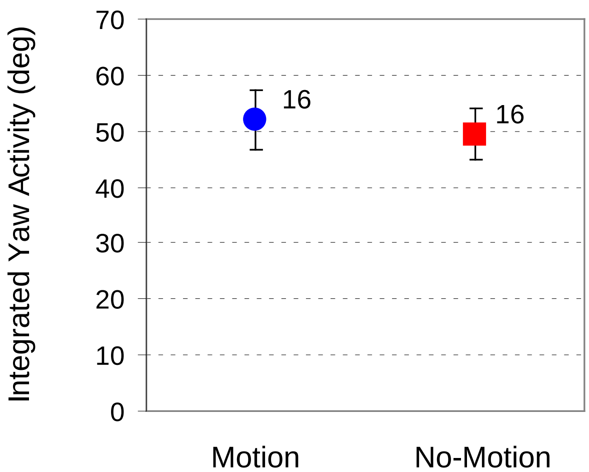
Measures	df_{within}	ss_{within}	$df_{between}$	$ss_{between}$	**F**	***p***
Mean Abs Lateral Deviation	30	2127.809	1	107.531	1.516	0.228
RMS Lateral Deviation	30	3817.301	1	180.671	1.420	0.243
Integrated Yaw Activity	30	11748.951	1	53.713	0.137	0.714
RMS Heading Deviation	30	178.429	1	7.305	1.228	0.277
Max Heading Deviation	30	244.547	1	1.047	0.128	0.723
Power Lever Reaction Time	30	21.510	1	0.068	0.095	0.759
Pedal Reaction Time	30	5.982	1	0.002	0.011	0.919
STD Pedal Position	30	8.189	1	0.233	0.852	0.363
Pedal Reversals	30	135.000	1	0.500	0.111	0.741
Sum of Pedal Spectrum f>0.5 Hz	30	4.943	1	0.288	1.747	0.196

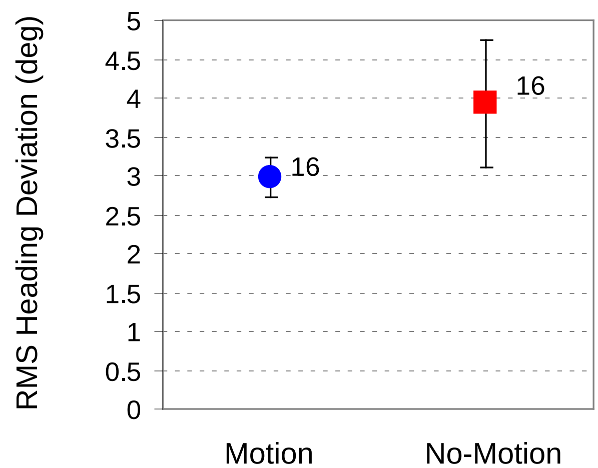
Table P-3. Group Differences: ANOVA Results

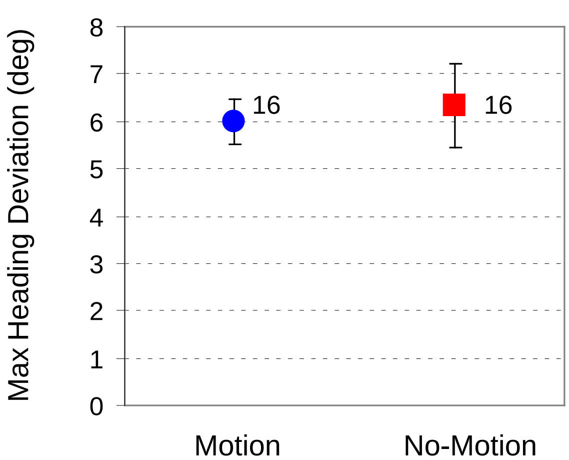
RTO Transfer
(Note: *Sample sizes are shown next to each data point*)

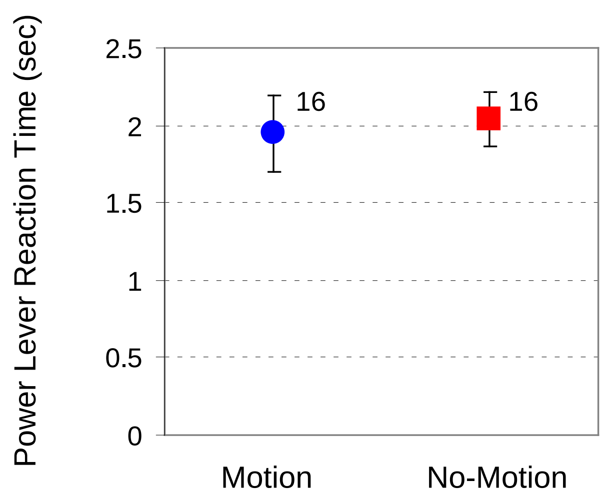




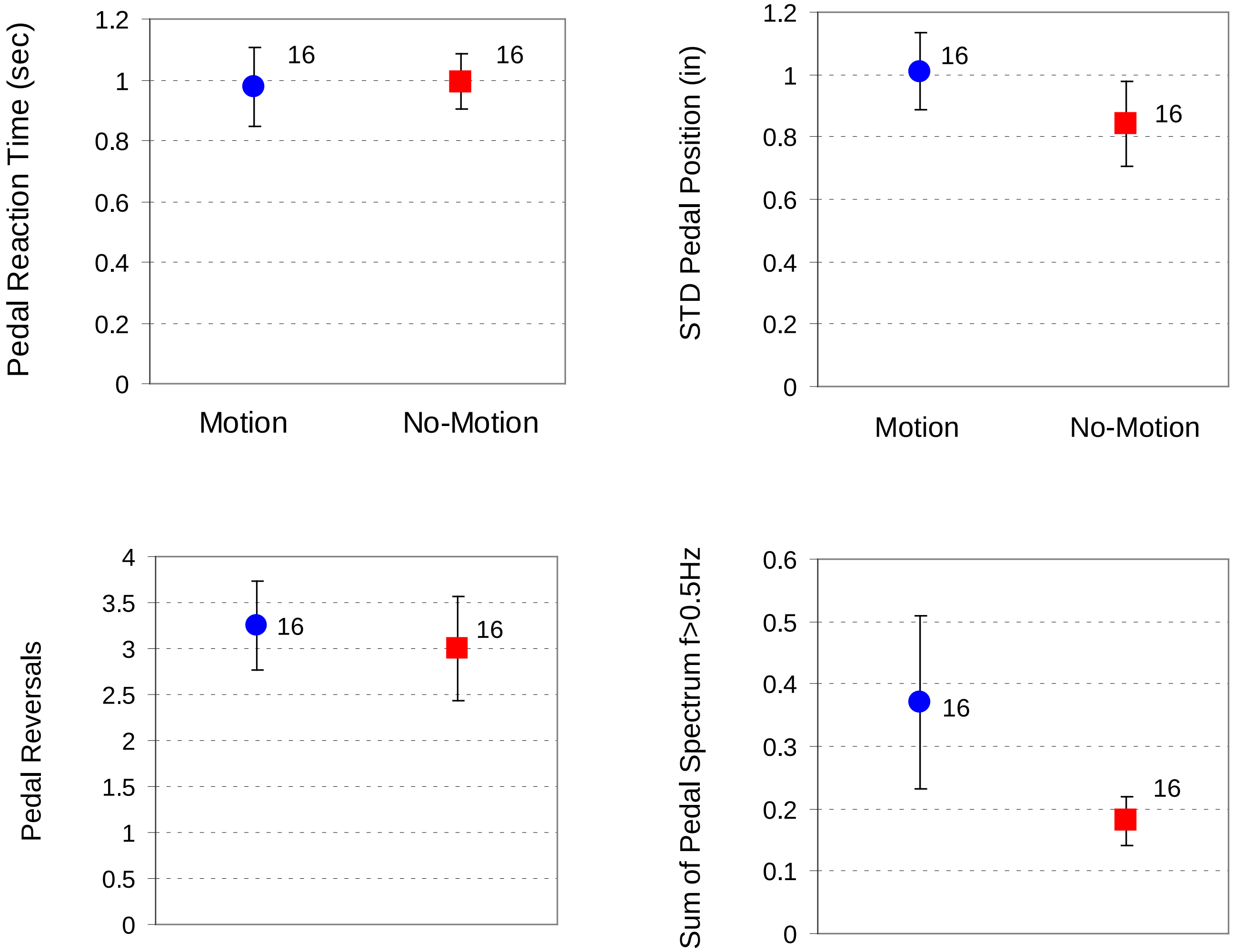








RTO Transfer
(Note: *Sample sizes are shown next to each data point*)



APPENDIX Q - V_1 CUT TRANSFER TESTING DATA SUMMARY

Motion Group		
Measures	**Mean**	**SE**
Max Abs Bank Angle (deg)	10.993	1.248
Integrated Bank Angle Exceedance (deg.sec)	12.646	3.236
Integrated Failure Induced Bank Angle (deg.sec)	28.387	4.290
Integrated Roll Activity (deg)	53.376	4.291
Integrated Yaw Activity (deg)	28.015	3.012
RMS Heading Deviation (deg)	6.444	0.889
Max Heading Deviation (deg)	11.231	1.604
Integrated Heading Exceedance (deg.sec)	31.820	8.746
Integrated Airspeed Exceedance (kts.sec)	11.670	2.920
Time to Reach 400 ft. Altitude (sec)	38.471	1.086
STD Pitch Angle (deg)	3.898	0.206
Wheel Reaction Time (sec)	2.443	0.318
Pedal Reaction Time (sec)	2.916	0.318
STD Column Position (in)	1.200	0.054
STD Wheel Position (deg)	9.475	0.702
STD Pedal Position (in)	0.903	0.076
Number of Column Reversals	3.750	0.281
Number of Wheel Reversals	3.063	0.295
Number of Pedal Reversals	1.563	0.157
Sum of Column Spectrum f>0.5 Hz (in.sec)	0.702	0.115
Sum of Wheel Spectrum f>0.5 Hz (deg.sec)	0.011	0.002
Sum of Pedal Spectrum f>0.5 Hz (in.sec)	0.240	0.039

Table Q-1. Motion Group: Means and Standard Errors

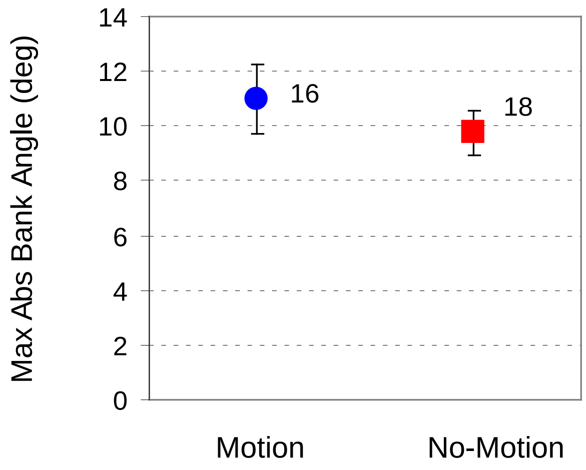
No-Motion Group		
Measures	**Mean**	**SE**
Max Abs Bank Angle (deg)	9.749	0.821
Integrated Bank Angle Exceedance (deg.sec)	8.283	2.032
Integrated Failure Induced Bank Angle (deg.sec)	34.524	4.577
Integrated Roll Activity (deg)	51.315	2.859
Integrated Yaw Activity (deg)	20.516	1.323
RMS Heading Deviation (deg)	5.405	0.678
Max Heading Deviation (deg)	8.856	0.882
Integrated Heading Exceedance (deg.sec)	21.136	7.219
Integrated Airspeed Exceedance (kts.sec)	36.498	7.413
Time to Reach 400 ft. Altitude (sec)	41.450	2.722
STD Pitch Angle (deg)	3.284	0.166
Wheel Reaction Time (sec)	2.527	0.140
Pedal Reaction Time (sec)	2.677	0.307
STD Column Position (in)	1.173	0.022
STD Wheel Position (deg)	8.588	0.478
STD Pedal Position (in)	0.743	0.070
Number of Column Reversals	4.111	0.290
Number of Wheel Reversals	4.056	0.400
Number of Pedal Reversals	0.944	0.151
Sum of Column Spectrum f>0.5 Hz (in.sec)	0.764	0.098
Sum of Wheel Spectrum f>0.5 Hz (deg.sec)	0.012	0.003
Sum of Pedal Spectrum f>0.5 Hz (in.sec)	0.201	0.036

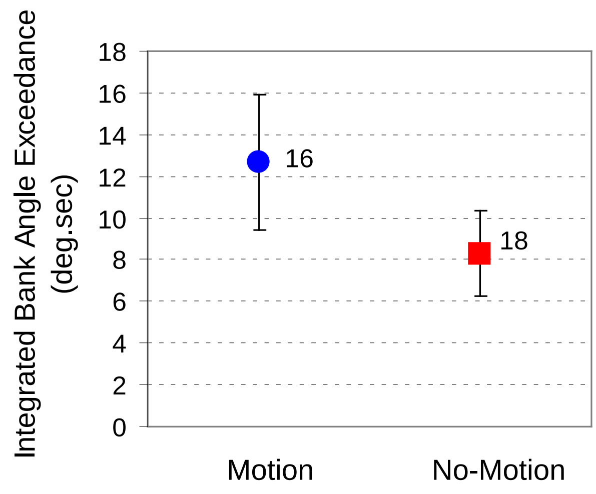
Table Q-2. No-Motion Group: Means and Standard Errors

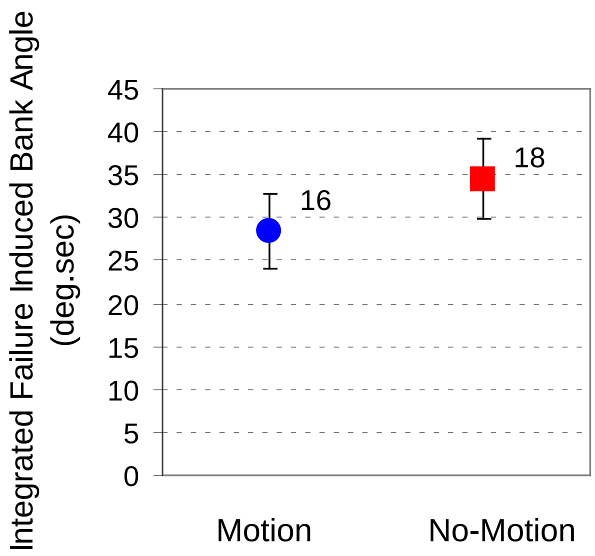
Measures	df within	ss within	df between	ss between	F	*p*
Max Abs Bank Angle	32	579.598	1	13.091	0.723	0.402
Integrated Bank Angle Exceedance	32	3777.058	1	161.193	1.366	0.251
Integrated Failure Induced Bank Angle	32	10827.440	1	319.022	0.943	0.339
Integrated Roll Activity	32	6921.369	1	35.991	0.166	0.686
Integrated Yaw Activity	32	2711.930	1	476.363	5.621	0.024
RMS Heading Deviation	32	331.384	1	9.142	0.883	0.354
Max Heading Deviation	32	855.526	1	47.782	1.787	0.191
Integrated Heading Exceedance	32	34306.322	1	966.857	0.902	0.349
Integrated Airspeed Exceedance	32	18860.699	1	5221.645	8.859	0.006
Time to Reach 400 ft. Altitude	32	2549.641	1	75.134	0.943	0.339
STD Pitch Angle	32	18.576	1	3.197	5.508	0.025
Wheel Reaction Time	32	30.310	1	0.060	0.063	0.803
Pedal Reaction Time	32	53.139	1	0.486	0.293	0.592
STD Column Position	32	0.855	1	0.006	0.234	0.632
STD Wheel Position	32	188.094	1	6.659	1.133	0.295
STD Pedal Position	32	2.859	1	0.216	2.418	0.130
Column Reversals	32	44.778	1	1.105	0.789	0.381
Wheel Reversals	32	69.882	1	8.353	3.825	0.059
Pedal Reversals	32	12.882	1	3.236	8.038	0.008
Sum of Column Spectrum f>0.5 Hz	32	6.096	1	0.032	0.170	0.683
Sum of Wheel Spectrum f>0.5 Hz	32	36841.275	1	245.093	0.213	0.648
Sum of Pedal Spectrum f>0.5 Hz	32	0.771	1	0.013	0.533	0.471

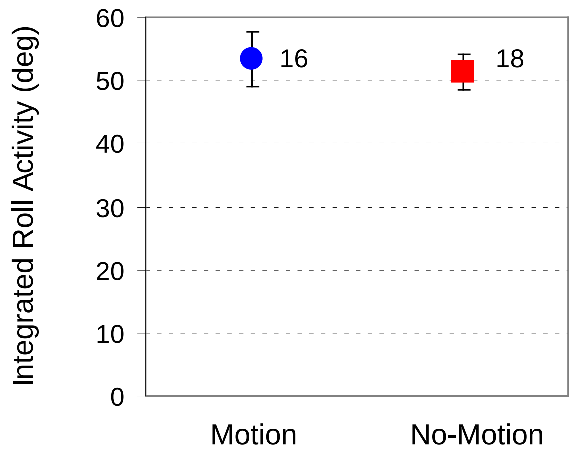
Table Q-3. Group Differences: ANOVA Results

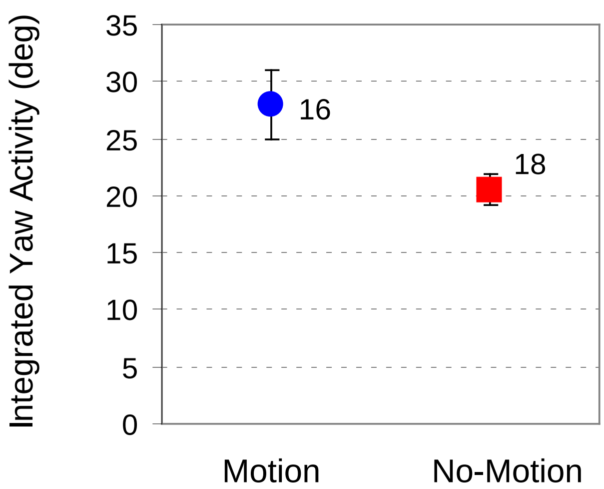
V1 Cut Transfer
(Note: *Sample sizes are shown next to each data point*)

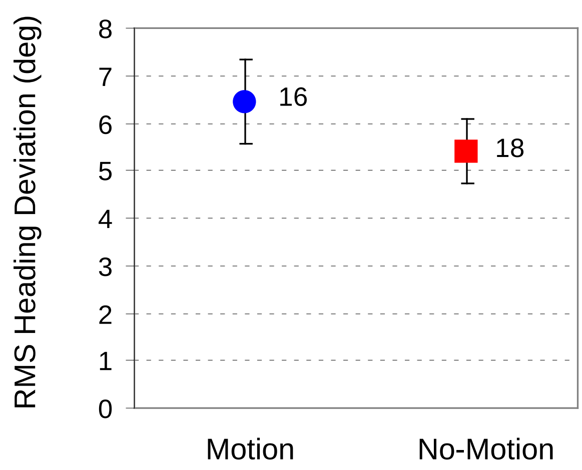




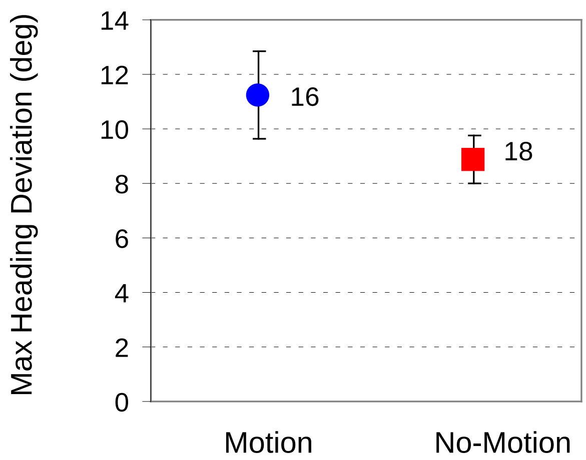


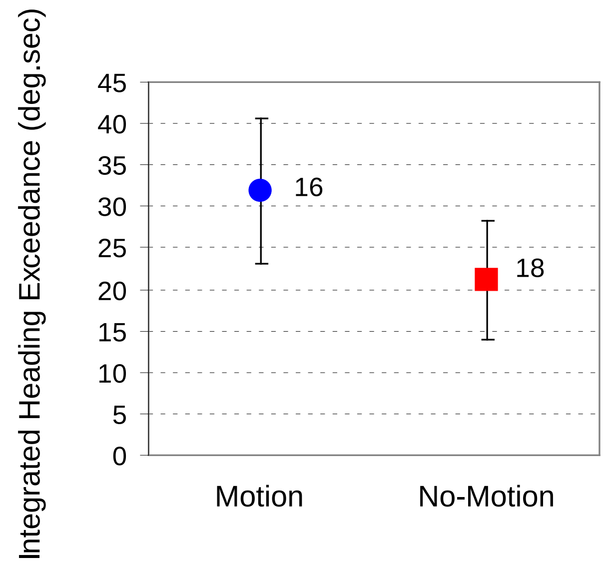


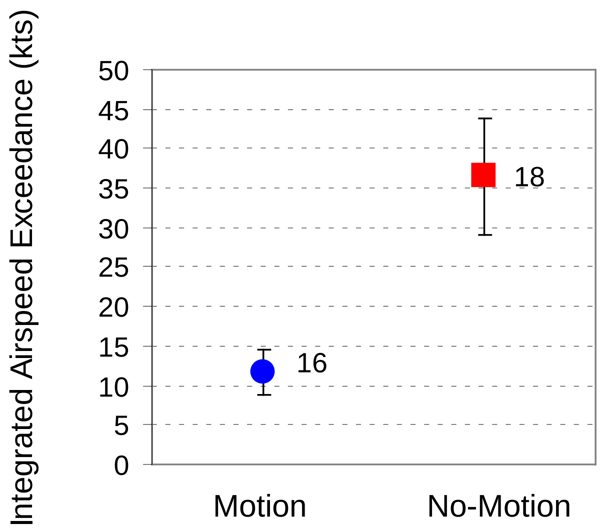


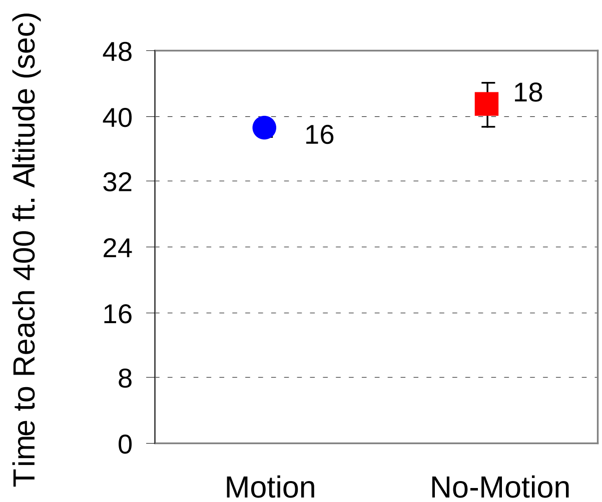


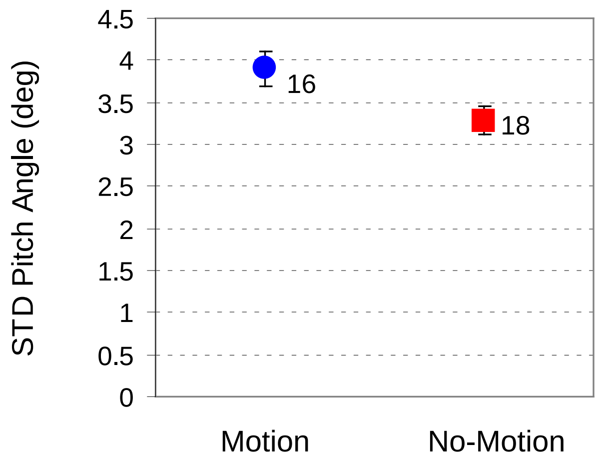
V1 Cut Transfer
(Note: *Sample sizes are shown next to each data point*)

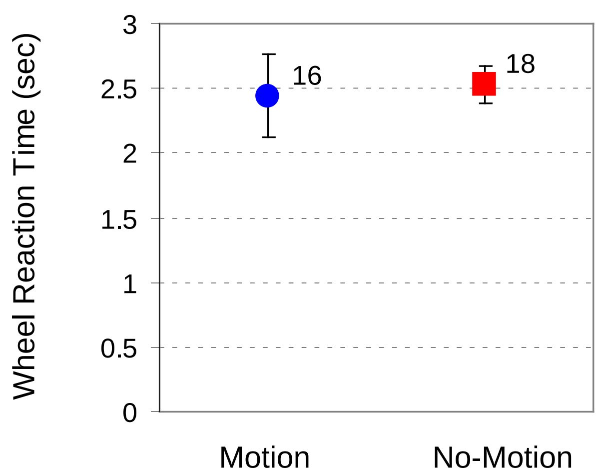




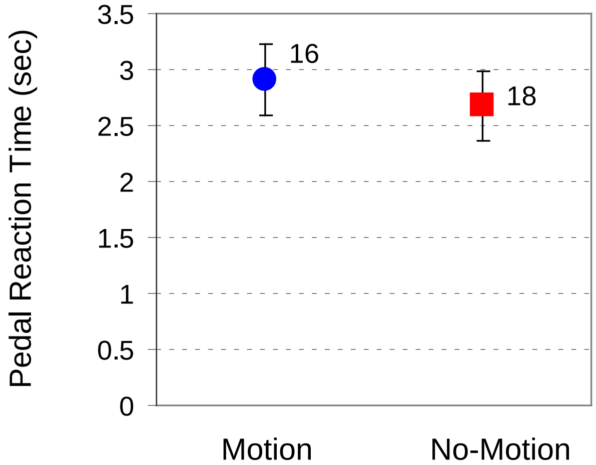


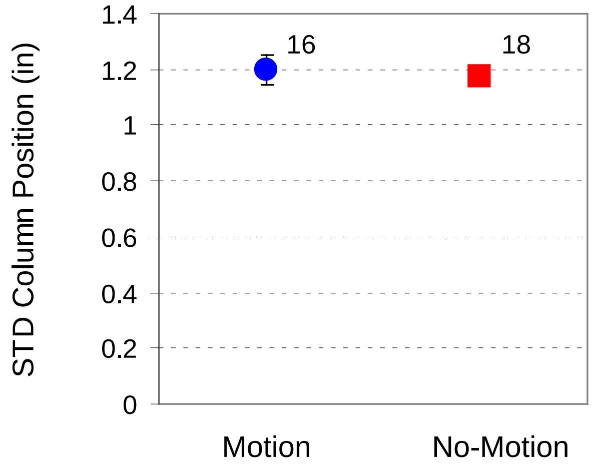


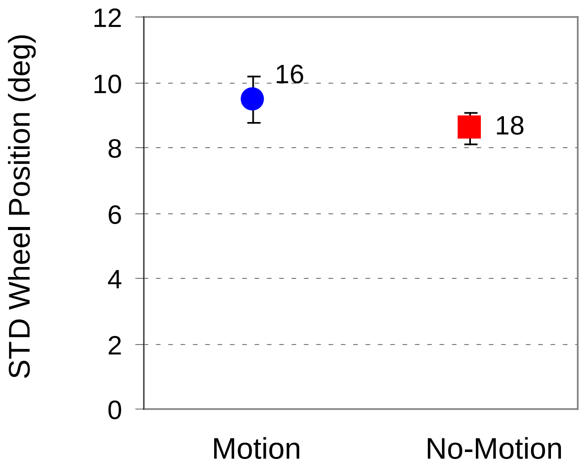


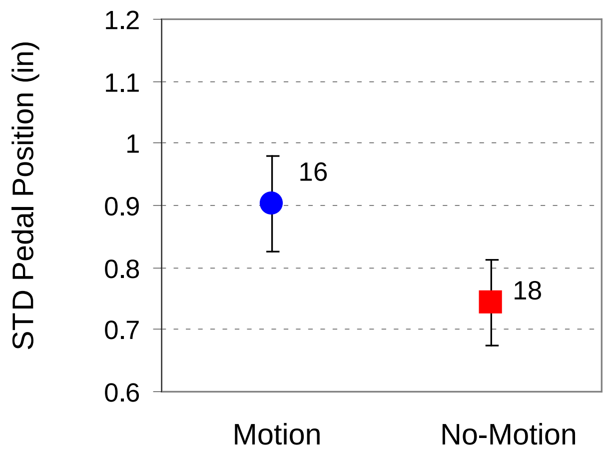


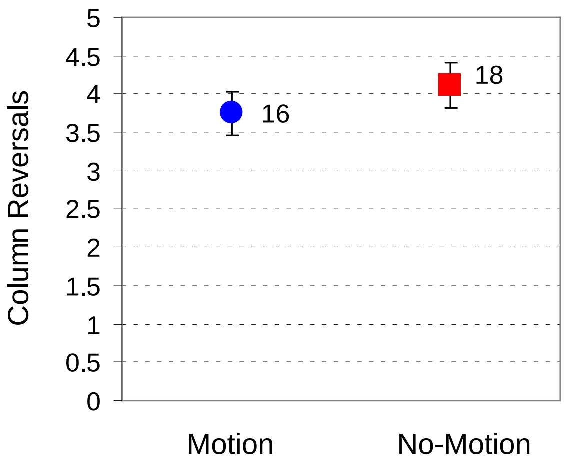
V1 Cut Transfer
(Note: *Sample sizes are shown next to each data point*)

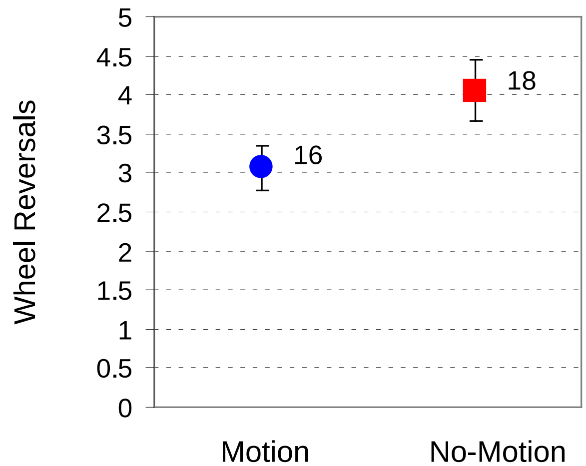




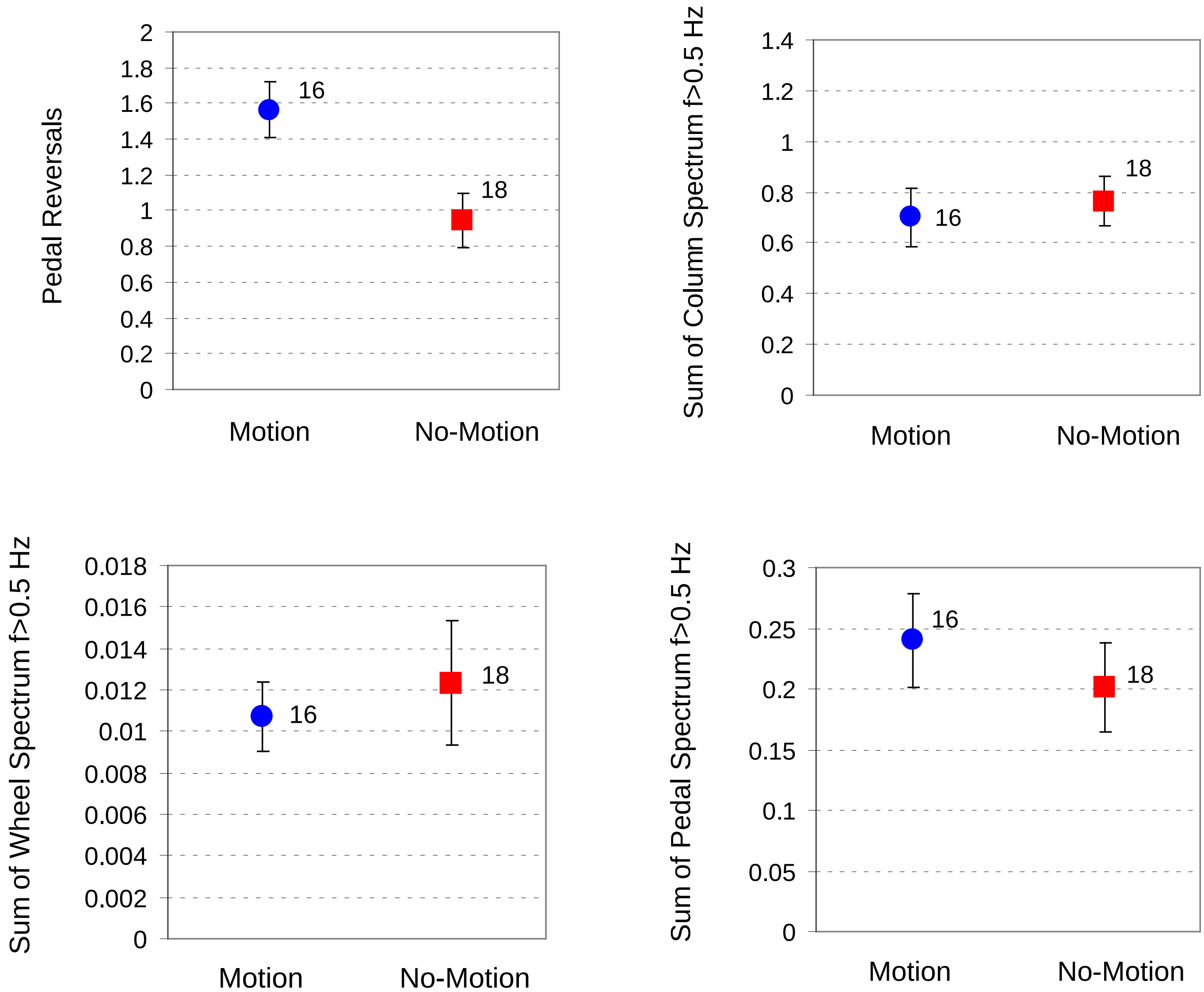








V1 Cut Transfer
(Note: *Sample sizes are shown next to each data point*)



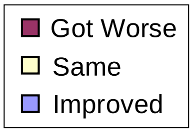
APPENDIX R – RTO FIRST VS LAST TRAINING COMPARISON

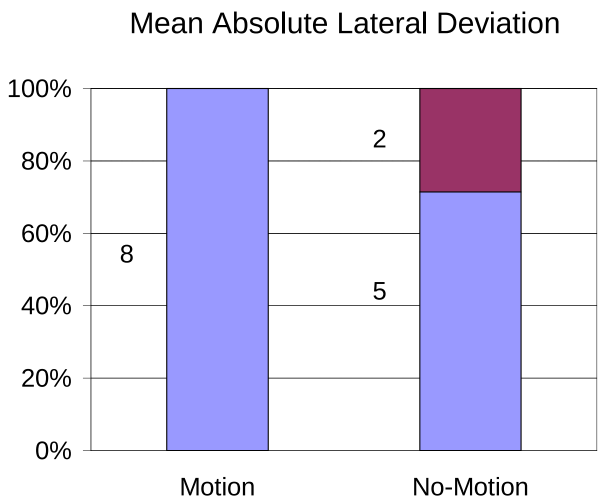
	GROUP DIFFERENCES IN IMPROVEMENT	OVERALL IMPROVEMENT		MOTION GROUP IMPROVEMENT		NO-MOTION GROUP IMPROVEMENT	
Variable Name	N = 15, Fisher Exact p	χ^2 (1, N = 15)	p	χ^2 (1, N = 8)	p	χ^2 (1, N = 7)	p
Mean Abs Lateral Deviation	0.2	8.067	0.005	8	0.005	1.286	0.257
RMS Lateral Deviation	0.467	11.267	0.001	8	0.005	3.571	0.059
Integrated Yaw Activity	0.622	0.6	0.439	0.5	0.480	0.143	0.706
RMS Heading Deviation	0.467	11.267	0.001	8	0.005	3.571	0.059
Max Heading Deviation	0.2	8.067	0.005	8	0.005	1.286	0.257
Power Lever Reaction Time	0.369	5.4	0.020	4.5	0.034	1.286	0.257
Pedal Reaction Time	0.294	0.286	0.593	1.286	0.257	0.143	0.705
STD Pedal Position	0.431	5.4	0.020	2	0.157	3.571	0.059
Pedal Reversals	0.662	3.267	0.071[31]	2	0.157	1.286	0.257
Sum of Pedal Spectrum f>0.5 Hz	0.1	1.667	0.197	4.5	0.034	0.143	0.705

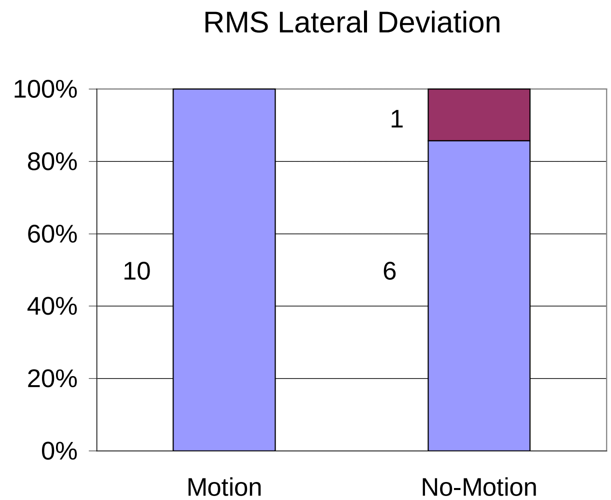
Table R-1. RTO First vs. Last Training Statistics

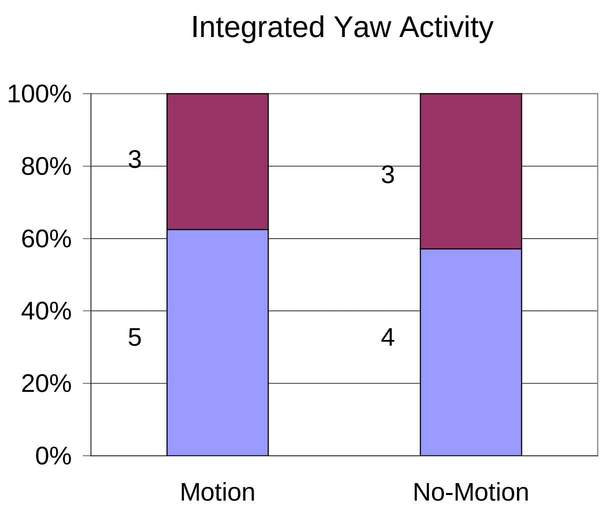
[31] Got worse

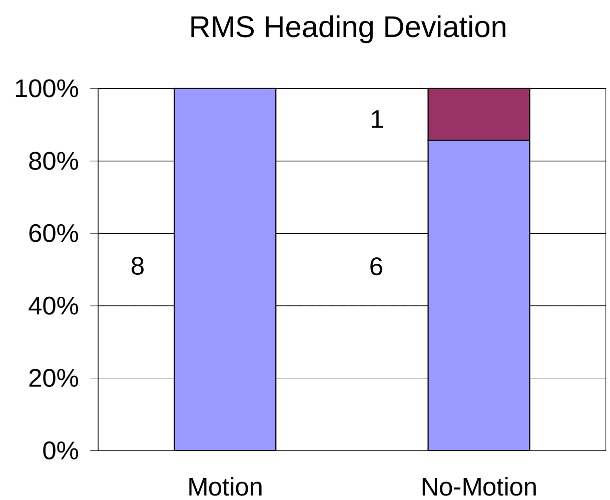
RTO – First vs. Last Training
(Note: *Numbers next to each bar indicate numbers of crews.*)

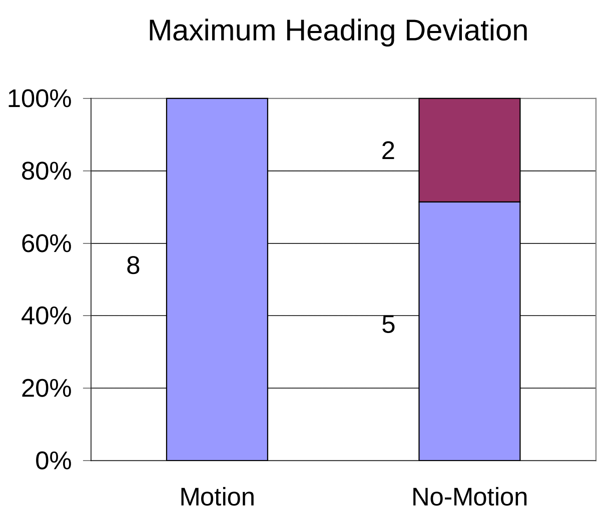


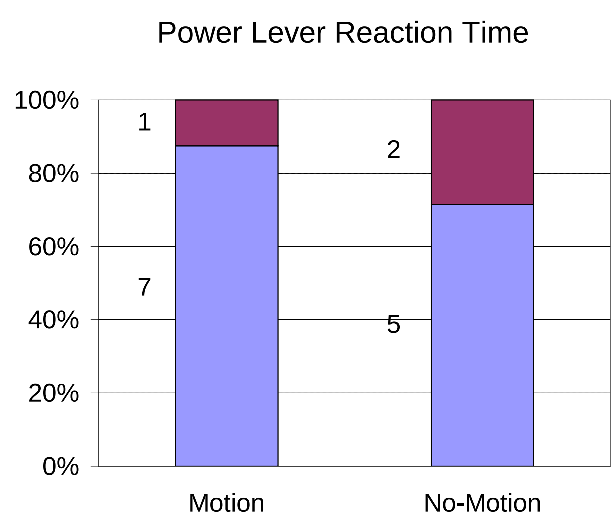




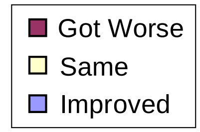


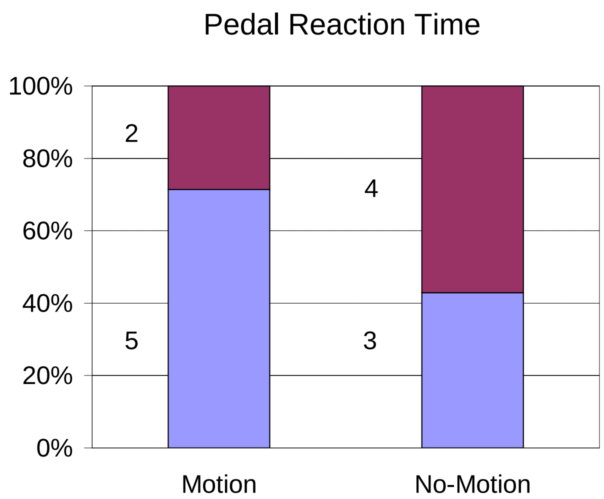


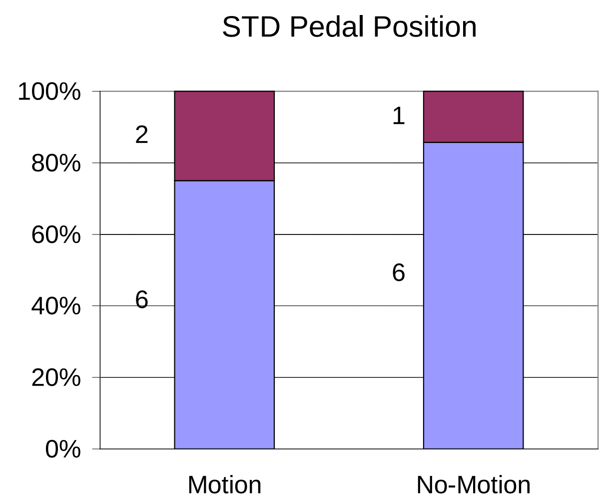


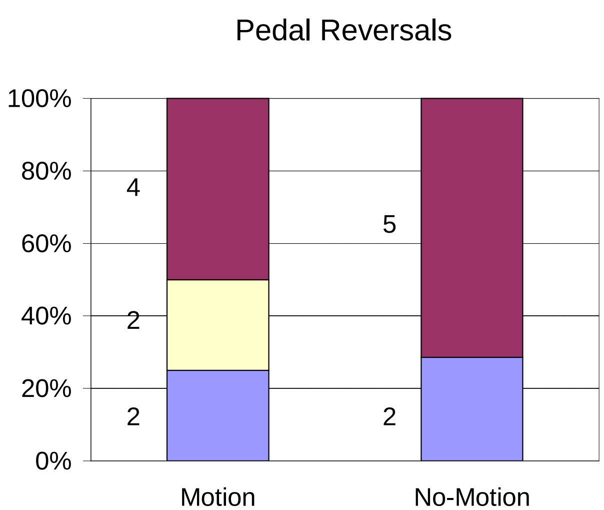


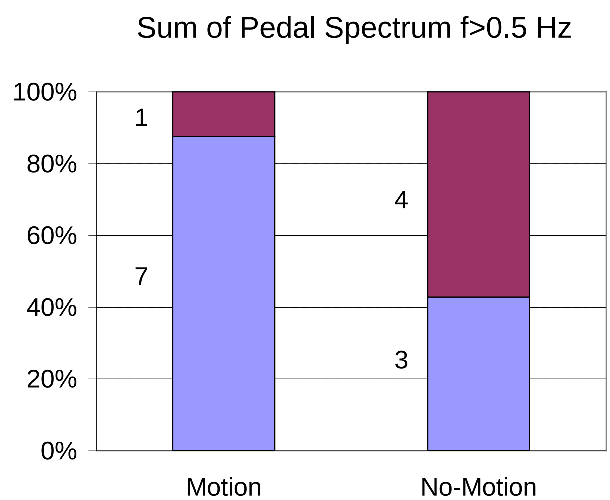
RTO – First vs. Last Training
(Note: *Numbers next to each bar indicate numbers of crews*.)











APPENDIX S - V_1 CUT FIRST VS LAST TRAINING COMPARISON

	GROUP DIFFERENCES IN IMPROVEMENT	OVERALL IMPROVEMENT		MOTION GROUP IMPROVEMENT		NO-MOTION GROUP IMPROVEMENT	
Variable Name	N = 29, Fisher Exact *p*	χ^2 (1, N = 29)	*p*	χ^2 (1, N = 16)	*p*	χ^2 (1, N = 13)	*p*
Max Abs Bank Angle	0.434	0.034	0.853	0.25	0.617	0.077	0.782
Integrated Bank Angle Exceedance	0.135	1.690	0.194	0	1	3.769	0.052
Integrated Failure Induced Bank Angle	0.18	0.034	0.853	1	0.317	0.692	0.405
Integrated Roll Activity	0.566	0.034	0.853	0	1	0.077	0.782
Integrated Yaw Activity	0.507	2.793	0.095[32]	1	0.317	1.923	0.166
RMS Heading Deviation	0.018	0.034	0.853	2.25	0.134	3.769	0.052
Maximum Heading Deviation	0.076	0.862	0.353	0.25	0.617	3.769	0.052
Integrated Heading Exceedance	0.018	0.034	0.853	2.25	0.134	3.769	0.052
Integrated Airspeed Exceedance	0.023	1.690	0.194	6.25	0.012	0.692	0.405
Time to Reach 400 ft. Altitude	0.282	0.034	0.853	0.25	0.617	0.692	0.405
STD Pitch Angle	0.089	4.172	0.041	0.25	0.617	6.231	0.013
Wheel Reaction Time	0.18	0.034	0.853	1	0.317	0.692	0.405
Pedal Reaction Time	0.253	0.862	0.353	0	1	1.923	0.166

Table S-1. V_1 Cut First vs. Last Training Statistics

[32] Got worse

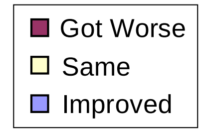
V_1 Cut: First vs. Last Training Statistics (cont.)

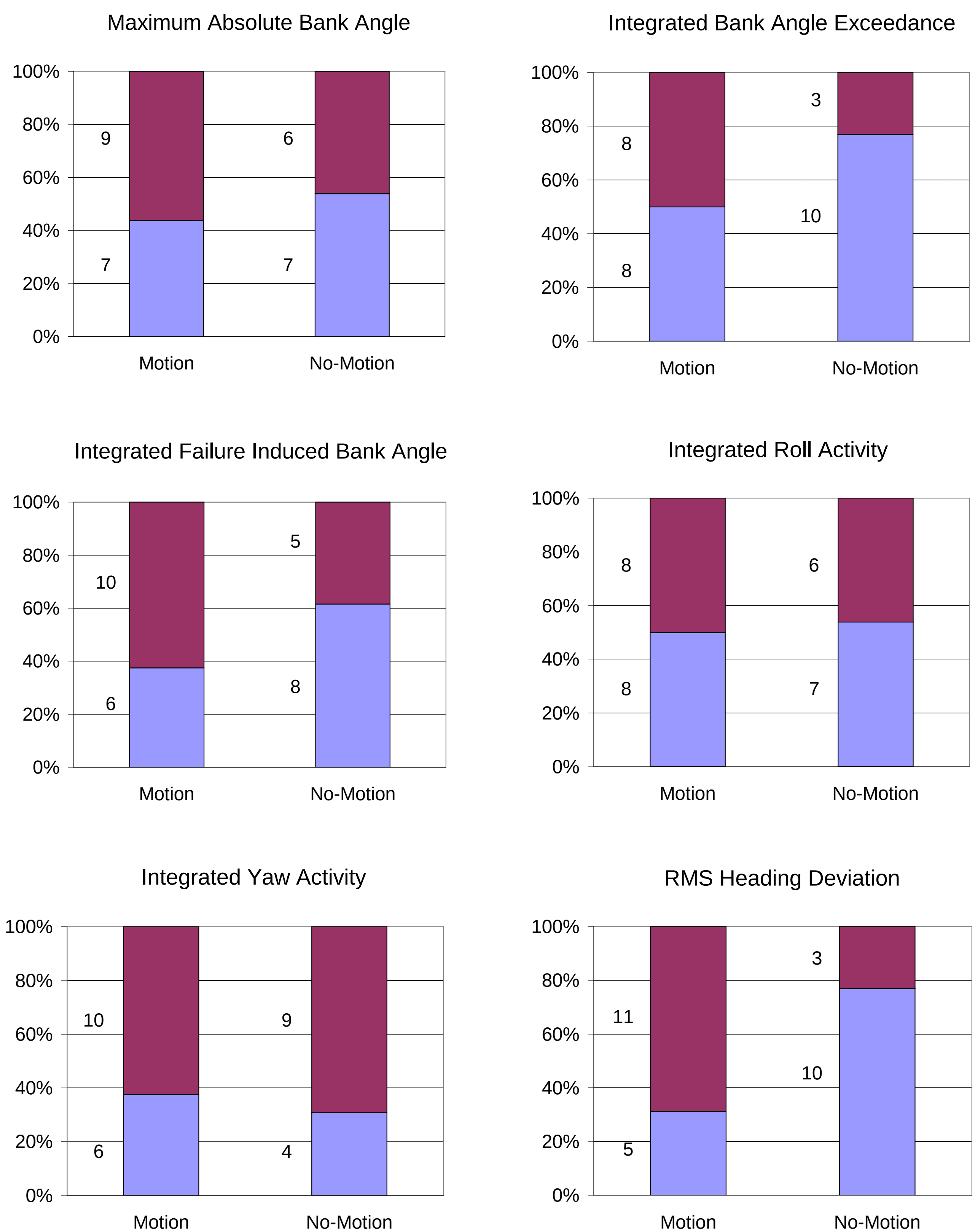
	GROUP DIFFERENCES IN IMPROVEMENT	**OVERALL IMPROVEMENT**		**MOTION GROUP IMPROVEMENT**		**NO-MOTION GROUP IMPROVEMENT**	
Variable Name	N = 29, Fisher Exact *p*	χ^2 (1, *N* = 29)	*p*	χ^2 (1, *N* = 16)	*p*	χ^2 (1, *N* = 13)	*p*
STD Column Position	0.526	5.828	0.016	4	0.046	1.923	0.166
STD Wheel Position	0.566	0.034	0.853	0	1	0.077	0.782
STD Pedal Position	0.566	0.034	0.853	0	1	0.077	0.782
Column Reversals	0.074	7.759	0.005[34]	1	0.317	9.308	0.002[34]
Wheel Reversals	0.322	7.759	0.005[33]	2.25	0.134	6.231	0.013[34]
Pedal Reversals	0.58	18.241	0.001[34]	9	0.003[34]	9.308	0.002[34]
Sum of Column Spectrum f>0.5 Hz	0.596	0.310	0.577	0.25	0.617	0.077	0.782
Sum of Wheel Spectrum f>0.5 Hz	0.307	0.310	0.577	1	0.317	0.077	0.782
Sum of Pedal Spectrum f>0.5 Hz	0.330	1.690	0.194	2.25	0.134	0.077	0.782

Table S-1 cont. V_1 Cut First vs. Last Training Statistics

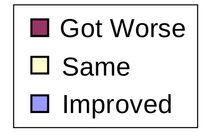
[33] Got worse/stayed the same

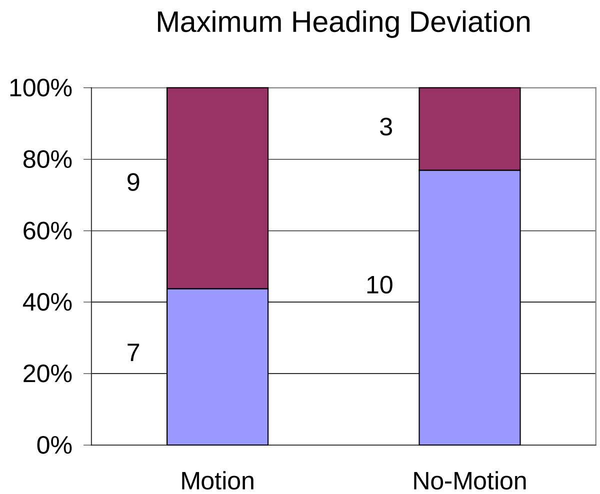
V_1 Cut – First vs. Last Training
(Note: *Numbers next to each bar indicate numbers of crews*.)

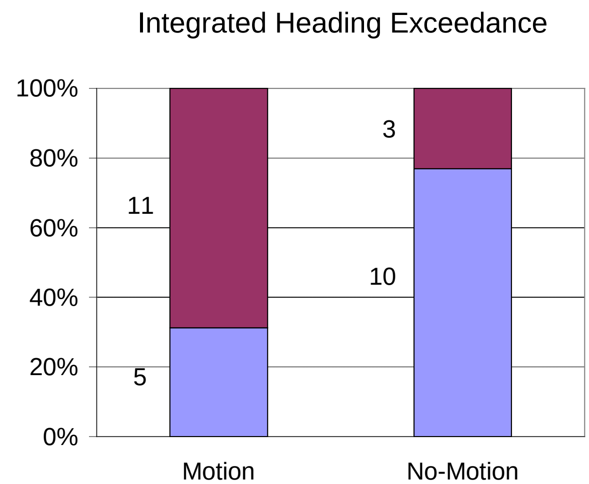


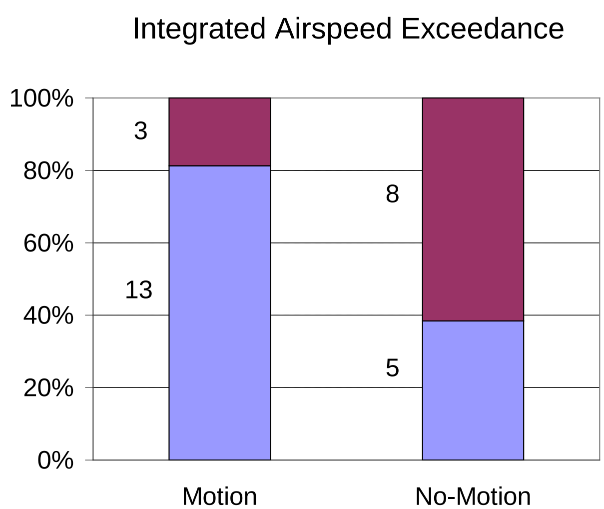


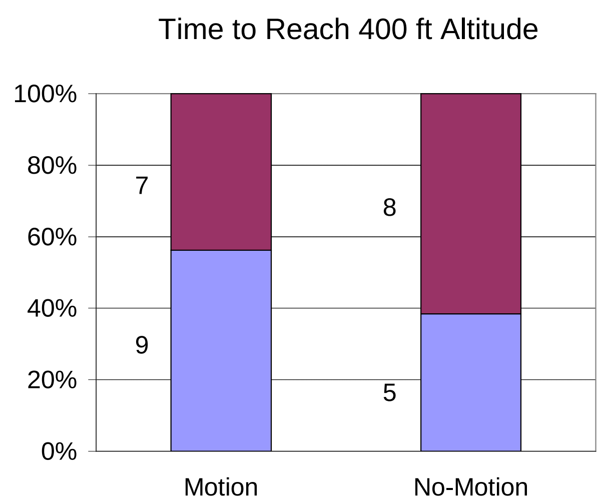
V_1 Cut – First vs. Last Training
(Note: *Numbers next to each bar indicate numbers of crews*.)

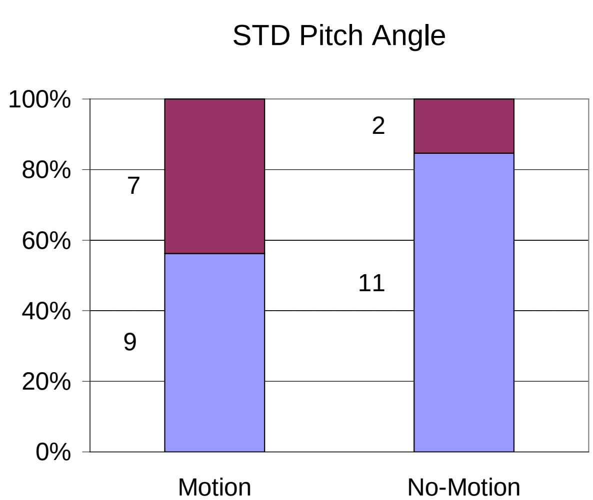


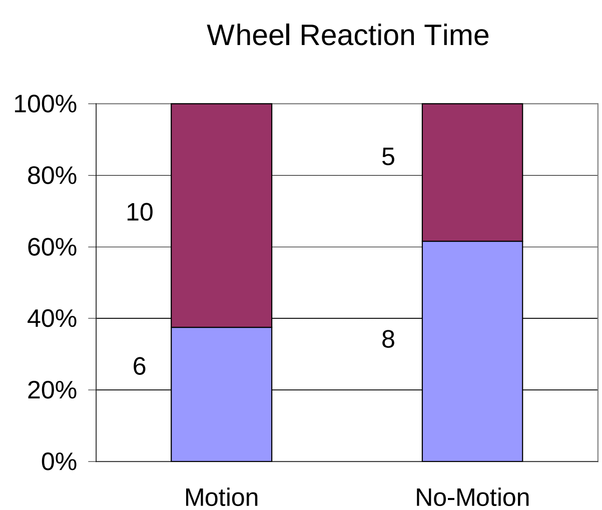




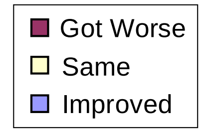


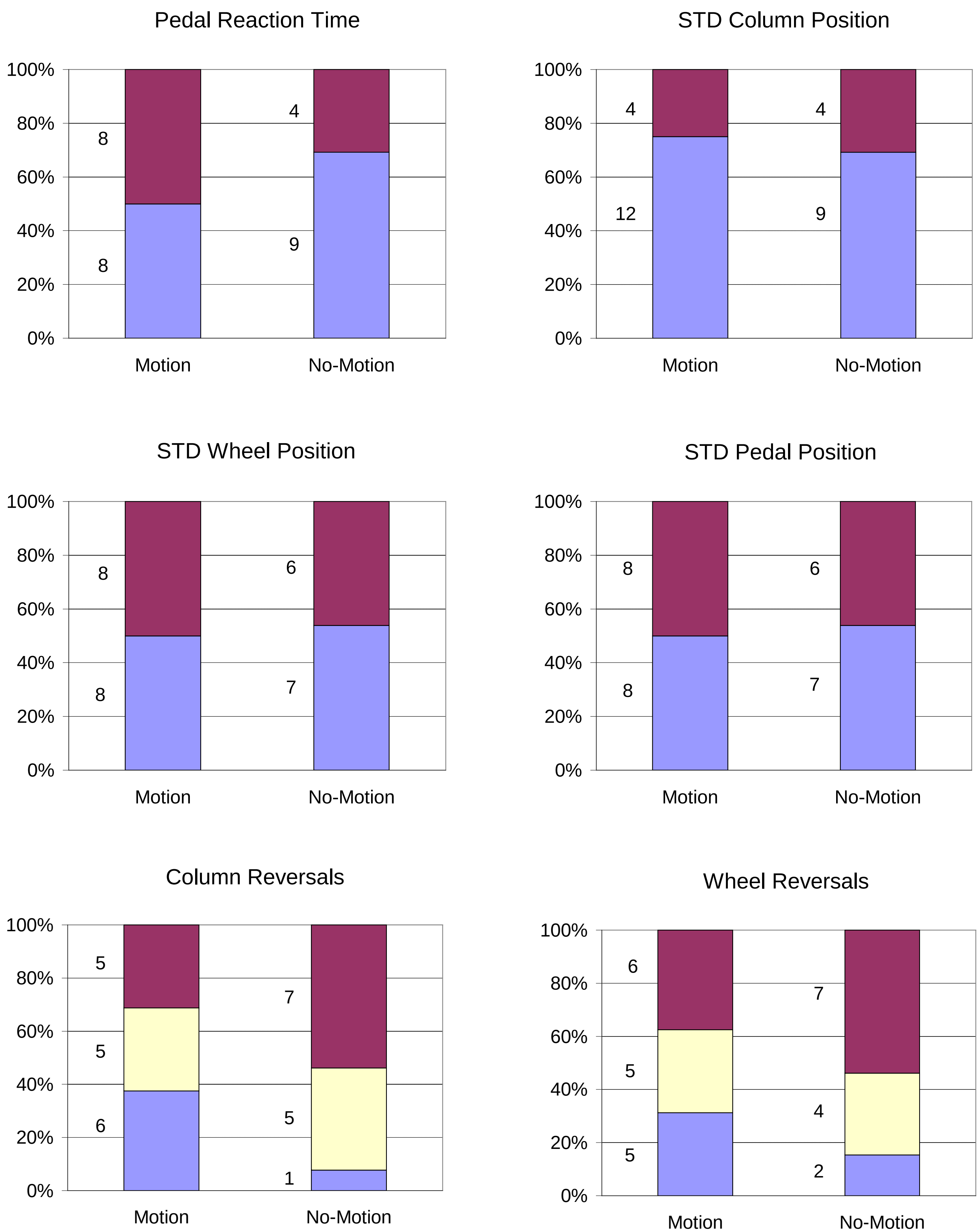




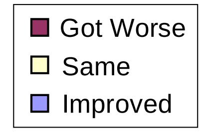


V_1 Cut – First vs. Last Training
(Note: *Numbers next to each bar indicate numbers of crews*.)





V_1 Cut – First vs. Last Training
(Note: *Numbers next to each bar indicate numbers of crews*.)

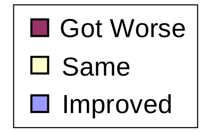


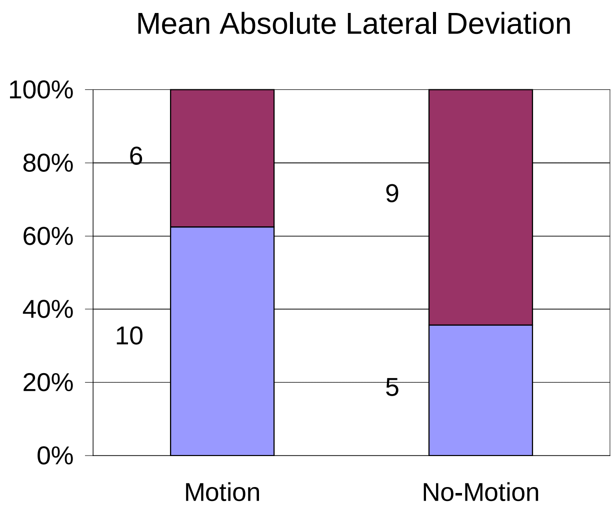
APPENDIX T – RTO LAST TRAINING VS TRANSFER COMPARISON

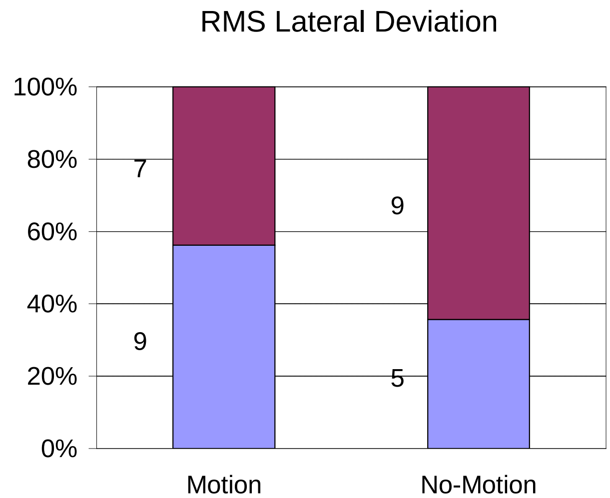
	GROUP DIFFERENCES IN IMPROVEMENT
Variable Name	N= 30, Fisher Exact *p*
Mean Abs Lateral Deviation	0.136
RMS Lateral Deviation	0.225
Integrated Yaw Activity	0.014
RMS Heading Deviation	0.374
Max Heading Deviation	0.033
Power Lever Reaction Time	0.374
Pedal Reaction Time	0.610
STD Pedal Position	0.206
Pedal Reversals	0.550
Sum of Pedal Spectrum f>0.5 Hz	0.206

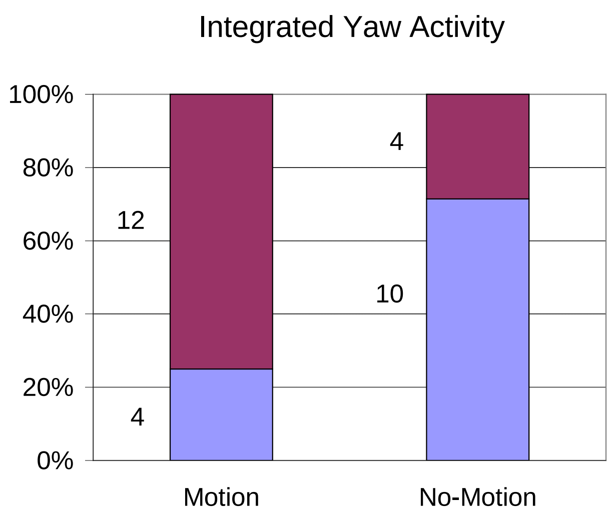
Table T-1. RTO Last Training vs. Transfer Statistics

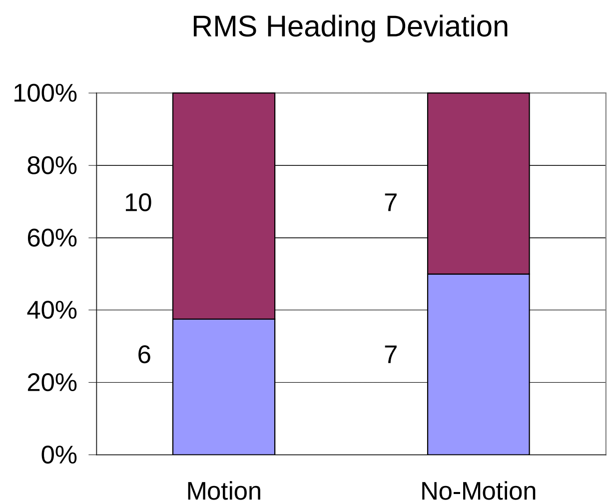
RTO – Last Training vs. Transfer
(Note: *Numbers next to each bar indicate numbers of crews*.)

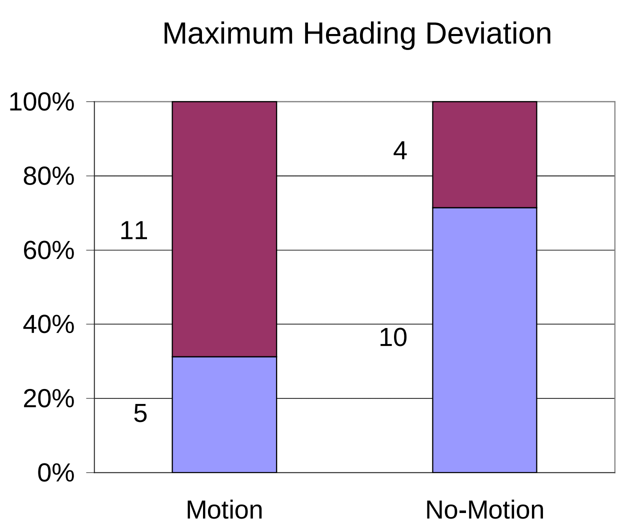


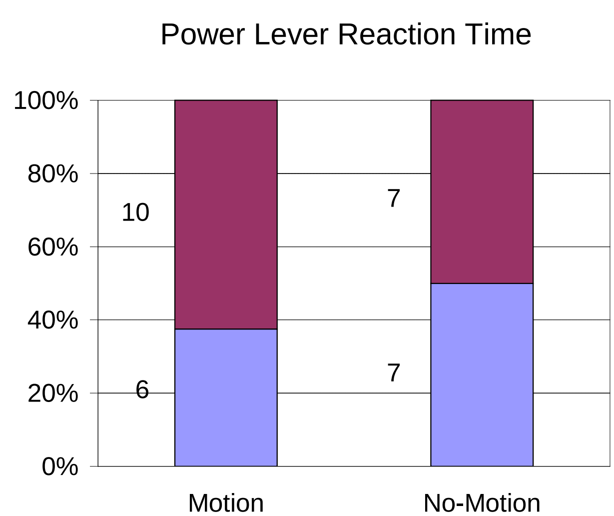












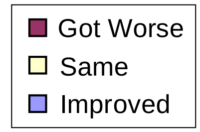
RTO – Last Training vs. Transfer
(Note: *Numbers next to each bar indicate numbers of crews*.)

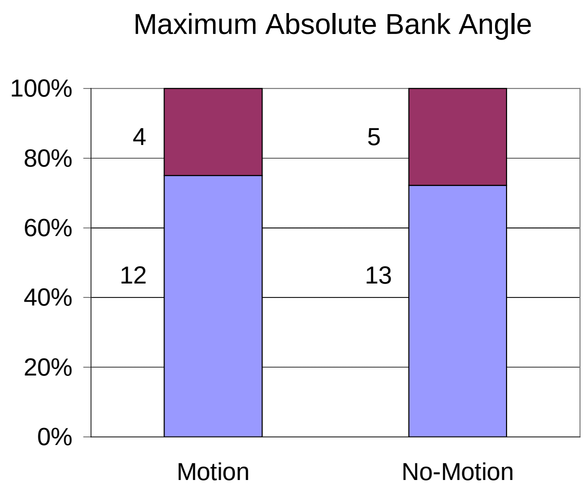
APPENDIX U - V_1 CUT LAST TRAINING VS TRANSFER COMPARISON

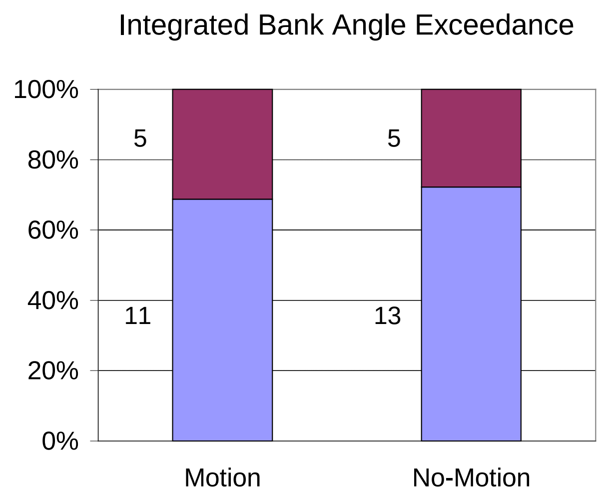
	GROUP DIFFERENCES IN IMPROVEMENT
Variable Name	*N* = 34, Fisher Exact *p*
Max Abs Bank Angle	0.583
Integrated Bank Angle Exceedance	0.56
Integrated Failure Induced Bank Angle	0.583
Integrated Roll Activity	0.541
Integrated Yaw Activity	0.02
RMS Heading Deviation	0.332
Maximum Heading Deviation	0.393
Integrated Heading Exceedance	0.332
Integrated Airspeed Exceedance	0.164
Time to Reach 400 ft Altitude	0.508
STD Pitch Angle	0.652
Wheel Reaction Time	0.087
Pedal Reaction Time	0.311
STD Column Position	0.126
STD Wheel Position	0.541
STD Pedal Position	0.017
Column Reversals	0.262
Wheel Reversals	0.594
Pedal Reversals	0.285
Sum of Column Spectrum f>0.5 Hz	0.020
Sum of Wheel Spectrum f>0.5 Hz	0.459
Sum of Pedal Spectrum f>0.5 Hz	0.560

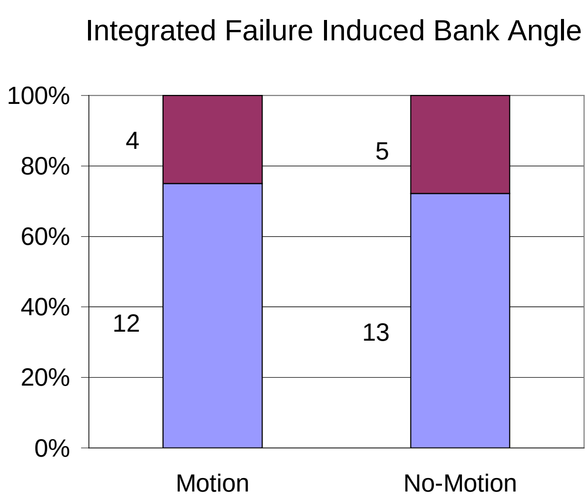
Table U-1. V_1 Cut Last Training vs. Transfer Statistics

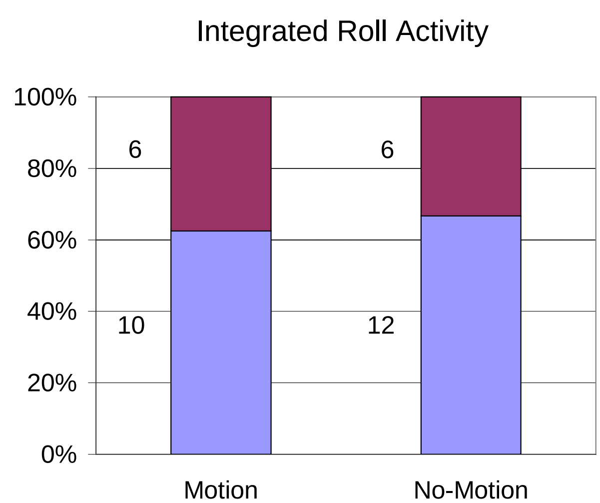
V_1 Cut – Last Training vs. Transfer
(Note: *Numbers next to each bar indicate numbers of crews.*)

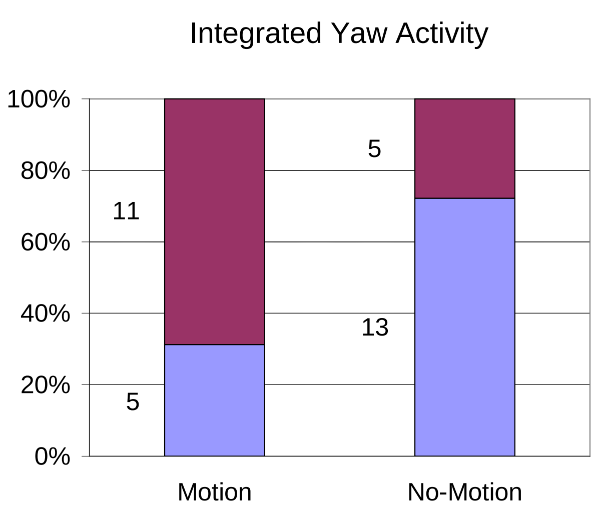


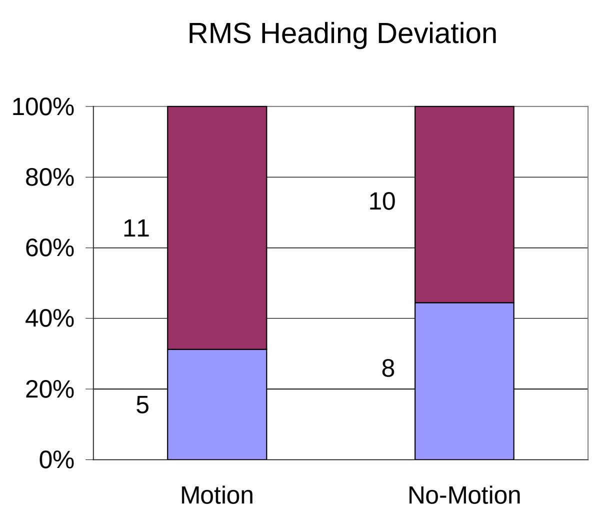




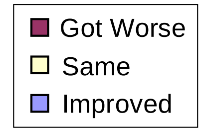


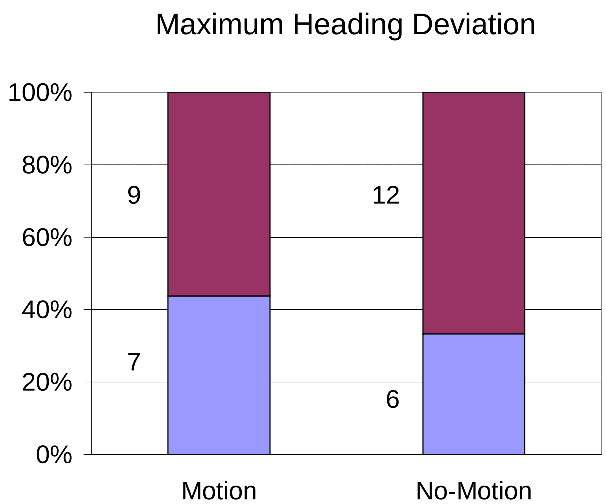


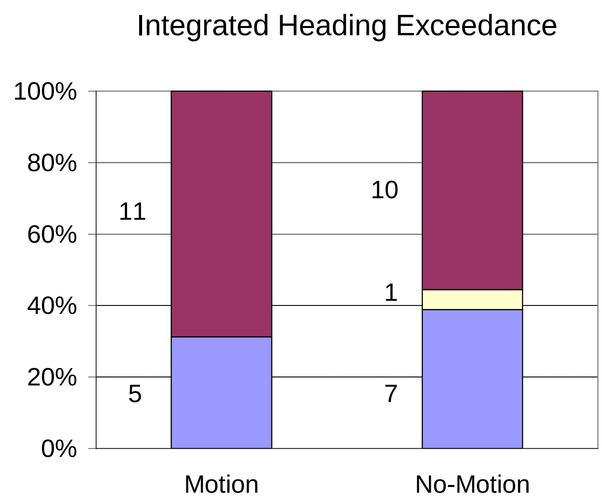


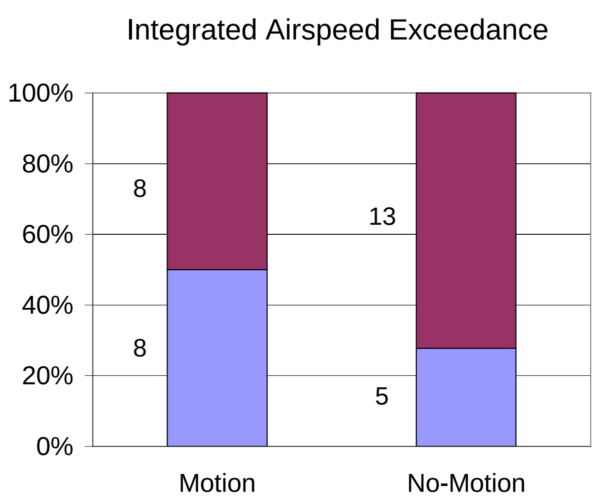


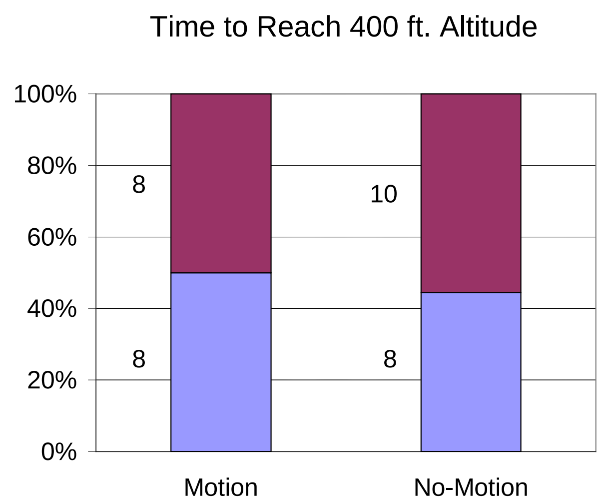
V$_1$ Cut – Last Training vs. Transfer
(Note: *Numbers next to each bar indicate numbers of crews*.)

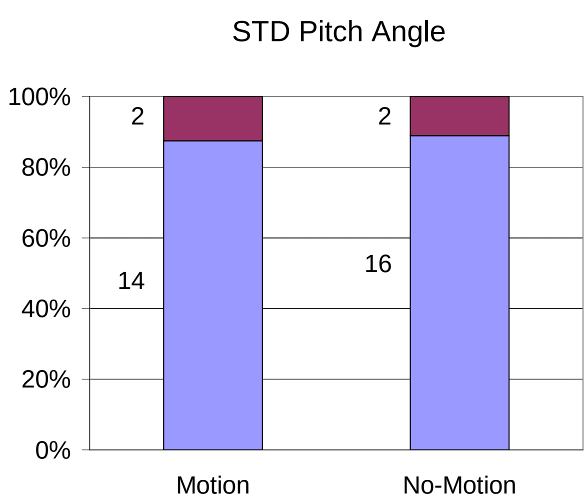


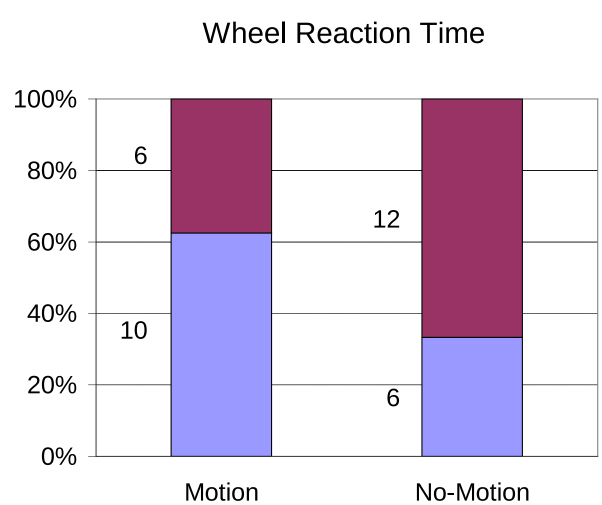




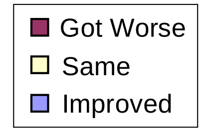


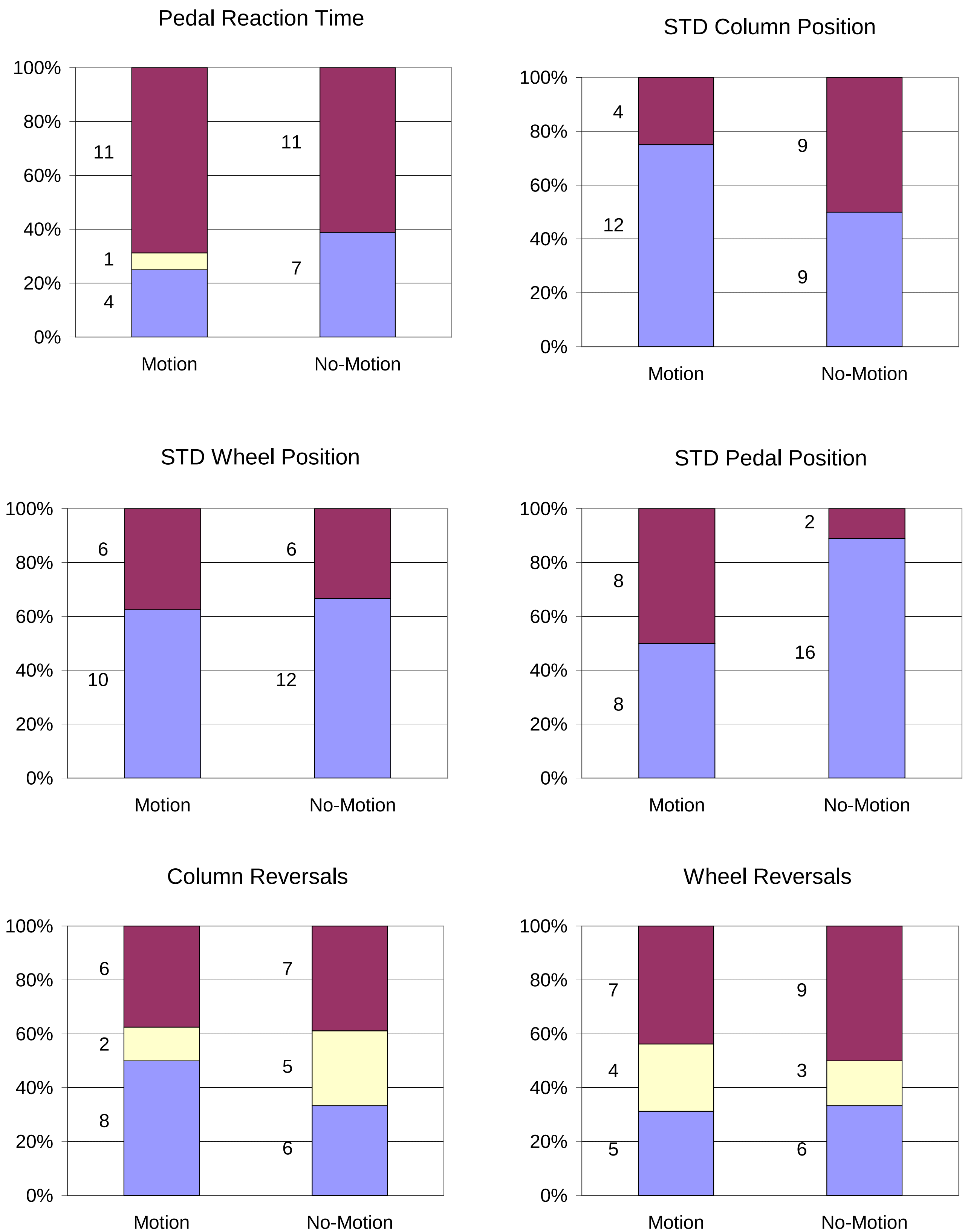






V_1 Cut – Last Training vs. Transfer
(Note: *Numbers next to each bar indicate numbers of crews*.)





V_1 Cut – Last Training vs. Transfer
(Note: *Numbers next to each bar indicate numbers of crews*.)