

# Transfer of Training from a Full-Flight Simulator vs. a High Level Flight Training Device with a Dynamic Seat

Andrea L. Sparko\* and Judith Bürki-Cohen†  
*USDOT/RITA/Volpe National Transportation Systems Center, Cambridge, MA 02142*

Tiau H. Go‡  
*Nanyang Technological University, Singapore 639798*

This paper summarizes the most recent study conducted by the Federal Administration Administration/Volpe Center Flight Simulator Fidelity Requirements Program. For many smaller airlines, access to qualified simulators is limited due to the availability of simulators for certain airplanes and the costs of equipment acquisition, leasing, personnel travel, operation, and maintenance. The FAA/Volpe Center Flight Simulator Fidelity Requirements Program has endeavored to address this situation for more than a decade, first examining the most costly aspects of flight simulation in subject-matter-expert workshops and then conducting a series of empirical investigations of the effect of simulator hexapod-platform motion on training effectiveness. This paper is the sequel to our 2007 AIAA Modeling-and-Simulation-Technologies Conference paper. In the earlier paper, we provided the scientific, technical, and operational background behind innovative solutions to provide motion cues in the simulator during airline-pilot training. We summarized three previous studies by the FAA/Volpe Center investigating the effect of hexapod-platform motion cues on training and evaluation of airline pilots in Full Flight Simulators (FFS). This research did not find operationally relevant differences in performance or behavior of pilots tested in the FFS with motion after having been trained in the same FFS with the motion system turned on or off – despite selection of maneuvers that require motion cues, at least theoretically. It made no difference whether the FFS represented a small turboprop “power house” or a sluggish four-engine jumbo jet, or whether the training in question was initial or recurrent training. Our 2007 paper also described a newly developed simulator, the Full-Flight Trainer FFT-X™ (FFT), able to simulate motion without a hexapod-motion platform. The paper concluded by reporting a proof-of-concept study culminating in the successful type rating of six pilots on a twin-engine turboprop after training in the FFT only. The present paper reports the results of our successor study, comparing the training effectiveness of the FFT, the “motion-cueing simulator without a motion base,” with its FFS equivalent. Not only does this study differ from the earlier studies by comparing FFS motion with an alternative method of motion cueing, but also by including pilots with minimal prior flight experience of fewer than 500 hours. Pilots were divided into two groups to be trained either in the FFS or the FFT on maneuvers determined, from the literature and our previous studies, to be most likely to require platform-motion cues, namely, engine-failures on takeoff and hand-flown engine-out landings with severe weather. Both groups were then tested in the same maneuvers in the FFS with the full-motion platform as a stand-in for the airplane. Results are presented on pilot control-input behavior and flight precision recorded from the simulator. Pilot and instructor opinions regarding the simulator and/or pilot behavior and performance, both after training and after testing, are also reported.

---

<sup>1</sup> Engineering Psychologist, Human Factors Research and System Applications Center of Innovation, RVT-81, 55 Broadway. Member AIAA.

<sup>2</sup> Engineering Psychologist, Human Factors Research and System Applications Center of Innovation, RVT-81, 55 Broadway. Senior Member AIAA.

<sup>3</sup> Assistant Professor, School of Mechanical and Aerospace Engineering, Senior Member AIAA.

## Nomenclature

<i>A.F.C.S</i>	=	Automatic Flight Control System
<i>AGL</i>	=	Above Ground Level
<i>ANOVA</i>	=	Analysis of Variance
<i>C.I.</i>	=	Confidence Interval
<i>DMRM</i>	=	Doubly Multivariate Repeated Measures
<i>EGPWS</i>	=	Enhanced Ground Proximity Warning System
<i>ETOPS</i>	=	Extended-range Twin-engine Operations
<i>FAA</i>	=	United States Federal Aviation Administration
<i>FFS</i>	=	Full Flight Simulator
<i>FFT</i>	=	Full-Flight Trainer FFT-X™
<i>FOV</i>	=	Field-of-View
<i>GNSS</i>	=	Global Navigation Satellite System
<i>GPS</i>	=	Global Positioning System
° <i>H</i>	=	Degrees horizontal
<i>IAS</i>	=	Indicated Airspeed
<i>ILS</i>	=	Instrument Landing System
<i>JAR</i>	=	Joint Aviation Requirement
<i>MANOVA</i>	=	Multivariate Analysis of Variance
<i>MSL</i>	=	Mean Sea Level
<i>NAA</i>	=	National Aviation Authorities
<i>nm</i>	=	Nautical Miles
<i>PIA</i>	=	Precision Instrument Approach
<i>RVR</i>	=	Runway Visible Range
<i>STD</i>	=	Standard Deviation
<i>SSL</i>	=	Sidestep Landing
<i>TCAS</i>	=	Traffic Collision Avoidance System
° <i>V</i>	=	Degrees vertical
$V_1$	=	Critical engine failure recognition speed
$V_2$	=	The speed at which a plane can safely take off with one inoperable engine
$V_{mca}$	=	Minimal Controllable Airborne speed

## I. Introduction

THIS paper summarizes the latest study of the Federal Administration Administration/Volpe Center Flight Simulator Fidelity Requirements Program. For many smaller airlines, access to qualified simulators is limited due to the availability of simulators for certain airplanes and the costs of equipment acquisition, leasing, personnel travel, operation, or maintenance. The FAA/Volpe Center Flight Simulator Fidelity Requirements Program has endeavored to address this situation for over a decade, by first examining the most costly aspects of flight simulation in subject-matter-expert workshops and then conducting a series of empirical investigations of the effect of simulator hexapod-platform motion on training effectiveness.

The current study builds on the work presented in our 2007 AIAA Modeling-and-Simulation-Technologies Conference paper.<sup>1</sup> That paper described a newly-developed simulator, the Full-Flight Trainer FFT-X™ (FFT), which offers an alternative to the hexapod-motion systems that are standard in Full-Flight Simulators (FFS) used for airline-pilot training. The FFT simulates motion via a high-level visual system and a dynamic seat with heave-motion and vibration cues only. At the time the 2007 paper was published, the FFT had been successfully used in the type-rating of six pilots – a world premiere. In the paper, we described a “proof-of-concept” study in which we collected opinions from these pilots and their instructors, both after training in the FFT and after transferring to the airplane. Opinion data revealed that the pilots’ experience in the FFT was perceived to be similar or equivalent to the airplane. After flying the airplane, trainees agreed that the simulator cues (visual, sound, and motion) were equivalent to the cues experienced in the airplane. During a debriefing of participants after the first phase, trainees and instructors agreed that the transition from the FFT to the airplane was seamless.

With the success of this first phase of the FFT type-rating program, we proceeded to perform a formal evaluation of the FFT. The current paper focuses on this formal evaluation, which divided the pilots into two groups: an FFS- and an FFT-trained group. Both groups first received training on the type-rating-program maneuvers in the

FFT-1, which has no physical motion system. Pilots were then trained in their assigned simulators, either the FFS or the FFT with dynamic-seat motion, in maneuvers that had been determined, from the literature and our previous studies, to be most likely to require motion. All pilots were then quasi-transferred to the FFS as a stand-in for the airplane, before flying the actual airplane to obtain credit. Extensive measures of pilot-vehicle performance were recorded from the simulator. Pilots and instructors also filled out detailed questionnaires, rating themselves, the pilot flying, and/or the simulator as appropriate.

The design of the formal evaluation phase was established to match that of our previous studies.<sup>2-4</sup> These studies trained two groups of pilots: one in the FFS with the motion on and the other in the same FFS with the motion turned off. All pilots were then transferred to the FFS with the motion on for testing. Across these three studies, we found no operationally relevant differences in the training and evaluation of pilots trained in the FFS with the motion on vs. with the motion off, with no subjective preferences for either motion configuration throughout. The maneuvers flown in these studies were carefully selected for their potential to require motion, and these studies obtained similar results with both initial and recurrent pilots and in FFS representing transport airplanes ranging from a small turboprop “power house” to a sluggish four-engine jumbo jet. The concept for the FFT was developed in view of the results of this work, with the expectation that the FFT would help curtail the potential reduction in training opportunities due to the pressing need for pilots and the high cost of qualified training tools. The advent of the FFT and its approval by the National Aviation Authorities (NAA) for type-rating training gave the FAA/Volpe Center Flight Simulator Human Factors Program the opportunity to build on our previous work with two fundamental differences: Not only does this study differ from the earlier studies by comparing the training-effectiveness of FFS motion with that of an alternative method of motion cueing, but also by including pilots with minimal prior flight experience of fewer than 500 hours. We are presenting the results of this evaluation in the current paper.

## II. Method

### A. Participants

#### 1. *Trainees*

A total of 92 pilots were successfully type-rated across the three phases of the program: 6 pilots in Phase 1 (“proof-of concept” phase) and 16 in Phase 2 (run-through-of-procedures phase, with limited data collection) (see Ref. 1 for overview of Phase 1), and 70 in Phase 3. Phase 3 was the official evaluation/study phase during which extensive data were collected. Of the 70 pilots in this final phase, data were collected from 49 low-experienced pilots (28 FFS, 22 FFT), 3 medium-experienced pilots (2 FFS, 1 FFT), and 17 high-experienced pilots (6 FFS, 11 FFT). All of the low-experienced trainees were training to be first officers and had no JAR-25 (Joint Aviation Requirements – Large Airplanes<sup>§§</sup>) experience and fewer than 500 total flight hours. The medium-experienced trainees were two captain-trainees with fewer than 2,000 hours JAR-25 experience and one first-officer trainee with no JAR-25 experience. The high-experienced trainees were all captains in training with more than 2,000 hours JAR-25 experience. Trainees were either private candidates or pilots sent from a regional airline.

For this paper, we restricted the analysis to data from the low-experienced pilots only. Due to restricted pilot and simulator availability, the experience levels of the pilots in the FFS- and FFT-trained groups were unbalanced; having more high-experienced pilots in one group vs. more low-experienced pilots in the other might influence performance, making a fair comparison of the two groups impossible. In total, data from 40 low-experienced pilots were included in the analysis (20 FFS-trained pilots and 20 FFT-trained pilots).<sup>\*\*</sup> All were flying in the right seat.

#### 2. *Instructors*

All instructors were employed by the training center and certified by the NAA. They were assigned to trainees based on the scheduling needs of the training center. Fifteen instructors participated in this phase of the study. Five “low-experienced” instructors had fewer than 500 total instruction flight hours in company airplanes. Four “medium-experienced” instructors had between 500 and 2,000 instruction flight hours; two with fewer than 500 hours in company airplanes, one with between 500 and 2,000 in company airplanes, and one with greater than 2,000 in company airplanes. Five “high-experienced” instructors had greater than 2,000 instruction flight hours and all but one of these instructors had greater than 2,000 hours company-airplane experience (one had fewer than 500).

---

<sup>§</sup> European equivalent of the U.S. FAA’s Federal Aviation Regulation (FAR) Part 25, Airworthiness Standards for Transport Category Airplanes

<sup>\*\*</sup> Technical problems (explained later) led to missing or eliminated data, so not all of the low-experienced trainees could be included in the analysis.

Data from 11 instructors were included in the analysis (only instructors who were matched with trainees whose data were analyzed). Four instructors had low experience and were matched with six FFS- and two FFT-trained pilots, four had medium experience and were each matched with two FFS- and five FFT-trained pilots, and four had high experience and were matched with 12 FFS- and 13 FFT-trained pilots.

## B. Equipment

The two simulators represented a turboprop airplane with two wing-mounted engines. The airplane is 27,166 meters long with a wingspan of 27,050 meters and has a zero fuel weight of 20,800 kg. It is intended to carry approximately 68 passengers and two flight-deck and two cabin crew.

The FFS was manufactured in 1998 by Thomson Training & Simulation; it is qualified annually. The FFT was manufactured in 2005 by Mechtronix. Both devices have a 180 x 40 deg Field-of-View (FOV) with a spherical collimated screen. The FFS motion provided mainly vestibular cues generated by a hydraulic six degrees-of-freedom hexapod platform on 60-inch legs. Specifications for both simulators are shown in Table 1.

**Table 1. FFS and FFT Specifications.**

	<b>FFS</b>	<b>FFT</b>
Autopilot/Automatic Flight Control System (AFCS)	Honeywell A.F.C.S. SPZ-6000	Honeywell A.F.C.S. SPZ-6000
Simulator model	FFS C98	ASCENT FFT (Full Flight Trainer)
Manufacturer	Thomson Training & Simulation	Mechtronix (Canada)
Date of manufacture	1998	2005
Visual System	Tropos R300	Tropos 6000 Medallion
Manufacturer	CAE	CAE
Projectors	3x BARCO	3x JVC HDRV 2K LCOS 1920x1080
Field-of-View	180°H x 40°V	180°H x 40°V
Screen	Spherical collimated with back projection screen	Spherical collimated with front projection screen
Motion System	Hydraulic, 6 degrees-of-freedom, 60-inch legs	Seat cueing system
Host Computer	Harris Night Hawk series NH5800	QNX PC with 3GHz INTEL Processor
Control Loading	Digital hydraulic system	Digital electric Fokker system
Crew	Airplane seats + 1 instructor + 2 observers	Airplane seats + 1 instructor
Autocoupled Approach	Yes	Yes
Traffic Collision Warning System (TCAS)	TCAS Change 7	TCAS Change 7
Weather Radar	Primus 800	Primus 800; 2 storms correlated; ground map not simulated
Enhanced Ground Proximity Warning System (EGPWS)	Honeywell Mark VII	Honeywell Mark VII with 3 enhanced scenarios simulated on LMML, LFKJ, LFKC
Extended-range Twin-engine Operations (ETOPS) Capability	Yes	Yes
Global Positioning System (GPS)/Global Navigation Satellite System (GNSS)	Yes (HT1000)	Not yet implemented (HT1000)
Smoke Generator	Yes	Yes

The FFT has a dynamic seat providing some vestibular, proprioceptive, and tactile motion stimulation from heave and vibration cues. Each pilot's seat is mounted on its own fixed platform which is moved vertically by actuators run by electrical rotary motors. A bass-speaker-shaker system underneath the cockpit provides additional vibration cues. The different specifications for the seat-cueing system, which are determined by the aerodynamic model of the airplane being simulated, are presented in Table 2. These specifications are based on the model's estimation of the airplane's combined translational and rotational accelerations. The airplane accelerations are corrected to account for differences in airplane vs. crew-seat center of gravity. Specific motion effects, provided by the manufacturer, include:

- Basic airplane accelerations (as described)
- Ground effects due to oleo deflections, groundspeed, runway roughness, and centerline lights
- Buffet when on the ground due to thrust reversal
- Thrust effect with brakes set
- Tail strike during takeoff
- Buffet in the air due to flap and gear extension
- Approach-to-stall buffet
- Buffet due to ice build-up
- High-speed buffet
- Engine vibration and engine stall malfunctions
- Turbulence effects
- Touchdown cues for main and nose gear, as a function of rate of descent
- Ground impact cues (e.g. during gear failures)
- Structural damage cues (e.g. during crash conditions)

**Table 2. Specifications of FFT seat-cueing system.**

Frequency (Hz)	Amplitude (in)	Velocity (in/s)	Acceleration (g)
2.5	0.800	10.00	1.33
5.0	0.521	10.00	1.33
10	0.130	8.18	1.33
15	0.058	5.45	1.33
20	0.033	4.09	1.33

### C. Maneuvers

The maneuvers trained and tested in the current study were the same maneuvers used in our previous studies. These maneuvers were carefully determined, based on literature reviews and expert opinion, to be maneuvers that critically depend on motion cues.

#### 1. Continued Takeoffs with Engine Failure

Two takeoff maneuvers were flown. In each, an engine failure occurred after  $V_1$  had been reached and the takeoff must be continued. The failure represented an engine flame-out with failure profile showing exponential loss of 94-99% of initial thrust in about one second. All takeoffs occurred at an altimeter altitude reading of 29.92 inches of mercury. There was a constant 10-knots tailwind. The runway visible range (RVR) was 600 meters with fog-top height of 500 ft. The takeoffs occurred during the day and the sky conditions were overcast with broken clouds at 3,000 feet. The engine-failure-triggering variables were varied to generate two maneuvers with different visual reference, motion stimulation, and workload as described below.

- “ $V_1$  cut”: For the  $V_1$  cut maneuver, an engine was failed when the simulator reached the minimal controllable airborne speed  $V_{mca}$  (104 knots). Compared to a  $V_2$  cut (see below), fewer axes need to be controlled when the wheels are still on the ground, but the asymmetry introduced by the engine failure can be larger than at higher speeds. Moreover, application of the rudder is less effective at a lower speed. Finally, fast action is critical to avoid tail or wing strikes.
- “ $V_2$  cut”: For the  $V_2$  cut, an engine failure was triggered after  $V_2$  (110 kts). Because the airplane was pitched up at this point in the takeoff envelope, pilots could no longer refer to the runway visually and had to fully rely on their instruments and motion perception, if available. Compared to earlier engine failures where the wheels are

still on the ground or the airplane is just rotating, pilots need to control pitch in addition to heading, resulting in high workload.

## 2. *Engine-out Landing Maneuvers*

Both landing maneuvers included an engine failure and were hand-flown without autopilot and flight director. Both require the pilot to tightly control the simulator in all three translational (sway, surge, heave) and rotational (pitch, roll, yaw) degrees of freedom. Under these circumstances, it was expected to be difficult, without motion feedback, to follow a narrow flight path and land softly in a tight box, especially with weather disturbances added to divert from the flight path.

- "ILS": For previous studies, the FAA's National Simulator Program manager recommended a hand-flown out-board engine-out precision instrument approach with shifting crosswind as particularly hard to fly without motion feedback. This maneuver consisted of landing the airplane guided by the instrument landing system (ILS, localizer and glide slope). The maneuver was flown in daylight and the visibility was kept low with 500 ft cloud ceiling and 2,500 meters RVR. A disturbance from a 13-14 knots terminal area crosswind shifting counterclockwise from a quartering headwind at 3,000 ft Mean Sea Level (MSL) to a quartering tailwind on the ground added to the challenges of this hand-flown approach and landing.
- "SSL": Ref. 5 used an offset approach followed by an S turn onto the runway at a very low altitude "to generate the kind of high-gain pilot behaviour which is necessary to bring out vehicle or simulator deficiencies." Sidestep-Landing (SSL) maneuvers with a vertical upward gust have a long history of use (e.g., see Refs. 3, 6, 7). In the current study, the pilot had to switch landing from runway 32 Right to the 300-meter-apart parallel runway 32 Left at the relatively low altitude of 1,000 ft AGL. The SSL was flown in daylight with a visibility of 12,000 miles and a cloud ceiling of 1,100 ft. The wind was a constant 13-14 knots at 270 degrees, with the exception of a vertical upward gust peaking at 25 ft/s applied just before pilots initiated the sidestep maneuver to increase the workload of the pilot during this critical phase of the maneuver. The gust profile started at approximately two nautical miles (nm) from the runway 32 Right threshold at an altitude of 677 ft AGL and died out by 629 ft AGL. The gust was programmed so that all pilots would experience the same wind strength regardless of their lateral deviation from the runway centerline.

## D. Design and Procedure

The experiment involved a between-groups comparison, so pilots needed to be counterbalanced - to the extent possible - across the two simulator conditions (FFS- and FFT-trained) to avoid any spurious effects due to differences in experience (low vs. medium or high). As described previously, pilot and simulator scheduling limitations led to an imbalance of experience levels between groups which was remedied by analyzing data from low-experienced pilots only.

All pilots received theoretical instruction in the first two weeks of the type-rating program. Over the next three weeks, they were trained with 10 sessions in the FFT-1, which has no physical motion. During the last week, pilots began the experimental part of the program: During the training phase, approximately half of the pilots trained the target maneuvers (described above) in the FFT-X and the other half trained the maneuvers in the FFS. Usually the next day,<sup>††</sup> all pilots flew the same maneuvers in the FFS (quasi-transfer phase). After the experimental phase and also within the last week, all crews flew two more 3-hour sessions in the FFS, followed by flight training in the airplane, and then performed the type-rating skill test.

During the experimental phase, all pilots first flew two takeoffs and landings in their assigned training simulator in order to familiarize themselves with the simulator before training. Next, pilots trained two scenarios in their assigned training simulator (training phase): (1) Takeoff with engine 1 failed at  $V_2$ , followed by a hand-flown ILS approach and landing; (2) Takeoff with engine 2 failed at  $V_1$ , followed by a side-step landing with microburst. Each scenario was trained three times in a row, after which the pilot and instructor filled out questionnaires about the maneuvers just flown. During the quasi-transfer phase (also referred to as the testing phase), pilots flew the same target maneuvers again, this time all in the FFS. This time, each maneuver/scenario was flown only once, and the maneuver and engine-failed order was varied as follows: (1)  $V_1$  cut (engine 2), followed by ILS approach/landing; (2)  $V_2$  cut (engine 1), followed by SSL. Immediately following the transfer/testing phase, pilots and instructors filled out more questionnaires about the maneuvers just flown.

---

<sup>††</sup> The majority of trainees were tested one day after they were trained, although in a few cases scheduling difficulties resulted in up to four days in between. There were no overall differences between the FFS and FFT groups in terms of the number of days between training and testing.

### III. Data Analysis and Results

#### A. Simulator Data

During each run of the experiment, variables that were considered useful for assessing the simulator performance, pilot performance, and pilot-control activities were recorded from the simulator computer at sampling rates of 30 Hz for the FFS and 50 Hz for the FFT. A total of 109 variables were recorded in the FFS and 90 variables were recorded in the FFT (FFS data included actuator excursions). From these raw simulator recordings, we generated a number of variables to be included in the data analysis (see Table 3 for a list of variables analyzed). The variables were calculated within the critical segments of each maneuver, specifically:

- From engine failure to 1,000 ft AGL ( $V_1$  and  $V_2$ )
- From approach fix to decision height (ILS1/Approach)
- From decision height to touchdown (ILS2/Landing)
- At touchdown (ILS & SSL)

Only touchdown data were analyzed for the sidestep landing. During this maneuver, technical difficulties in the FFS resulted in significant wind-setting differences between the FFS and the FFT, making for an unfair comparison between groups and across phases. Examination of the data, as well as formal discussions with the training center's Head of Master Reference and Flight Crew Training Manager, determined that the wind differences had not affected pilots at touchdown.

**Table 3. Variables used in the data analysis.**

Type	Variable	Maneuver(s)	Definition
Control Inputs	Pedal Reaction Time (s)	Takeoffs	Time for the pedal position to exit a 2-deg band about its initial position in response to engine failure
	Wheel Response (deg)	Takeoffs & ILS landing	RMS of wheel response, calculated by taking the square root of the total area under the wheel position power spectral density curve
	Column Response (deg)	Takeoffs & ILS	RMS of wheel response, calculated by taking the square root of the total area under the column position power spectral density curve
	Pedal Response (deg)	Takeoffs & ILS	RMS of wheel response, calculated by taking the square root of the total area under the pedal position power spectral density curve
Flight Precision	Heading StdDev (deg)	Takeoffs & ILS	STD of the deviation around the desired heading
	Airspeed Exceedance (kt)	Takeoffs & ILS	Average of absolute deviation outside 5 kts about the desired IAS
	Yaw Activity (deg/s)	Takeoffs	Mean absolute yaw rate
	Pitch Standard Deviation (deg)	Takeoffs	STD of pitch angle
	Localizer Deviation (dot)	ILS	STD of horizontal deviation of the airplane from the localizer centerline
	Glideslope Deviation (dot)	ILS	STD of vertical deviation from the glide slope reference path
	Roll Activity (deg/s)	ILS	Mean absolute roll rate
Touchdown	Vertical Speed (ft/min)	Landing only (incl. SSL)	Absolute value of vertical speed at touchdown
	Touchdown Precision (ft)	Landings only (incl. SSL)	Absolute distance from desired landing point when plane first touches ground
	IAS when Beginning Flare (kt)	ILS landing only	Average IAS at 50 ft AGL

Other technical problems<sup>‡‡</sup> led to the elimination of one or more pilot trials. As a result, some pilots had incomplete training data. To remedy this, we used the mean of available training trials for each pilot in the statistical analyses; in other words, each pilot in the analysis had one data point for the training phase (rather than three). We considered that the number of trials might affect the representativeness each pilot's mean, so we ensured that there were no differences in (1) the number of pilots in each group (FFS or FFT) that had only one or two trials included for each maneuver, and (2) which of the three trials for each maneuver were included for each group (to avoid practice effects). We did this first by comparing the trials for the two groups, and found no significant group differences. Second, we ran our main analysis (described below) with only the data from pilots with all three trials, and found the results to be very similar to the results with data from all pilots.

Our chosen analysis procedure (described below) requires that all variables be normally distributed. Examination of the data revealed that the distributions of many of the variables were asymmetrical or contained outliers that were not operationally large enough to justify deleting. Therefore, we transformed all variables to approximate normal distributions using the Box-Cox power-transform procedure. The statistics reported in the following sections are those obtained with the transformed data.

Each maneuver flight segment (defined above) was analyzed in a separate multivariate analysis of variance (MANOVA). MANOVA essentially combines all dependent variables into one variable; this method is prudent for the current study because, given the physics of airplane motion and the characteristics of human-pilot control, the performance and behavior measures discussed above are interrelated. The use of MANOVA instead of multiple univariate analyses of variance (ANOVA) was also intended to reduce the possibility of Type I error, i.e., a false *rejection* of the null hypothesis that FFS-motion has no effect. MANOVAs were conducted using Statistical Analysis Software (SAS) and employed a doubly multivariate repeated measures (DMRM) design<sup>§§</sup> using the general linear model (GLM). The DMRM design treats all dependent variables as repeated (within-subject) responses,<sup>\*\*\*</sup> with independent variables group (FFS- or FFT-trained) as the between-subjects factor and phase (training or quasi-transfer testing) as a within-subjects factor. If the overall MANOVA effect was statistically significant, additional univariate contrasts (akin to ANOVA) tested for group x phase effects on each variable to see whether any variable(s) contributed most to the effects found by the MANOVA. If the contrasts revealed any significant effects, post-hoc ANOVAs and paired t-tests were conducted to test the effect of group at each level of phase and the effect of phase on each group, respectively. Bonferroni p-value corrections<sup>†††</sup> were used to determine statistical significance on all post-hoc tests to correct for the effect of multiple tests, i.e., to reduce the probability of finding a significant effect by chance alone. Any difference with a probability to have occurred by chance of lower than 5 percent ( $p < 0.05$ ) was considered a statistically significant effect (note that statistical significance is not necessarily synonymous with operational relevance). Probabilities of lower than 10 percent were considered a trend ( $p < 0.10$ ).

It is also very important for consideration of the experimental results to know whether the data gathered was consistent enough to reveal an operationally relevant effect (resolution of the study). For this study, a rough estimate of the detectable effect sizes was found using confidence intervals. Confidence intervals (C.I.'s) were calculated on the mean differences found between each level of group and phase for each dependent variable in raw form, i.e., *before the data transformation*. The C.I.'s reveal the highest and lowest mean difference that we might possibly have found, given our small sample size, with a confidence level of 95%. To put it another way, we only have a 5% chance of finding an effect larger or smaller than the highest and lowest C.I.'s, respectively. In this paper we are reporting the upper confidence limit for each observed mean difference to show the probable highest mean difference that we could expect to find in the data. Note, however, that by calculating the confidence interval on the untransformed data, we have been trading off operational interpretability for statistical robustness, because confidence intervals also rely on normally distributed data.

---

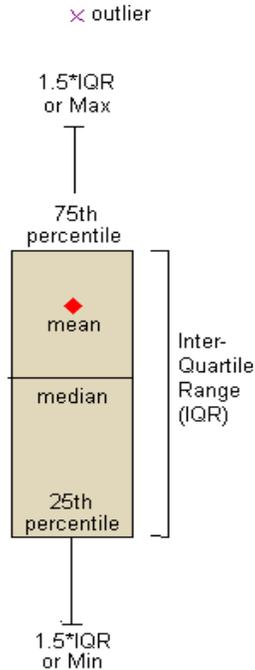
<sup>‡‡</sup> Problems included: incorrect wind settings, incorrect sound settings, a turn indicator malfunction, no yaw damper, a gas freeze, and data-recording errors.

<sup>§§</sup> The doubly multivariate repeated measures design was chosen due to the structure of the dataset, SAS capabilities, and the within-subjects comparison.

<sup>\*\*\*</sup> A significant effect of response variable indicates that the effects are different for each dependent variable. In this paper, we are not reporting response effects, but all were significant at a  $p < 0.001$  level.

<sup>†††</sup> The p-value criteria were divided by the number of tests. The probability of rejecting a true null hypothesis increases exponentially with each test; e.g., if there is a 0.95 probability that one test will be non-significant, the probability of two tests each being non-significant goes down to 0.90 ( $0.95 \times 0.95$ ), of three tests  $0.95^3 = 0.86$ , and so on. The p-value adjustment used in this study corrects for this by setting a stricter criterion for statistical significance.

As mentioned, all statistical analyses reported in this paper were run on the data after it had been transformed to a normal distribution. While the data transformation is necessary in order to assess *statistical* differences, we are also interested in looking at *operational* differences. In order to assess operationally relevant differences as well as statistical ones, we chose to report the effect sizes and confidence limits for the raw data, i.e., before the data transformation. We also display statistically significant effects (as determined using the transformed data) using box-plots of the raw data, before the transformation. Not only do these graphs depict flight-operation relevant values using raw data, but they also demonstrate the asymmetry of the raw data which justified the data transformation. For an explanation of the values reported in our box-plots, see Figure 1.



**Figure 1. Box-plot explanation.**

Each box contains 50% of the data, with the lower boundary of the box representing the 1<sup>st</sup> quartile, or 25% of the data, and the higher boundary representing the 3<sup>rd</sup> quartile, or 75% of the data; the middle line is the median, or 50% of the data. The whiskers are 1.5 times the range of the box, a common standard for box-plots. Outliers are X's and are defined as any values outside of the whiskers. If there are no outliers, the whiskers represent the minimum and maximum values of the distribution. Note: When the data is symmetrical, the mean and median should be similar; our data-transformation makes the mean and median more aligned, so the best way to interpret statistical results from our box-plots is to look at the median.

1.  $V_1$  cut

a. Flight precision for  $V_1$  cut

Table 4 shows the upper confidence limits and mean differences for all analyzed  $V_1$  cut flight-precision variables (refer to Table 3 for variable definitions). For heading standard deviation, we observed effects ranging from 0.80 to 0.98 deg, but to be 95% certain that we have included all possible effects, we need to assume effects as large as between 2.31 and 3.04 deg. For airspeed exceedance, actual effects ranged from 0.25 to 0.95 kt, with less than a 5% chance of finding an effect greater than 1.90 kt across groups and phases. Observed effects for yaw rate ranged from 0.01 to 0.30 deg/s, with a largest probable effect of 0.53 deg/s. Pitch standard deviation effects were between 0.02 and 0.05 deg, but we could have found effects as large as 0.35 to 0.51 deg with 95% certainty.

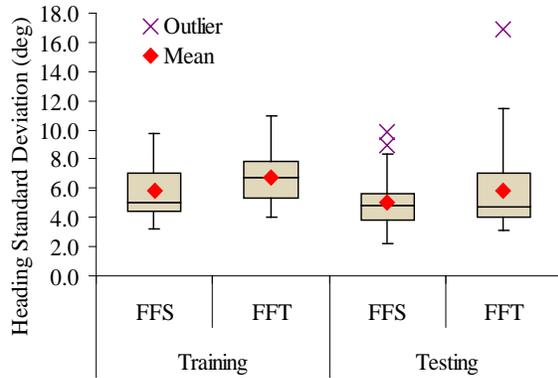
**Table 4. Upper confidence bounds (95% C.I.) and raw-data mean differences by group and phase for all examined  $V_1$  cut flight-precision variables.**

Variable	Mean Difference	Phase Difference (Trg vs. Tst)		Group Difference (FFS vs. FFT)	
		FFS	FFT	Trg	Tst
Heading StdDev (deg)	largest	2.31	3.04	2.41	2.72
	<i>observed</i>	$0.80^{TR}$	$0.92^{TR}$	$0.98^T$	$0.86^T$
Airspeed Exceeded (kt)	largest	1.19	1.90	1.11	1.63
	<i>observed</i>	$0.25^{TS}$	$0.95^{TS}$	$0.28^S$	$0.42^T$
Yaw rate (deg/s)	largest	0.28	0.53	0.47	0.21
	<i>observed</i>	$0.11^{TR}$	$0.30^{TR}$	$0.17^T$	$0.01^S$
Pitch StdDev (deg)	largest	0.35	0.51	0.44	0.44
	<i>observed</i>	$0.02^{TR}$	$0.05^{TR}$	$0.05^T$	$0.02^T$

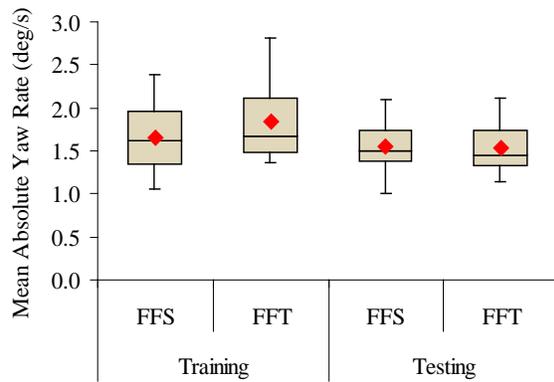
Note: Mean difference is an absolute value. Superscript notations indicate the larger of the two values that make up the mean difference (<sup>TR</sup>Training; <sup>TS</sup>Testing; <sup>S</sup>FFS; <sup>T</sup>FFT). Shaded cells represent significant differences found in the main analysis using the transformed data.

A DMRM MANOVA for  $V_1$  cut flight precision was conducted with heading standard deviation, pitch standard deviation, yaw rate, and average airspeed exceedance as dependent variables. The p-value correction for the eight post-hoc tests of each type was  $p < 0.006$  (0.05/8; to be significant at a  $p < 0.05$  equivalent). Actual p-values from tests meeting the  $p < 0.006$  criterion are reported.

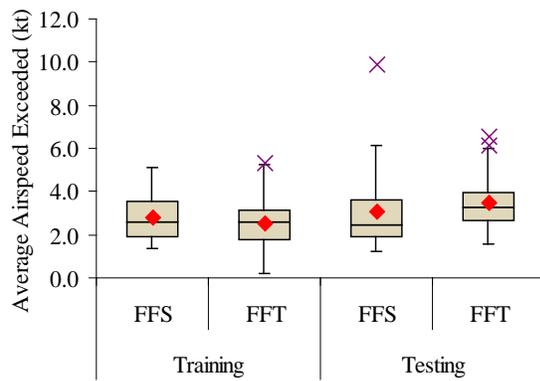
The MANOVA showed a significant overall effect of phase only [Wilks' Lambda = 0.67,  $F(4,30) = 3.76$ ,  $p < 0.05$ ]. Contrast analyses revealed significant training vs. testing differences in heading standard deviation, yaw rate, and velocity exceedance when data from both groups were pooled [heading:  $F(1,33) = 4.17$ ,  $p < 0.05$ ; yaw:  $F(1,33) = 8.01$ ,  $p < 0.01$ ; velocity:  $F(1,33) = 4.15$ ,  $p < 0.05$ ]. There were overall decreases in heading standard deviation and yaw rate and an overall increase in velocity exceeded from training to testing. When phase effects were tested separately within each group, there were no significant changes from training to testing for any variable, so neither group alone altered their flight precision from training to testing to a statistically significant degree. The significant  $V_1$  cut control-behavior variables are shown in Fig. 2 and the results are summarized below.



a) Heading Standard Deviation



b) Yaw Rate



c) Airspeed Exceedance

**Figure 2.  $V_1$  cut flight-precision variables.** Box-plots depicting the raw data (before the Box-Cox transformation) of the  $V_1$  cut precision variables for which there were statistically significant effects.

b. Control-inputs for  $V_1$  cut

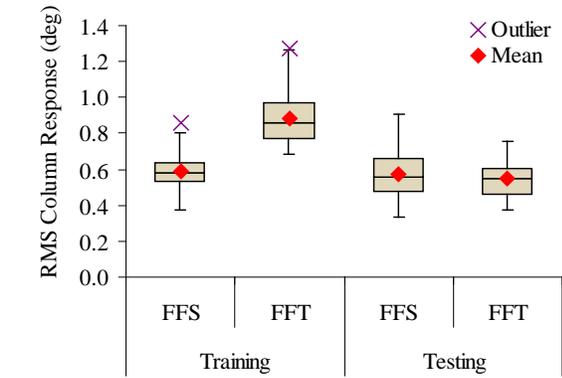
The upper confidence limits and mean differences for all analyzed  $V_1$  cut control-input variables are shown in Table 5 (refer to Table 3 for variable definitions). For pedal reaction time, the observed effect sizes (mean differences) across both groups and phases ranged from 0.16 to 0.26 seconds, and the largest probable effect that we could have found across groups and phases was 0.58 seconds. In other words, there is only a 5% chance that we would have found an effect larger than 0.58 s. For wheel response, the observed effect sizes ranged from 0.10 to 0.41 deg, with a largest probable effect of 2.39 deg across groups and phases. Column response effect sizes ranged from 0.01 to 0.34, and the largest probable effect was 0.42 deg. For pedal response, the observed effects were between 0.02 and 0.45 deg, with a largest probable effect of 0.65 deg.

For the  $V_1$  control-input variables, a DMRM MANOVA was conducted with response variables wheel, column, and pedal responses, and pedal reaction time as dependent variables. The p-value correction for the eight post-hoc tests of each type was  $p < 0.006$ . Actual p-values from tests meeting this criterion are reported.

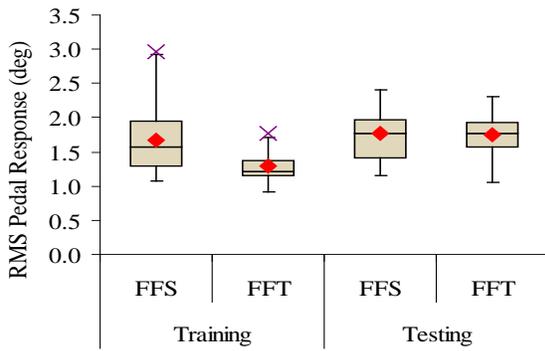
**Table 5. Upper confidence bounds (95% C.I.) and raw-data mean differences by group and phase for all examined  $V_1$  cut control-input variables.**

Variable	Mean Difference	Phase Difference (Trg vs. Tst)		Group Difference (FFS vs. FFT)	
		FFS	FFT	Trg	Tst
Pedal R.T. (s)	largest	0.58	0.56	0.49	0.51
	<i>observed</i>	$0.26^{TS}$	$0.21^{TS}$	$0.20^T$	$0.16^T$
Wheel RMS (deg)	largest	1.87	1.92	1.98	2.39
	<i>observed</i>	$0.10^{TR}$	$0.40^{TR}$	$0.41^T$	$0.11^T$
Column RMS (deg)	largest	0.07	0.42	0.39	0.12
	<i>observed</i>	$0.01^{TR}$	$0.34^{TR}$	$0.30^T$	$0.03^S$
Pedal RMS (deg)	largest	0.31	0.64	0.65	0.26
	<i>observed</i>	$0.09^{TS}$	$0.45^{TS}$	$0.38^S$	$0.02^S$

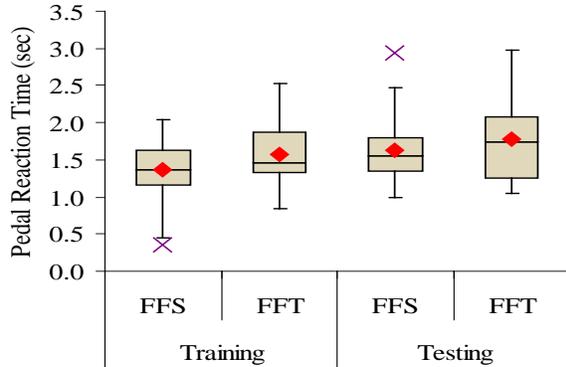
Note: Mean difference is an absolute value. Superscript notations indicate the larger of the two values that make up the mean difference (<sup>TR</sup>Training; <sup>TS</sup>Testing; <sup>S</sup>FFS; <sup>T</sup>FFT). Shaded cells represent significant differences found in the main analysis using the transformed data.



a) *Column Response*



b) *Pedal Response*



c) *Pedal Reaction Time*

**Figure 3.  $V_1$  cut control-input variables.** Box-plots depicting the raw data (before the Box-Cox transformation) of the  $V_1$  cut control variables for which there were statistically significant effects.

All MANOVA effects for the  $V_1$  cut control-input variables were significant: there were significant overall effects of group, phase, and an interaction of group and phase [group: Wilks' Lambda = 0.46,  $F(4,30) = 8.68$ ,  $p < 0.001$ ; phase: Wilks' Lambda = 0.17,  $F(4,30) = 36.31$ ,  $p < 0.001$ ; interaction: Wilks' Lambda = 0.26,  $F(4,30) = 21.21$ ,  $p < 0.001$ ]. The contrast analyses showed that the MANOVA effects were driven primarily by three variables: column response, pedal response, and pedal reaction time.

The significant group effect was driven mostly by column response. There was a significant overall group difference in column response, with the FFS-trained group using less column than the FFT-trained group [ $F(1,33) = 10.89$ ,  $p < 0.01$ ]. This effect was qualified by the difference being significant only during training [ $F(1,33) = 45.74$ ,  $p < 0.001$ ]. There was no significant group difference at testing, and thus the group difference did not transfer, and both groups used the column similarly at testing. There was also a trend for a group difference for pedal response, with FFS-trained pilots being overall more responsive than FFT-trained pilots [ $F(1,33) = 3.68$ ,  $p = 0.06$ ]. Again, this difference only remained significant during the training phase, and thus did not transfer [ $F(1,33) = 8.95$ ,  $p < 0.01$ ].

The MANOVA phase effect was driven by column response, pedal response, and pedal reaction time. There were significant overall phase differences in column and pedal responses, with column response decreasing and pedal response increasing from training to testing [column:  $F(1,33) = 54.13$ ,  $p < 0.001$ ; pedal:  $F(1,33) = 21.23$ ,  $p < 0.001$ ]. Post-hoc analyses revealed that these differences were only significant for the FFT-trained group [column:  $t(15) = 8.56$ ,  $p < 0.001$ ; pedal:  $t(15) = -5.84$ ,  $p < 0.001$ ]. There was a significant difference in pedal reaction time between training and testing phases, but only with both FFS- and FFT-trained groups combined; reaction time was slower at testing than it was at training [ $F(1,33) = 4.49$ ,  $p < 0.05$ ].

The interaction effect was driven by column and pedal responses [column:  $F(1,33) = 43.47$ ,  $p < 0.001$ ; pedal:  $F(1,33) = 8.23$ ,  $p < 0.01$ ]. The interactions show that the presence of a group effect depends on phase, and the presence of a phase effect depends on group. The results of all significant  $V_1$  cut control-input effects are depicted in Figure 3 and summarized below.

c. Summary of results for  $V_1$  cut

- There were no group differences in flight precision. Overall, the FFS-trained group used the column less and the pedal more than the FFT-trained group. These effects were significant during training, but disappeared when both groups transferred to the FFS.
- There was an overall decrease in heading standard deviation and yaw rate and an overall increase in airspeed exceedance at testing. Column response decreased and pedal response and pedal reaction time increased overall from training to testing. For column and pedal responses, these changes were only significant for the FFT-trained group.
- There were no interactions for flight precision variables, but for column and pedal responses, the group effects depended on phase (training only) and the phase effects depended on group (FFT-trained group only).

2.  $V_2$  cut

a. Flight precision for  $V_2$  cut

The upper confidence limits and observed mean differences for the  $V_2$  cut flight-precision variables are shown in Table 6 (refer to Table 3 for variable definitions). For heading standard deviation, we observed effects ranging from 0.21 to 1.52 deg. The highest probable effect we could have found across groups and phases was 2.43 deg. Effect sizes for airspeed exceedance ranged from 0.32 to 0.97 kt, with the largest probable effect size across groups and phases being 2.33 kt. For yaw rate, we observed effects between 0.18 and 0.47 deg/s, with the highest probable effect being 0.72 deg/s. Pitch standard deviation effects ranged from 0.11 to 0.22 deg and the highest effect we could have found with 95% confidence was 0.60 deg across both groups and phases.

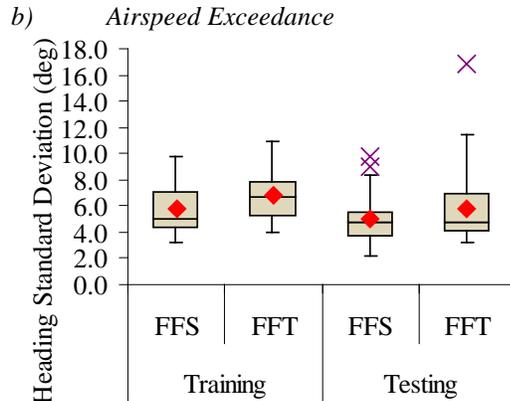
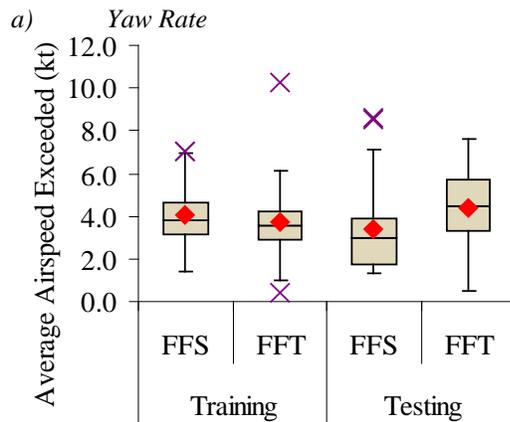
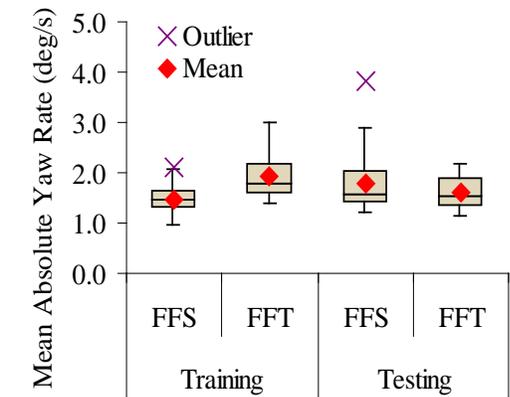
**Table 6. Upper confidence bounds (95% C.I.) and raw-data mean differences by group and phase for all examined  $V_2$  cut flight-precision variables.**

Variable	Mean Difference	Phase Difference (Trg vs. Tst)		Group Difference (FFS vs. FFT)	
		FFS	FFT	Trg	Tst
Heading StdDev (deg)	largest	2.43	1.43	3.02	1.64
	<i>observed</i>	$1.32^{TS}$	$0.40^{TR}$	$1.52^T$	$0.21^S$
Airspeed Exceeded (kt)	largest	1.62	1.76	1.47	2.33
	<i>observed</i>	$0.65^{TR}$	$0.64^{TS}$	$0.32^S$	$0.97^T$
Yaw rate (deg/s)	largest	0.57	0.57	0.72	0.52
	<i>observed</i>	$0.33^{TS}$	$0.32^{TR}$	$0.47^T$	$0.18^S$
Pitch StdDev (deg)	largest	0.43	0.55	0.60	0.51
	<i>observed</i>	$0.11^{TS}$	$0.20^{TS}$	$0.22^S$	$0.13^S$

Note: Mean difference is an absolute value. Superscript notations indicate the larger of the two values that make up the mean difference (<sup>TR</sup>Training; <sup>TS</sup>Testing; <sup>S</sup>FFS; <sup>T</sup>FFT). Shaded cells represent significant differences found in the main analysis using the transformed data.

The same DMRM MANOVA was run for the  $V_2$  cut flight-precision variables as for the  $V_1$  cut. Actual post-hoc  $p$ -values are reported but we only report those that are significant at the criterion of  $p < 0.006$ , as mentioned above.

The only significant MANOVA effect was the interaction [Wilks' Lambda = 0.63,  $F(4,33) = 4.89$ ,  $p < 0.01$ ]. This effect was driven primarily by yaw rate and airspeed exceedance, as well as by heading standard deviation. There was a significant interaction of yaw rate [ $F(1,36) = 16.87$ ,  $p < 0.001$ ]. This interaction is qualified by an overall trend for the FFS-trained group to exhibit lower yaw rates than the FFT-trained group [ $F(1,36) = 3.08$ ,  $p = 0.09$ ]. Although this was only a trend, post-hoc ANOVAs suggested that the difference was only significant during training [ $F(1,36) = 15.44$ ,  $p < 0.001$ ]. Thus, group only had an effect on yaw rate during training and not during testing. The interaction for airspeed exceedance was also significant [ $F(1,36) = 4.84$ ,  $p < 0.05$ ]. This interaction shows that the group effect was smaller during training, when the FFS-trained group had higher airspeed exceedance, than it was at testing, when the FFS-trained group had lower exceedance than the FFT-trained group. This effect, however, is attenuated by the fact that there were no significant post-hoc group or phase effects. A trend for an interaction for heading standard deviation showed that group had a larger effect on flight precision during training, when the FFS-trained group had lower heading standard deviation than the FFT-trained group, than it did at testing, when the FFS-trained group had higher heading standard deviation than the FFT-trained group. Also, phase changes in heading standard deviation from training to testing were smaller for the FFS-trained group, who increased, than for the FFT-trained group, who decreased [ $F(1,36) = 3.94$ ,  $p = 0.06$ ]. Again, this effect is attenuated by a lack of significant post-hoc group or phase effects. The results are depicted in Figure 5 and are summarized at the end of the  $V_2$  cut results section.



c) *Heading Standard Deviation*

**Figure 5.  $V_2$  cut flight-precision variables.** Box-plots depicting the raw data (before the Box-Cox transformation) of the  $V_2$  cut precision variables for which there were statistically significant effects.

b. Control-inputs for  $V_2$  cut

The  $V_2$  cut control-input mean differences and upper confidence limits are shown in Table 7 (see Table 3 for variable definitions). For pedal reaction time, the observed mean differences ranged from 0.03 to 0.39 s. The highest mean difference that we could have found with 95% confidence across both groups and phases was 0.91 s. For wheel reaction time, we found effect sizes between 0.41 and 1.51 deg, with the highest probable effect size of 2.94 deg, suggesting that it is unlikely that we would find anything larger than this across groups and phases. The range of effect sizes for column response was 0 to 0.17 deg, with the largest probable effect size of 0.26 deg. For pedal response, we observed effect sizes of 0.12 to 0.58, and the largest probable effect we could have found was 0.77 deg across groups and phases.

**Table 7. Upper confidence bounds (95% C.I.) and raw-data mean differences by group and phase for all examined  $V_2$  cut control-input variables.**

Variable	Mean Difference	Phase Difference (Trg vs. Tst)		Group Difference (FFS vs. FFT)	
		FFS	FFT	Trg	Tst
Pedal R.T. (s)	largest	0.91	0.48	0.28	0.71
	<i>observed</i>	$0.39^{TS}$	$0.17^{TS}$	$0.03^T$	$0.19^S$
Wheel RMS (deg)	largest	2.59	2.05	2.94	2.02
	<i>observed</i>	$1.23^{TS}$	$0.68^{TR}$	$1.51^T$	$0.41^S$
Column RMS (deg)	largest	0.10	0.18	0.26	0.10
	<i>observed</i>	$0.05^{TS}$	$0.13^{TR}$	$0.17^T$	$0.00^S$
Pedal RMS (deg)	largest	0.71	0.77	0.46	0.50
	<i>observed</i>	$0.44^{TS}$	$0.58^{TS}$	$0.26^S$	$0.12^S$

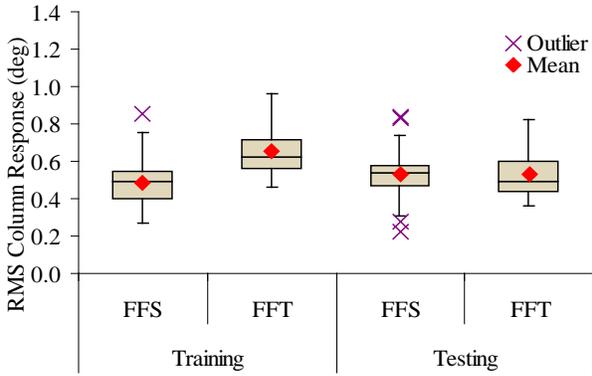
*Note: Mean difference is an absolute value. Superscript notations indicate the larger of the two values that make up the mean difference (<sup>TR</sup>Training; <sup>TS</sup>Testing; <sup>S</sup>FFS; <sup>T</sup>FFT). Shaded cells represent significant differences found in the main analysis using the transformed data.*

The same DMRM MANOVAs were run for the  $V_2$  cut control inputs as for the  $V_1$  cut. Again, the p-value correction for the eight post-hoc tests of each type was  $p < 0.006$  and the actual p-values from tests meeting this criterion are reported. All MANOVA effects for the  $V_2$  cut control-input variables were significant. There were significant overall effects of group, phase, and an interaction of group and phase [group: Wilks' Lambda = 0.74,  $F(4,33) = 2.89$ ,  $p < 0.05$ ; phase: Wilks' Lambda = 0.21,  $F(4,33) = 30.45$ ,  $p < 0.001$ ; interaction: Wilks' Lambda = 0.34,  $F(4,33) = 16.13$ ,  $p < 0.001$ ]. The contrasts revealed that the significant effects were driven mostly by column, pedal, and wheel response variables.

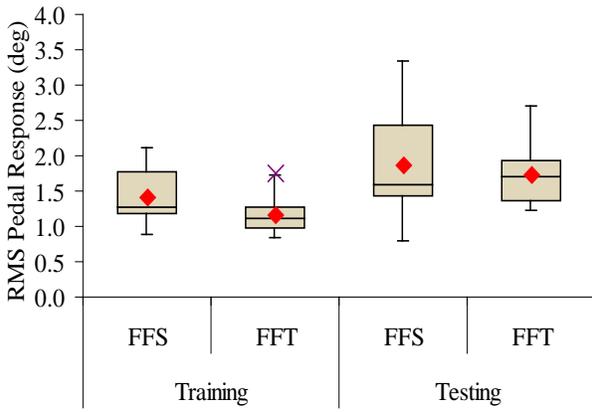
The significant group effect was driven by column response; overall, the FFS-trained group exhibited less column activity than the FFT-trained group [ $F(1,36) = 4.58$ ,  $p < 0.05$ ]. However, the post-hoc ANOVAs revealed that this difference was only significant during training and not during testing; thus, the group difference did not transfer [training:  $F(1,36) = 18.74$ ,  $p < 0.001$ ].

Phase differences were found for both column and pedal responses, with a significant overall decrease in column response and a significant overall increase in pedal response from training to testing [column:  $F(1,36) = 4.28$ ,  $p < 0.05$ ; pedal:  $F(1,36) = 48.22$ ,  $p < 0.001$ ]. Post-hoc t-tests revealed that these phase differences were only significant for the FFT-trained group [column:  $t(17) = 5.57$ ,  $p < 0.001$ ; pedal:  $t(17) = -7.34$ ,  $p < 0.001$ ]. For pedal response, however, there was a trend for the FFS-trained pilots to also increase their pedal activity [ $t(19) = -3.05$ ,  $p = 0.007$ , adjusted  $p = 0.053$ ].

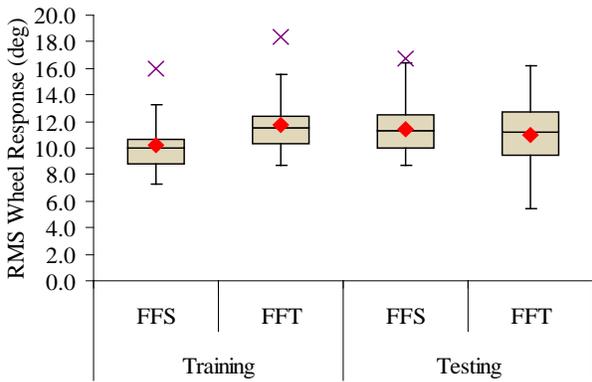
There were significant interactions of group and phase for column response and for pedal response [column:  $F(1,36) = 22.47$ ,  $p < 0.001$ ; pedal:  $F(1,36) = 5.32$ ,  $p < 0.05$ ]. For column response, there was an effect of group during training only and an effect of phase only for the FFT-trained group (statistics reported above). For pedal response, the increase in pedal activity from training to testing was smaller for the FFS-trained group than for the FFT-trained group. There was also a significant interaction for wheel response, indicating that the group effect was greater during training and the phase effect was greater for the FFS-trained group [ $F(1,36) = 5.38$ ,  $p < 0.05$ ]. Box-plots depicting significant effects of  $V_2$  cut control inputs are displayed in Figure 6 and results are summarized below.



a) Column Response



b) Pedal Response



c) Wheel Response

**Figure 6.  $V_2$  cut control-input variables.** Box-plots depicting the raw data (before the Box-Cox transformation) of the  $V_2$  cut control variables for which there were statistically significant effects.

- c. Summary of results for  $V_2$  cut
- There were no group effects for flight precision. For control inputs, the FFS-trained group showed overall less column activity than the FFT-trained group, but this effect was only significant during training.
  - There were no phase changes for flight precision, but for control inputs there were overall changes in column response and pedal response. Column response decreased at testing, but this effect was only significant for the FFT-trained group. Pedal response increased at testing; this effect was a trend for the FFS-trained group and was significant for the FFT-trained group.
  - For flight precision, there were significant interactions of group and phase for yaw rate and airspeed exceedance and an interaction trend for heading standard deviation. For yaw rate, the group effect depended on phase. For airspeed exceedance, the magnitude of group differences depended on phase. For heading standard deviation, the direction and the magnitude of group differences depended on phase and the direction and magnitude of phase differences depended on group. For control inputs, the significant interaction on column response showed that the group effect depended on phase and the phase effect depended on group. For pedal response, the interaction showed that the magnitude of the phase effect depended on group. The significant interaction on wheel response showed that the direction and magnitude of group differences depended on phase, and vice-versa.

3. ILS 1: Approach Fix to Decision Height

a. Flight precision for ILS1

The upper confidence intervals and observed effect sizes for the ILS1 flight-precision raw data are in Table 8 (see Table 3 for variable definitions). For localizer standard deviation, we found effect sizes ranging from 0.01 to 0.07 dot, but we are 95% confident that we could have found effects as large as 0.09 to 0.20 dot. The effect sizes for airspeed exceedance were between 0.08 and 1.72 kt, with a highest probable effect size across groups and phases of 4.89 kt. For roll rate, we observed effects between 0.04 and 0.34 deg/s. The largest probable effect of roll rate that we could have found across groups and phases was 0.59 deg/s. For glideslope standard deviation, we found effects ranging from 0.02 to 0.26 dot and a probable largest effect size of 0.68 across groups and phases.

**Table 8. Upper confidence bounds (95% C.I.) and raw-data mean differences by group and phase for all examined ILS1 flight-precision variables.**

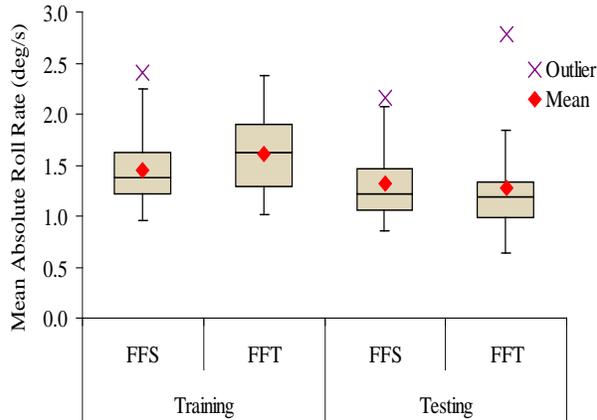
Variable	Mean Difference	Phase Difference (Trg vs. Tst)		Group Difference (FFS vs. FFT)	
		FFS	FFT	Trg	Tst
Loc. StdDev (dot)	largest	0.14	0.20	0.18	0.09
	observed	0.05 <sup>TR</sup>	0.07 <sup>TR</sup>	0.03 <sup>T</sup>	0.01 <sup>T</sup>
IAS exceeded (kt)	upper	2.90	1.86	4.41	4.89
	mean diff.	0.24 <sup>TS</sup>	0.08 <sup>TR</sup>	1.40 <sup>S</sup>	1.72 <sup>S</sup>
Roll rate (deg/s)	upper	0.38	0.59	0.46	0.38
	mean diff.	0.14 <sup>TR</sup>	0.34 <sup>TR</sup>	0.15 <sup>T</sup>	0.04 <sup>S</sup>
G.S. StdDev (dot)	upper	0.37	0.68	0.59	0.32
	mean diff.	0.12 <sup>TS</sup>	0.26 <sup>TR</sup>	0.11 <sup>T</sup>	0.02 <sup>S</sup>

Note: Mean difference is an absolute value. Superscript notations indicate the larger of the two values that make up the mean difference (<sup>TR</sup>Training; <sup>TS</sup>Testing; <sup>S</sup>FFS; <sup>T</sup>FFT). Shaded cells represent significant differences found in the main analysis using the transformed data.

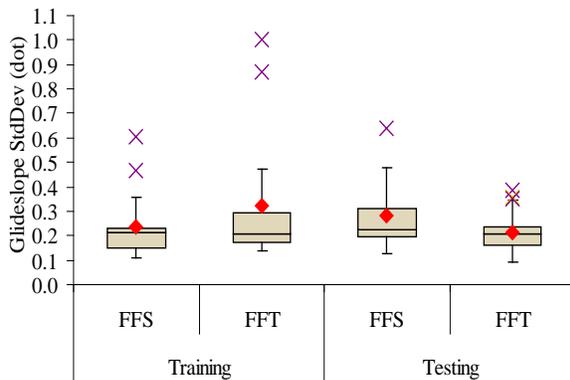
The MANOVA for ILS1 flight-precision variables was conducted with localizer and glideslope standard deviations, roll activity, and airspeed exceedance as dependent variables. The  $p$ -value correction for the eight post-hoc tests of each type was  $p < 0.006$  and the actual  $p$ -values from tests meeting this criterion are reported. The MANOVA yielded significant effects of phase and an interaction [phase: Wilks' Lambda = 0.69,  $F(4,26) = 2.88$ ,  $p < 0.05$ ; interaction: Wilks' Lambda = 0.63,  $F(4,26) = 3.74$ ,  $p < 0.05$ ]. These effects were driven by roll rate and glideslope standard deviation.

The phase effect was driven by roll rate, for which there was a significant overall decrease from training to testing [ $F(1,29) = 9.34$ ,  $p < 0.05$ ]. In the post-hoc tests, this decrease in roll rate emerged as a trend for the FFT-trained group but was not significant for the FFS-trained group [FFT:  $t(15) = 2.92$ ,  $p = 0.01$ , adjusted  $p = 0.08$ ].

The MANOVA interaction effect was driven by glideslope standard deviation, for which there was a significant interaction of group and phase [ $F(1,29) = 6.34, p < 0.05$ ]. This interaction shows that there was a group difference in glideslope standard deviation during training, when the FFS-trained group had lower glideslope standard deviation, but there was no difference during testing. Also, decreases in glideslope standard deviation from training to testing were smaller for the FFS-trained group than for the FFT-trained group. This effect, however, is attenuated by the fact that no post-hoc effects emerged. All significant results are in Figure 7 and summarized below.



a) Roll Rate



b) Glideslope Standard Deviation

**Figure 7. ILS1 flight-precision variables.** Box-plots depicting the raw data (before the Box-Cox transformation) of the ILS1 control variables for which there were statistically significant effects.

b. Control-inputs for ILS1

The observed raw mean differences and upper confidence limits for the ILS1 are presented in Table 9 (for definitions of the variables, refer to Table 3). The observed effect sizes for wheel response ranged from 0.13 to 1.60 deg. Across both groups and phases, the highest probable effect size that we could have found for wheel response was 3.06 deg. For column response, we found effects between 0.01 and 0.04 deg, and the highest effect size we could have found with 95% certainty was 0.13 deg. The observed effect sizes for pedal response ranged from 0.03 to 0.24 deg, with a highest probable effect of 0.55 deg across groups and phases.

**Table 9. Upper confidence bounds (95% C.I.) and raw-data mean differences by group and phase for all examined ILS1 control-input variables.**

Variable	Mean Difference	Phase Difference (Trg vs. Tst)		Group Difference (FFS vs. FFT)	
		FFS	FFT	Trg	Tst
Wheel RMS (deg)	largest	1.90	2.86	3.06	1.59
	<i>observed</i>	<i>0.46<sup>TR</sup></i>	<i>1.27<sup>TS</sup></i>	<i>1.60<sup>T</sup></i>	<i>0.13<sup>T</sup></i>
Column RMS (deg)	largest	0.08	0.13	0.12	0.13
	<i>observed</i>	<i>0.02<sup>TR</sup></i>	<i>0.04<sup>TR</sup></i>	<i>0.01<sup>S</sup></i>	<i>0.04<sup>S</sup></i>
Pedal RMS (deg)	largest	0.20	0.30	0.49	0.55
	<i>observed</i>	<i>0.03<sup>TS</sup></i>	<i>0.04<sup>TS</sup></i>	<i>0.24<sup>S</sup></i>	<i>0.23<sup>S</sup></i>

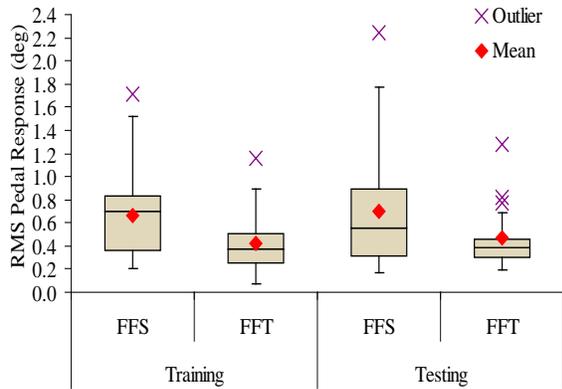
*Note: Mean difference is an absolute value. Superscript notations indicate the larger of the two values that make up the mean difference (<sup>TR</sup>Training; <sup>TS</sup>Testing; <sup>S</sup>FFS; <sup>T</sup>FFT). Shaded cells represent significant differences found in the main analysis using the transformed data.*

For ILS1 control-input variables, a DMRM MANOVA was conducted with wheel, column, and pedal responses as dependent variables. The  $p$ -value correction for the six post-hoc tests of each type was  $p < 0.008$  ( $0.05/6$ ; to be significant at  $p < 0.05$ ); actual  $p$ -values from tests meeting this criterion are reported. The MANOVA was significant for the group effect and there were trends for phase and the interaction [group: Wilks' Lambda = 0.65,  $F(3,27) = 4.86$ ,  $p < 0.01$ ; phase: Wilks' Lambda = 0.79,  $F(3,27) = 2.32$ ,  $p = 0.10$ ; interaction: Wilks' Lambda = 0.77,  $F(3,27) = 2.72$ ,  $p = 0.06$ ]. These effects were driven by pedal response and wheel response.

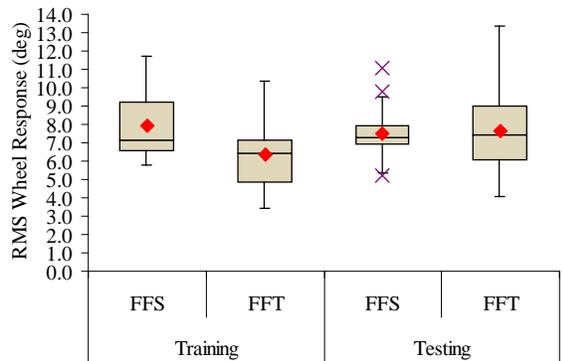
The MANOVA group effect was driven mostly by pedal response; contrasts revealed a significant overall group effect, with FFS-trained pilots showing overall more pedal activity than FFT-trained pilots [ $F(1,29) = 4.76$ ,  $p < 0.05$ ]. Post-hoc ANOVAs revealed no differences in group effects for pedal activity when differences during training and testing phases were looked at separately; thus, the group effect only remained significant when results of training and testing were combined. There was also a trend for an overall group difference in wheel response, with FFS-trained pilots exhibiting overall more wheel activity than the FFT-trained group, with no significant post-hoc differences [ $F(1,29) = 3.01$ ,  $p = 0.09$ ].

The MANOVA trend for phase disappeared when contrasts analyzed the phase effect for each variable separately. Thus, there were no significant phase differences for any one variable alone.

The MANOVA interaction trend was driven by a trend for wheel response [ $F(1,29) = 3.07$ ,  $p = 0.09$ ]. The interaction showed that group differences in wheel response were larger during training than during testing, with the FFS-trained group having lower wheel activity than the FFT-trained group in both phases. There was also a smaller phase change for the FFS-trained group, who decreased their wheel activity, than for the FFT-trained group, who increased their wheel activity. These effects are however attenuated by the fact that no post-hoc effects emerged. Results for the ILS1 control-input variables are shown in Figure 8 and are summarized below.



a) Pedal Response



b) Wheel Response

**Figure 8. ILS1 control-input variables.** Box-plots depicting the raw data (before the Box-Cox transformation) of the ILS1 control variables for which there were statistically significant effects.

c. Summary of results for ILS1

- There were no group differences in flight precision. For control inputs, the FFS-trained group had overall more pedal activity than the FFT-trained group. FFS-trained pilots also had overall more wheel activity (trend only).
- There was an overall decrease in roll rate from training to testing, with a trend of this effect for the FFT-trained group only. There were no phase differences for individual control inputs.
- An interaction for glideslope standard deviation showed that group differences depended on phase and that the magnitude of phase differences depended on group. There was also an interaction of wheel response, showing that the magnitude of the group effect depended on phase and that the direction and magnitude of the phase effect depended on group.

4. ILS2: Decision Height to Touchdown

a. Flight-precision for ILS2

The upper confidence limits and observed raw-data effect sizes for the ILS2 flight-precision variables are shown in Table 10 (refer to Table 3 for variable definitions). The effect sizes for localizer standard deviation ranged from 0.03 to 0.14 dot, with the highest probable effect size being 0.25 dot across both groups and phases. For airspeed exceedance, the range of observed effects was 0.07 to 0.82 kt, with the highest upper limit of 2.08 kt. Roll rate ranged in observed effect size from 0.12 to 0.54 deg/s and the highest probable effect was 1.25 deg/s across groups and phases. For glideslope standard deviation, the observed effects were between 0.01 and 0.10 dot, with the highest probable effect being 0.42 dot.

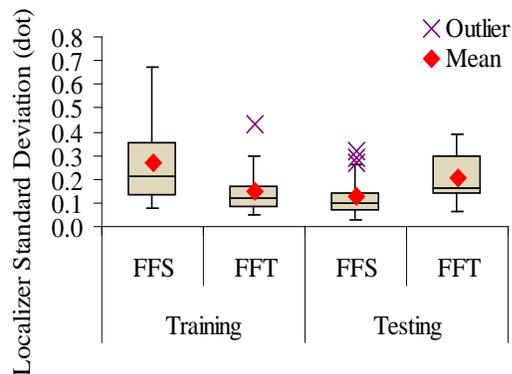
**Table 10. Upper confidence bounds (95% C.I.) and raw-data mean differences by group and phase for all examined ILS2 flight-precision variables.**

Variable	Mean Difference	Phase Difference (Trg vs. Tst)		Group Difference (FFS vs. FFT)	
		FFS	FFT	Trg	Tst
Loc. StdDev (dot)	largest	0.25	0.13	0.22	0.15
	<i>observed</i>	<i>0.14<sup>TR</sup></i>	<i>0.03<sup>TS</sup></i>	<i>0.10<sup>S</sup></i>	<i>0.08<sup>T</sup></i>
IAS exceeded (kt)	largest	1.02	1.15	1.72	2.08
	<i>observed</i>	<i>0.14<sup>TS</sup></i>	<i>0.07<sup>TR</sup></i>	<i>0.61<sup>S</sup></i>	<i>0.82<sup>S</sup></i>
Roll rate (deg/s)	largest	0.83	0.70	0.96	1.23
	<i>observed</i>	<i>0.28<sup>TR</sup></i>	<i>0.12<sup>TR</sup></i>	<i>0.38<sup>T</sup></i>	<i>0.54<sup>T</sup></i>
G.S. StdDev (dot)	largest	0.42	0.37	0.31	0.39
	<i>observed</i>	<i>0.10<sup>TS</sup></i>	<i>0.03<sup>TS</sup></i>	<i>0.01<sup>S</sup></i>	<i>0.07<sup>S</sup></i>

Note: Mean difference is an absolute value. Superscript notations indicate the larger of the two values that make up the mean difference (<sup>TR</sup>Training; <sup>TS</sup>Testing; <sup>S</sup>FFS; <sup>T</sup>FFT). Shaded cells represent significant differences found in the main analysis using the transformed data.

The MANOVA for ILS2 flight-precision variables was conducted with localizer and glideslope standard deviations, roll rate, and airspeed exceedance as dependent variables. The p-value correction for the eight post-hoc tests of each type was  $p < 0.006$  (0.05/8; to be significant at  $p < 0.05$ ); actual p-values from tests meeting this criterion are reported. The only significant MANOVA effect was the interaction [Wilks' Lambda = 0.61,  $F(4,25) = 3.93$ ,  $p < 0.05$ ].

The MANOVA interaction effect was influenced by a significant group by phase interaction on localizer standard deviation [ $F(1,28) = 13.56$ ,  $p < 0.001$ ]. The interaction is qualified by a significant post-hoc showing that there was a significant decrease in localizer standard deviation from training to testing for the FFS-trained group only [ $t(15) = 3.72$ ,  $p < 0.01$ ]. Thus, there was an effect of phase for the FFS-trained group but not for the FFT-trained group. The results are displayed in Figure 9 and are summarized below.



Localizer Standard Deviation

**Figure 9. ILS2 flight-precision.** Box-plots depicting the raw data (before the Box-Cox transformation) of the ILS2 flight-precision variables for which there were statistically significant effects.

b. Control-inputs for ILS2

The upper confidence limits and raw-data effect sizes for the ILS2 control-inputs are in Table 11 (see Table 3 for variable definitions). For wheel response, we observed effect sizes ranging from 1.28 to 2.26 deg, with the largest probable effect size of 4.10 deg across groups and phases. Column response effects ranged from 0.03 to 0.26 deg and the greatest effect we could expect to find across groups and phases was 0.37 deg. Pedal response effects were between 0.06 and 0.68 deg, with the highest probable effect size being 0.92 deg.

A DMRM MANOVA investigated group, phase, and interaction effects on ILS2 control variables wheel, column, and pedal response. The p-value correction for the six post-hoc tests of each type was  $p < 0.008$  ( $0.05/6$ ; to be significant at  $p < 0.05$ ) and the actual p-values from tests meeting this criterion are reported. The MANOVA yielded significant effects of phase and an interaction [phase: Wilks' Lambda = 0.43,  $F(3,26) = 11.46$   $p < 0.001$ ; interaction: Wilks' Lambda = 0.40,  $F(3,26) = 12.89$ ,  $p < 0.001$ ]. All three control-input variables influenced the MANOVA effects.

**Table 11. Upper confidence bounds (95% C.I.) and raw-data mean differences by group and phase for all examined ILS2 control-input variables.**

Variable	Mean Difference	Phase Difference (Trg vs. Tst)		Group Difference (FFS vs. FFT)	
		FFS	FFT	Trg	Tst
Wheel Resp. (deg)	largest	3.39	4.10	2.94	4.90
	observed	1.41 <sup>TR</sup>	2.12 <sup>TS</sup>	1.28 <sup>S</sup>	2.26 <sup>T</sup>
Column Resp. (deg)	largest	0.12	0.37	0.33	0.14
	observed	0.03 <sup>TR</sup>	0.26 <sup>TR</sup>	0.21 <sup>T</sup>	0.03 <sup>S</sup>
Pedal Resp. (deg)	largest	0.41	0.92	0.66	0.74
	observed	0.06 <sup>T<sup>R</sup></sup>	0.68 <sup>T<sup>S</sup></sup>	0.40 <sup>S</sup>	0.33 <sup>T</sup>

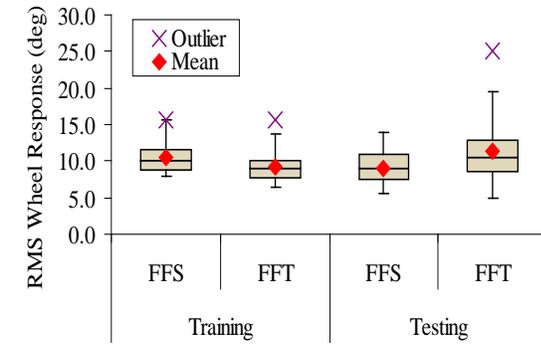
Note: Mean difference is an absolute value. Superscript notations indicate the larger of the two values that make up the mean difference (<sup>TR</sup>Training; <sup>TS</sup>Testing; <sup>S</sup>FFS; <sup>T</sup>FFT). Shaded cells represent significant differences found in the main analysis using the transformed data.

The significant phase effect was driven by column response and pedal response. There was a significant overall decrease in column activity from training to testing [ $F(1,28) = 19.63$ ,  $p < 0.001$ ]. Post-hoc t-tests revealed that this decrease was only significant for the FFT-trained group [ $t(13) = 5.70$ ,  $p < 0.001$ ]. The significant phase effect for pedal response showed an overall increase in pedal activity from training to testing [ $F(1,28) = 6.93$ ,  $p < 0.05$ ], but the post-hoc t-tests also confirmed that this increase was only significant for the FFT-trained group [ $t(13) = -5.55$ ,  $p < 0.001$ ].

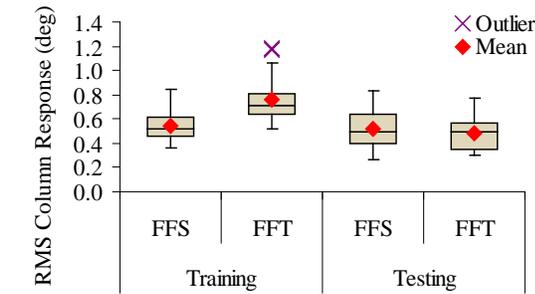
The MANOVA interaction effect was evident in all three control-input variables [wheel:  $F(1,28) = 5.60$ ,  $p < 0.05$ ; column:  $F(1,28) = 10.30$ ,  $p < 0.01$ ; pedal:  $F(1,28) = 13.16$ ,  $p < 0.01$ ]. For wheel response, a significant interaction showed the group difference was smaller during training, when the FFS-trained group had greater wheel activity, than during testing, when the FFS-trained group had less wheel activity than the FFT-trained group. The phase change was also smaller for the FFS-trained group, who decreased their wheel activity, than for the FFT-trained group, who increased it. The significant interaction for column response was qualified by a significant post-hoc effect revealing that the FFS-trained group had lower column activity than the FFT-trained group, but this effect was only significant during training [ $F(1,28) = 13.11$ ,  $p < 0.01$ ]. Group only had an effect on column response during the training phase and phase only had an effect for the FFT-trained group. For pedal response, the interaction was qualified by a finding of higher pedal activity for the FFS-trained group than for the FFT-trained group during training only [ $F(1,28) = 8.82$ ,  $p < 0.01$ ]. Group only had an effect on pedal response during the training phase and phase only had an effect on the FFT-trained group. The ILS2 control-input variables are depicted in Fig. 10 and are summarized below.

c. Summary of results for ILS2

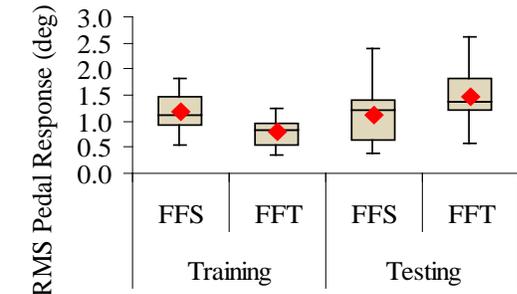
- There were no group effects on flight precision or control inputs.
- There were no phase effects on flight precision. There were overall decreases in column response and increases pedal response, but these effects were only significant for the FFT-trained group.
- For flight precision, there was an interaction of group and phase for localizer standard deviation, showing that the phase effect depended on group. For control inputs, the direction and magnitude of group differences in wheel response depended on phase, and the direction and magnitude of the phase differences depended on group. For column and pedal responses, the group effects depended on phase and the phase effects depended on group.



a) *Wheel Response*



b) *Column Response*



c) *Pedal Response*

**Figure 10. ILS2 control-input variables.** *Box-plots depicting the raw data (before the Box-Cox transformation) of the ILS1 control variables for which there were statistically significant effects.*

### 5. ILS Touchdown

The upper confidence limits and effect sizes for the ILS touchdown are listed in Table 12 (for a definition of the variables, refer to Table 3). For touchdown precision, effect sizes ranged from 54.48 to 173.40 ft and the largest probable effect size was 595.14 ft. The vertical velocity ranged in observed effect sizes from 7.44 to 143.58 ft/min, and the largest probable effect size that we could have found across groups and phases was 204.80 ft/min. For IAS at 50 ft AGL, we found effect sizes between 0.12 and 1.27 kt. The largest probable effect size we could expect to find for IAS was 5.01 kt.

A DMRM MANOVA was conducted with velocity at touchdown and touchdown precision as dependent variables. The  $p$ -value correction for the four post-hoc tests of each type was  $p < 0.013$  ( $0.05/4$ ; to be significant at  $p < 0.05$ ) and the actual  $p$ -values from tests meeting this criterion are reported. A two-way (group x phase) ANOVA analyzed the IAS at 50 ft variable (not included in the MANOVA due a lack of statistical and practical correlation with the other touchdown variables), but there were no significant effects. For the MANOVA with touchdown precision and velocity, all effects were significant; there was a significant effect of group, phase, and an interaction of group and phase [group: Wilks' Lambda = 0.60,  $F(2,27) = 9.05$ ,  $p < 0.001$ ; phase: Wilks' Lambda = 0.65,  $F(2,27)$

= 7.39,  $p < 0.001$ ; interaction: Wilks' Lambda = 0.78,  $F(2,27) = 3.90$ ,  $p < 0.05$ ]. All effects were driven by velocity at touchdown.

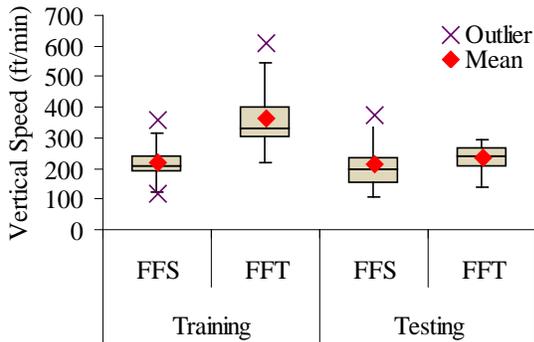
The significant group effect on touchdown velocity revealed that the FFS-trained group landed more softly (with a slower velocity) than the FFT-trained group when both training and testing data were combined [ $F(1,28) = 16.97$ ,  $p < 0.001$ ]. This difference, however, only remained significant during the training phase, and thus did not transfer [ $F(1,28) = 27.56$ ,  $p < 0.001$ ].

A significant phase effect for velocity at touchdown also emerged; overall, pilots landed more softly at testing [ $F(1,28) = 14.22$ ,  $p < 0.001$ ]. This difference, however, only held significant for the FFT-trained group [ $t(13) = 4.80$ ,  $p < 0.001$ ].

The significant interaction of group and phase on velocity at touchdown shows that group only had an effect during training, when the FFS-trained group had a lower touchdown velocity, and phase only had an effect on the FFT-trained group, who decreased their velocity at testing [ $F(1,28) = 8.06$ ,  $p < 0.01$ ]. Significant results are shown in Fig. 11 and are summarized below.

a. Summary of results for ILS touchdown

- FFS-trained pilots landed more softly than FFT-trained pilots at training, but the effect did not transfer. There were no group effects on touchdown precision or IAS at 50 ft.
- Trainees landed more softly at testing, but this effect was significant for the FFT-trained group only. There were no phase differences in touchdown precision or IAS at 50 ft.
- An interaction for touchdown velocity showed that the group effect depended on phase and the phase effect depended on group. There were no interaction effects for touchdown precision or IAS at 50 ft.



Touchdown Velocity

**Figure 11. ILS touchdown variables.** Box-plots depicting the raw data (before the Box-Cox transformation) of the ILS touchdown variables for which there were statistically significant effects

6. SSL Touchdown

The upper confidence limits and observed effect sizes for the SSL are in Table 13 (see Table 3 for variable definitions). The effects for touchdown precision ranged from 67.29 to 395.20 ft and the highest probable effect size we could have found across groups and phases was 1,135.10 ft. For vertical velocity, we found effects between 24.82 and 146.11 ft/min, with the highest probable effect of 261.70 ft/min.

**Table 13. Upper confidence bounds (95% C.I.) and raw-data mean differences by group and phase for all examined SSL touchdown variables.**

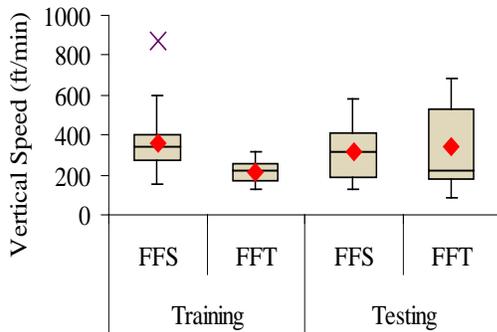
Variable	Mean Difference	Phase Difference (Trg vs. Tst)		Group Difference (FFS vs. FFT)	
		FFS	FFT	Trg	Tst
Vertical Velocity (ft/min)	largest	140.51	261.70	246.58	152.48
	observed	41.29 <sup>TR</sup>	129.60 <sup>TS</sup>	146.11 <sup>S</sup>	24.82 <sup>T</sup>
Touchdown Precision (ft)	largest	593.14	987.40	610.90	1135.10
	observed	137.64 <sup>TR</sup>	190.30 <sup>TS</sup>	67.29 <sup>T</sup>	395.20 <sup>T</sup>

Note: Mean difference is an absolute value. Superscript notations indicate the larger of the two values that make up the mean difference (<sup>TR</sup>Training; <sup>TS</sup>Testing; <sup>S</sup>FFS; <sup>T</sup>FFT). Shaded cells represent significant differences found in the main analysis using the transformed data.

The same MANOVA was run on the SSL touchdown data as on the ILS touchdown data. Again, the p-value correction for the four post-hoc tests of each type was  $p < 0.013$ ; actual p-values from tests meeting this criterion are reported. IAS at 50 ft AGL was not analyzed due to wind differences in the FFS and FFT. For the MANOVA on touchdown velocity and precision, there were no overall main effects of group or phase, but there was a trend for an interaction [Wilks' Lambda = 0.84,  $F(2,28) = 2.68$ ,  $p = 0.09$ ]. This effect was driven by velocity at touchdown. There was a trend for an interaction of group and phase on touchdown velocity [ $F(1,29) = 3.52$ ,  $p = 0.07$ ]. This effect was qualified by a significant effect of group during training only [ $F(1,28) = 11.66$ ,  $p < 0.01$ ]. The FFT-trained group landed more softly than the FFS-trained group at training, but this effect did not transfer. Thus, the interaction shows that effect of group depends on phase. These results are shown in Figure 12 and are summarized below.

a. Summary of results for SSL touchdown

- There were no overall group effects for touchdown velocity or precision.
- There were no phase differences.
- An interaction trend showed that the effect of group on touchdown velocity depended on phase. There were no interaction effects for touchdown precision.



Touchdown Velocity

**Figure 12. SSL touchdown variables.** Box-plots depicting the raw data (before the Box-Cox transformation) of the ILS touchdown variables for which there were statistically significant effects

**B. Opinion Data**

Pilots and instructors completed questionnaires immediately after each phase of the experiment, i.e., once after training and once after quasi-transfer testing. Participants were asked to relate the questionnaires only to the maneuvers flown or observed so far or since the last break.

1. Instructor Opinions

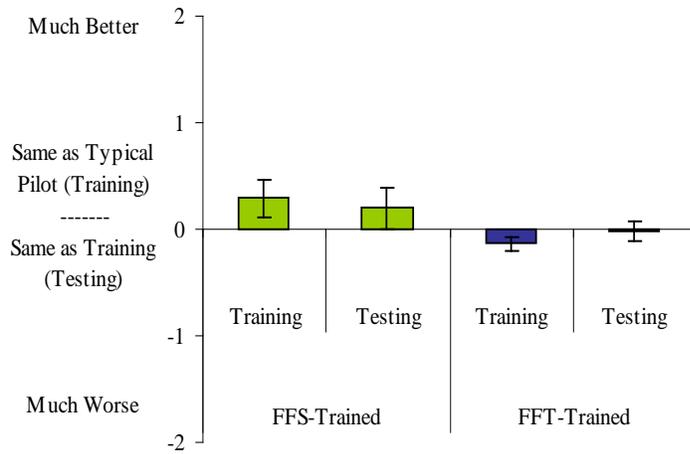
Instructors rated pilots' performance and control strategy and technique for each maneuver. They also rated pilots' ease of gaining proficiency in the simulator during training and their overall workload. Instructors also rated their

own overall comfort, in terms of a lack of nausea or simulator-induced disorientation. The questionnaires given after training asked instructors to compare the pilot flying the maneuvers in the simulator to a typical pilot flying the maneuvers in an FFS at this stage of training. After testing, instructors were asked the same questions as during training, but instead compared the pilot flying the maneuvers during testing to how he/she flew the maneuvers in the simulator at the end of training. Gaining proficiency was only rated during training since pilots would have gained proficiency at training. Most questions were asked on a scale from -2 to +2, with -2 anchored as “much worse,” 0 “the same” and +2 “much better.” In general, responses consisted of checkmarks with the opportunity to add comments.

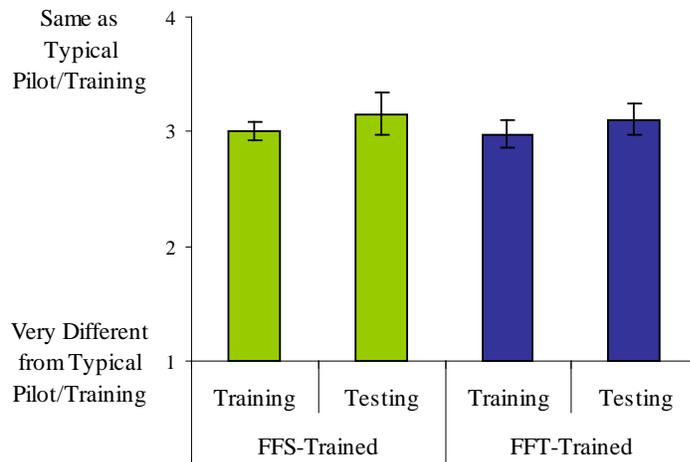
We only analyzed instructor opinions on trainees and maneuvers that were included in the simulator data analysis. Some ratings were overall ratings (across all four maneuvers), whereas others asked the instructor to rate each maneuver separately. Small sample sizes when ratings were separated by maneuver forced us to analyze the mean of the individual maneuver ratings (a composite “overall” rating) instead of analyzing each maneuver separately. The data were analyzed using one-way ANOVAs that tested for group differences at training and at testing, separately for each rating. Ratings at training were not compared to ratings at testing because the comparison was built into the questionnaires at testing, when instructors were asked to compare the pilot to him/herself at training. One-sample *t*-tests were also used to test whether the mean rating at each level of group and phase was significantly different from the best possible rating (usually 0, or “same as typical pilot/training”). Bonferroni *p*-value corrections were applied to the multiple ANOVAs and the one-sample *t*-tests. For the nine ANOVAs, the Bonferroni *p*-value criterion was  $p < 0.006$  ( $0.05/9$ ; to be significant at  $p < 0.05$ ). For the 18 one-sample *t*-tests, the *p*-value criterion was  $p < 0.003$ . In the following text, we report only the significant results, unless non-significant results warrant further explanation. We report the actual *p*-values for those tests that met the *p*-value criteria just described.

The ANOVAs revealed no group differences in instructors’ ratings of pilot performance, control strategy and technique, gaining proficiency, or pilot workload during training. There were also no differences in instructors’ own comfort in the FFS and the FFT. There were no group differences at testing, either. (We acknowledge that there appear to be effects in some of the Figures, but these were not significant with the *p*-value correction applied; e.g., pilot performance in Fig. 13 appears to depict a group difference during training, but this was non-significant with our *p*-value criterion and neither group mean was significantly different from 0.)

For the most part, instructors felt that pilots performed just as well and worked just as hard as a typical pilot would in an FFS, regardless of the training device. Once transferred, pilots also were perceived as performing and behaving no differently than they did at training. There was one exception, however, which was revealed in the one-sample *t*-tests that tested for differences from the best possible rating: for both groups of trainees, the control strategy and technique at testing was thought to be significantly different from the control strategy and technique at training [FFS-trained group:  $t(14) = -4.60$ ;  $p < 0.001$ ; FFT-trained group:  $t(16) = -6.48$   $p < 0.001$ ]. Thus, instructors believed that both groups of pilots altered their control strategy and technique once transferred. During training, both groups of pilots were rated significantly different than the rating of “same as a typical pilot” [FFS-trained group:  $t(14) = -12.55$ ,  $p < 0.001$ ; FFT-trained group:  $t(16) = -8.79$ ,  $p < 0.001$ ]. The results of all instructor ratings are depicted in Figures 13 (“performance”) and 14 (“experience”).



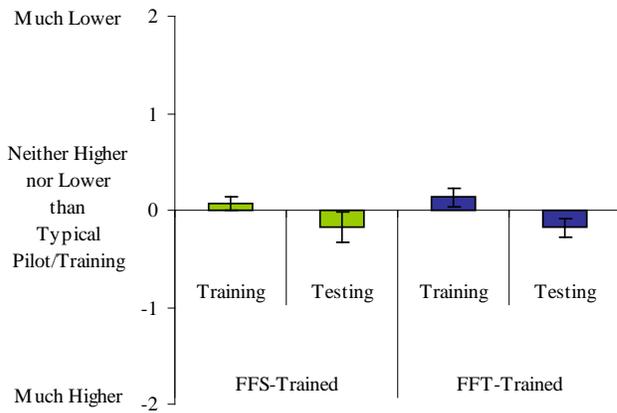
a) *Pilot Performance*



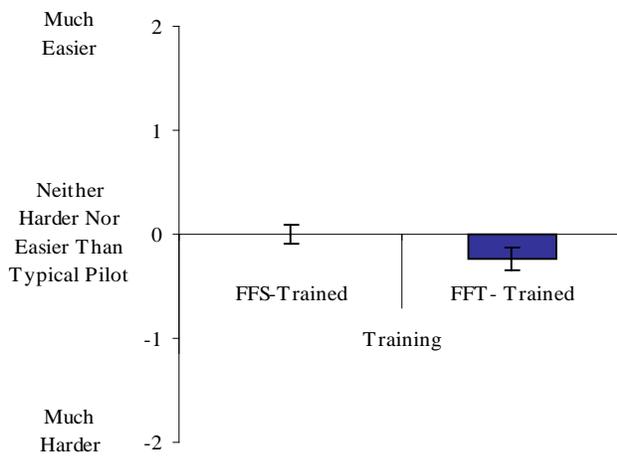
b) *Control Strategy & Technique*

**Figure 13. Instructor ratings of “performance.”**

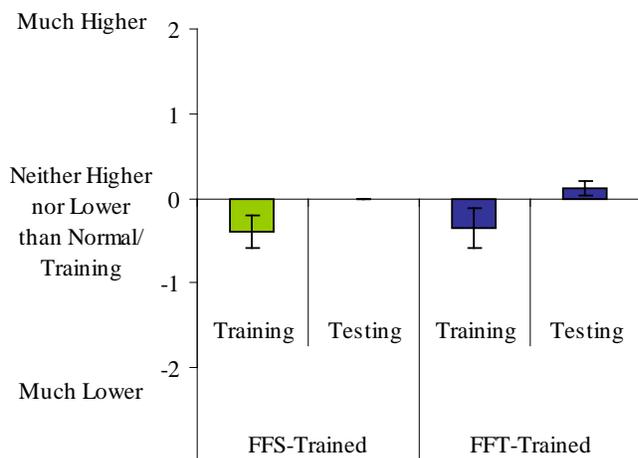
*Note: Recall that instructors made different comparisons at training and at testing (as indicated in the y-axis labels).*



a) *Workload*



b) *Gaining Proficiency*



c) *Own Comfort (lack of nausea/simulator-induced disorientation)*

**Figure 14. Instructor ratings of “experience.”**

*Note: Recall that instructors made different comparisons at training and at testing (as indicated in the y-axis labels).*

## 2. Trainee Opinions

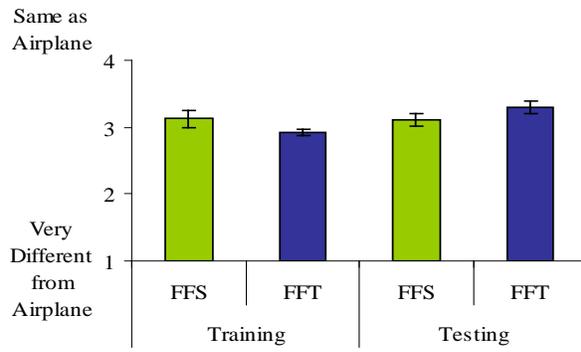
Trainees rated the handling qualities, feel and response of the simulator controls, and the control strategy and technique used in the simulator for each maneuver. Pilots also rated their overall workload and comfort, simulator acceptability, and the ease of gaining proficiency in the simulator. For all ratings, pilots were asked to compare their assessments of the simulator to what they would expect in the airplane. It was acknowledged that the pilots may not have flown the particular airplane previously or may not have experienced some of the maneuvers in the airplane, so they were asked to use their overall experience to develop an expectation of how the airplane would have “behaved.” Most questions were asked on a scale from -2 to +2, with -2 anchored as “much worse,” 0 “the same” and +2 “much better” (than the airplane). The questionnaires after training and quasi-transfer testing were identical. In general, responses consisted of checkmarks with the opportunity to add comments.

Only opinion data from pilots and maneuvers used in the simulator-data analysis were analyzed. Like the instructor ratings, some were overall ratings (across all four maneuvers), whereas others asked the trainee to rate each maneuver separately. Small sample sizes when ratings were separated by maneuver forced us to analyze the mean of the individual maneuver ratings (a composite “overall” rating) instead of doing the analyses separately for each maneuver. Trainees’ opinions were analyzed using two-way ANOVAs testing for main effects of group (FFS- or FFT-trained) and phase (training or quasi-transfer testing), as well as group and phase interactions, separately for each rating. If the ANOVA was significant, additional tests were used to clarify the effects; one-way ANOVAs tested the effect of group separately at each phase and paired t-tests tested the effect of phase separately for each group. One-sample t-tests were also used to test whether the mean rating at each level of group and phase was significantly different from the best possible rating. Bonferroni  $p$ -value corrections were applied to post-hoc tests and the one-sample t-tests. For the 14 post-hoc ANOVAs and 14 post-hoc t-tests, the Bonferroni  $p$ -value criterion was  $p < 0.004$  ( $0.05/14$ ; to be significant at  $p < 0.05$ ). For the 28 one-sample t-tests, the  $p$ -value criterion was  $p < 0.002$ . In the following text, we report only the significant results; for post-hoc and one-sample tests, we report the actual  $p$ -values for those tests that met the  $p$ -value criteria just described.

The ANOVAs for the trainee ratings yielded no group differences, across training and testing sessions, in ratings of handling qualities, feel and response of controls, control strategy and technique, workload, comfort, ease of gaining proficiency, or acceptability. There were only two phase effects found: (1) Pilots rated their control strategy and technique as more like the airplane at testing [ $F(1,40) = 5.62, p < 0.05$ ]; post-hoc t-tests qualified this effect by showing this trend only for the FFT-trained group [ $t(19) = -3.02, p = 0.007$ , adjusted  $p = 0.098$ ]. (2) There was an overall trend in the same direction for control feel and response at testing [ $F(1,40) = 3.65, p = 0.06$ ].

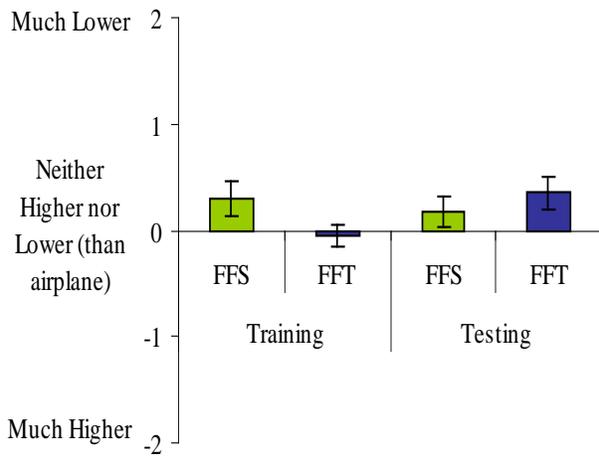
The t-tests examining differences from the “best” rating showed that, for control strategy and technique, ratings were significantly different from “same as the airplane” for both groups and across both phases [FFS-trained, training:  $t(19) = -6.87, p < 0.001$ ; FFT-trained, training:  $t(21) = -24.74, p < 0.001$ ; FFS-trained, testing:  $t(19) = -9.35, p < 0.001$ ; FFT-trained, testing:  $t(21) = -7.12, p < 0.001$ ]. Trainees also rated the acceptability to be significantly different from the best rating, or significantly less than “excellent” regardless of group and phase [FFS-trained, training:  $t(16) = -4.75, p < 0.001$ ; FFT-trained, training:  $t(20) = -9.07, p < 0.001$ ; FFS-trained, testing:  $t(17) = -4.51, p < 0.001$ ; FFT-trained, testing:  $t(19) = -4.77, p < 0.001$ ]. However both groups tended to rate the handling qualities of the FFS once transferred as higher than the best rating or, in other words, *more* satisfactory than the airplane [FFS-trained, testing:  $t(19) = 3.32, p = 0.0036$ , adjusted  $p = 0.10$ ; FFT-trained, testing:  $t(21) = 3.35, p = 0.003$ , adjusted  $p = 0.08$ ]. Workload at testing was perceived as no different from the workload expected in the airplane for both groups at both phases. The ease of gaining proficiency was significantly lower than the best rating of “very easy” during both testing and training, with no increase in ratings from either group once transferred to the FFS [FFS-trained, training:  $t(16) = -10.95, p < 0.001$ ; FFT-trained, training:  $t(20) = -9.65, p < 0.001$ ; FFS-trained, testing:  $t(17) = -17.00, p < 0.001$ ; FFT-trained, testing:  $t(19) = -11.92, p < 0.0001$ ].

In summary, trainees felt that their technique became more like in the airplane once transferred, although this effect was influenced by ratings made by the FFT-trained pilots only. All trainees thought that their technique was not the same as it would be in the airplane. There were no differences in workload or the ease of gaining proficiency, although all pilots agreed it was less easy to gain proficiency in the simulators than it would be in the airplane. Trainees rated FFS handling qualities once transfer-tested as better than the airplane, even though the simulator’s acceptability as an airplane stand-in was less than perfect at testing, and the same as it was in the FFS and the FFT during training. Figures 15-17 (“performance,” “experience,” and simulator ratings, respectively) depict the results of all trainee ratings.

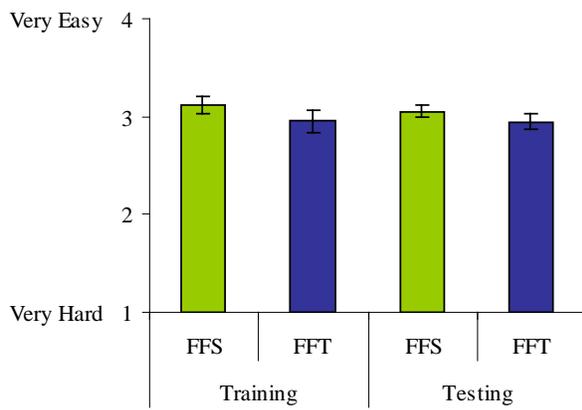


*Control Strategy & Technique*

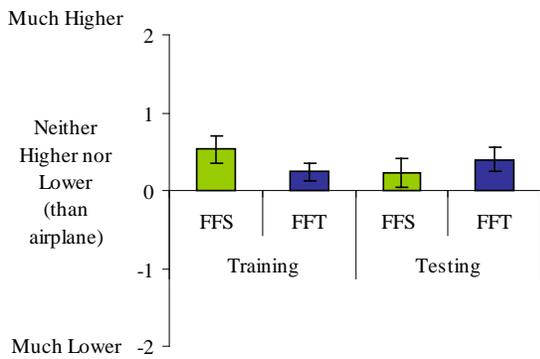
**Figure 15. Trainee ratings of “performance.”**



a) *Workload*

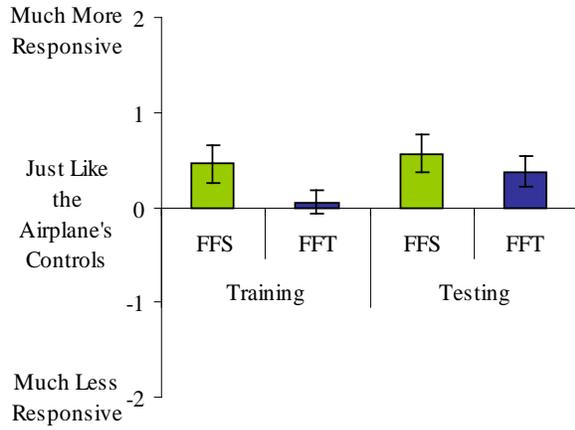


b) *Gaining Proficiency*

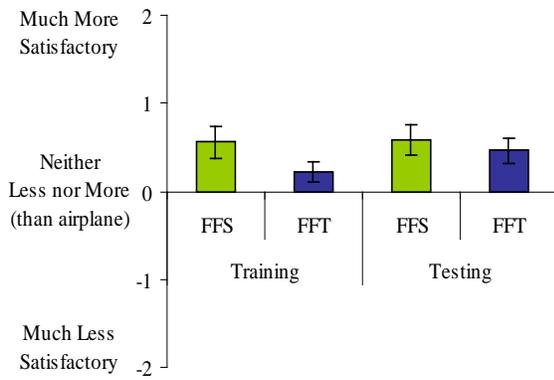


c) *Comfort (lack of nausea/simulator-induced disorientation)*

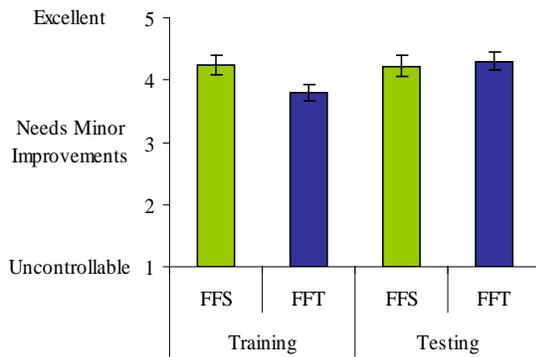
**Figure 16. Trainee ratings of “experience.”**



a) *Feel & Response of Controls*



b) *Handling Qualities*



c) *Acceptability (as a stand-in for the airplane)*  
**Figure 17. Trainee ratings of the simulator.**

## IV. Summary and Discussion

A total of 92 pilots were successfully trained in the simulator, followed by a final qualification flight in the airplane. Many of these pilots had fewer than 500-hours prior experience in any airplane and no JAR-25 experience. Of these, we compared, in the FFS, flight precision and control inputs of 20 pilots additionally trained in the FFS to those of 20 pilots additionally trained in the FFT. They were trained in the FFS or FFT and then compared in the FFS on three maneuvers identified by subject-matter experts and the literature as being diagnostic of different effects of the different motion-cueing methods provided in the two simulators. Moreover, we collected opinion data from the same 40 pilots and their instructors during training in the FFS/FFT and testing in the FFS.

### A. Effect of Alternative Motion System on Final Flight Precision and Control Inputs

#### 1. Flight Precision

What matters most for the safety of the flying public is whether the alternative motion system of the FFT reduced pilots' final flight precision. We already know that the check airmen who qualified the 34 FFT-trained pilots were obviously satisfied with the pilots' flight performance, but we will also consider the pilots' flight precision during quasi-transfer to the FFS. Once pilots transferred, there were no differences between the flight precision of the FFS-trained and the FFT-trained groups for the takeoff maneuvers; for both the  $V_1$  cut and the  $V_2$  cut, pilots showed no statistically significant differences in heading standard deviation, yaw rate, airspeed exceedance, or pitch standard deviation. Similarly, for the engine-out instrument landing with quartering head and tail winds, there were no significant differences between the two groups with respect to localizer, glideslope, airspeed exceedance, or roll activity. This was the case for both the initial approach segment, from approach fix to decision height, as well as for the landing. There were also no differences between the two groups in touchdown velocity or precision during quasi-transfer, and no differences in IAS at 50 ft AGL.

Consistent with the flight-precision simulator data, instructors and trainees perceived no differences in flight precision between the FFS- and FFT-trained groups once they transferred to the FFS. Instructors believed that the two groups of trainees performed equally and as well as a typical pilot would. Both instructors and trainees agreed that the pilots' control strategy and technique was equivalent between groups. Instructors believed that both groups achieved proficiency before transferring to the FFS, both groups doing so with the same amount of ease as a typical pilot. Both groups of trainees agreed with each other on the degree of ease with which they achieved proficiency during transfer-testing. Hence, opinions from both instructors and trainees indicate that there were no differences in how either group performed once transferred to the FFS as a stand-in for the airplane. Trainees in both groups did think that their control strategy and technique was different from what it would be in the airplane, and that their ability to gain proficiency at quasi-transfer was harder than they expected it would be in the airplane. Despite these opinions, however, instructors believed that all trainees behaved like a normal pilot would, and all trainees were successfully transferred to the airplane and obtained their type rating, so their flight precision was perceived as satisfactory.

#### 2. Control Inputs

We demonstrated that there were no differences between the two groups in terms of flight precision at quasi-transfer, but it is also important to know whether there was any indication that FFT-trained pilots had to work any harder to achieve this same precision. If so, pilots trained in the FFT might be more fatigued in the long run than the FFS-trained pilots. We can address workload in the FFT vs. the FFS by examining the pilots' control inputs. For these, there were some overall group differences (to be discussed below), but all of these differences disappeared when transferred to the FFS as a stand-in for the airplane. This was true across all maneuvers and especially for the landing, where there were no group effects at all. Thus, pilots put in the same effort to achieve the same results in both devices.

Perhaps the best measure of trainees' workload is self-report. Consistent with the above data, instructors and their trainees confirmed that the workload in the simulator at transfer-testing was normal (similar to a typical pilot and to the airplane, respectively). There were no group differences perceived by instructors or by the trainees themselves. Trainees also agreed on the ease with which they gained proficiency once transferred. Additionally, they agreed on the adequacy of the handling qualities and the responsiveness of the controls in the simulator once transferred, so neither group believed that their ability to control the simulator was any better or worse than the other. All-in-all, trainees did not perceive a difference in how they controlled the simulator or in their amount of workload, and instructors agreed that both groups of trainees worked equally hard.

## **B. Effect of Alternative Motion System on Flight Precision and Control Inputs During Training**

### *1. Flight Precision*

Although all group differences evidently disappeared at quasi-transfer, it is still pertinent to know what group differences were present during training. No group differences emerged for flight precision during the departure maneuvers, even during training. This was also true for the hand-flown one-engine ILS approach and landing. At touchdown, however, the group with only seat and visual motion landed less softly, that is, 144 ft/min faster. Again, this difference did not transfer, so it appears that full motion helps with controlling vertical speed at touchdown, but pilots do not need to be trained with full motion in order to use it.

Instructors observed no differences in pilot performance during training, with all pilots performing as a typical pilot would, regardless of the simulator they were training in. Both instructors and trainees agreed that the control strategy and technique used in the different simulators was equivalent, although different from a typical pilot and from the strategy to be expected in the airplane, respectively. Despite this, however, trainees' strategy and technique had no effect on their performance. It also had no effect on their ability to achieve proficiency during training; according to instructors, both groups equally gained proficiency as easily as a typical pilot would. Thus, opinions from both instructors and trainees agree with the simulator data, finding no differences in flight precision during training that could be attributed to the different training simulators.

### *2. Control Inputs*

There were some differences in control inputs during training only, but these all disappeared once pilots transferred. For both departure maneuvers, the FFT-trained group worked the column harder and, as a statistical trend during the  $V_1$  cut only, worked the pedal less hard than the FFS-trained group did during training, with no difference in flight precision. The FFT-trained pilots also worked the column harder and the pedal less hard than FFS-trained pilots during training of the ILS landing, but again, this had no effect on flight precision. For the ILS approach, there were no control-input group effects during training.

Despite some differences in control inputs recorded at training, neither instructors nor trainees perceived any group differences in pilot workload or in the ease of gaining proficiency. Trainees perceived the control feel and response and the handling qualities in both simulators to be no different from those expected in the airplane. Therefore, pilots may have behaved differently in the two simulators, but they did not notice it, and thus it had no effect on their workload. (Perhaps, then, the effect sizes found in the simulator-data analysis, while statistically significant, are not operationally significant; but we leave this to the reader to decide.) In both simulators, trainees thought that it was only somewhat easy to gain proficiency, even though the instructors thought they performed as a typical pilot would. This just means that the typical pilot would likely find it difficult to train to proficiency on these difficult maneuvers.

Taken together, it does not appear that the differences in control inputs matter given that the final flight performance appears to be unaffected. Moreover, these differences did not affect workload as instructors or trainees saw it. One could argue that these differences might affect the use of the alternative motion system during evaluation. However, a look at the size of the effects shows that they may indeed not be operationally relevant (at least two representatives of the airplane manufacturer, who are both pilots and instructors, did not think so), as evidenced by the lack of perceived differences in the opinion data. Second, these were trainees undergoing initial training, so any conclusion on evaluation is unwarranted. Finally, two earlier studies have investigated the issue of evaluation with recurrent pilots and found only minor differences.<sup>3,4</sup>

## **C. Flight Precision Improvement from Training to Testing**

Another question is whether the flight precision changed after transfer, and whether it mattered if the trainee transferred from the FFT to the FFS or from the FFS to the FFS; in other words, whether the so-called phase effect interacted with group.<sup>†††</sup> For the  $V_1$  cut, flight precision improved for all but the pitch variable, but only when precision data from both groups were combined; taken separately, neither group improved significantly. There was no change in  $V_2$  cut precision across phases. The ILS-approach roll rate improved only when both groups were combined (there was a trend for the FFT-trained but not the FFS-trained group) and glideslope compliance improved for the FFT-trained group only (although this interaction is attenuated by a lack of significant post-hoc effects). For the landing, localizer compliance improved for the FFS-trained group only, but the FFT-trained group already flew very precisely during training. Also during the landing, touchdown vertical speed improved for the FFT-trained group only. This is the one improvement that appears to be affected not only by the training process, but also by the change in device. For all other variables, improvement did not seem to depend on the type of device the pilots were

---

<sup>†††</sup>Because the earlier discussion has shown that there were only few behavioral differences without a clear trend of an effect of motion system on workload, we will not further discuss control inputs.

trained in. Trainees and instructors agreed, as training group did not interact with overall perceptions of improvement from training to testing. The one finding in the opinion data that does suggest some phase differences is that both groups of pilots – those trained the FFS and those trained in the FFT – rated the simulator handling qualities as better than the ones presumed in the airplane, but only when they transferred to the FFS. This may simply have been an expression of their increased familiarity with flying the difficult tasks.

#### **D. Differences between Groups for Pooled Training and Testing**

So far, we have only looked at the effect of motion system separately for training and transfer testing. When we look for overall group differences that might emerge when training and transfer testing are combined, we might find that one group performed better than the other. For flight precision, there were no overall group effects. For control behavior, however, there were a few effects that were only apparent when the results of both phases were combined. There were significant overall group differences for takeoff maneuvers when all control-input types were combined, but no one control-input type in itself yielded an overall effect. For the ILS approach, FFS-trained pilots used less pedal when the two phases were combined, and, as a trend only, less wheel than the FFT-trained pilots. For the ILS landing, however, there were no overall differences between the two devices. Taken together, this is further evidence that the alternative motion-stimulation system of the FFT does not affect workload negatively. Although FFT-trained pilots tended to work more with the pedal and wheel during the approach than the FFS-trained pilots, any hint of workload difference between groups disappeared during landing. Importantly, there were also no overall group differences in impressions of workload or resulting performance from instructors or from trainees.

#### **V. Conclusion**

In three previous studies, we investigated the transfer (to the FFS as a stand-in for the airplane) of training in the FFS with visual and physical motion vs. the same FFS with visual motion only. Across all three studies, there were no operationally-relevant group differences that transferred to the simulator as a stand-in for the airplane. Our first study,<sup>4</sup> which tested recurrent pilots in a Level C FFS of a turboprop “powerhouse” airplane with wing-mounted engines, investigated engine-failures with rejected and continued takeoffs. All we found was very few and very small control-input differences, and none of them affected control precision. We found, however, that especially the sway motion of the simulator used in the study was rather benign, although a follow-up study comparing it to eight other simulators found that it was quite typical for other FAA-qualified Level C and D simulators. We therefore undertook a second study<sup>3</sup> in the FAA-NASA Level D B747-400 simulator, adjusted to provide motion cues best suited for the test maneuvers. This study included also hand-flown one-engine-out approach and landing maneuvers with weather similar to the present study. For the ILS approach, all group differences transferred. The only potentially operationally relevant difference, according to our subject-matter-experts, was that the motion-trained group had worse horizontal directional control than the no-motion trained group during ILS1 (0.19 dots higher localizer STD and 0.16 dots higher localizer exceedance), however, this effect is not reflected in the current study. All other effects were very small and also did not reflect anything found in the current study. Interestingly for the current study, for the SSL, the motion group had lower pedal bandwidth and landed more softly (by 42 ft/min) but less precisely (by 225 ft, but both groups landed well within the landing box) than the group with visual motion only, and unlike in the current study, where this effect occurred only with the ILS, this effect did transfer. None of the small differences found with the  $V_1$  cut transferred, but one of them is very interesting: Unlike in the current study with seat motion, we found a  $V_1$  cut pedal reaction time disadvantage without any motion that may have arisen from the lack of vestibular cues. It was robust, but it was less than half a second (0.39), had minimal effect on flight precision (0.76 deg lower heading standard deviation), and, most importantly, did not transfer to the simulator with motion. During the  $V_2$  cut, interestingly, where the motion alert should be more important due to the lack of visual cues, the motion-trained pilots had a higher pedal reaction time at transfer than the group trained without physical motion. This effect was also not replicated in the current study.

Our third study,<sup>2</sup> which is probably most relevant to the current work because it tested initial pilots, tested new hires in a Level D B717-200 and results mostly confirmed those of the second study. During training for the  $V_1$  cut, we found a trend for a faster pedal response with motion (again, less than half a second), without affecting flight precision. Also, the visual-motion-only trained pilots immediately caught up once exposed to physical motion at transfer.

In conclusion, like our previous work, the current study does not seem to support an operational benefit of FFS motion. Not only did most group effects disappear in this and previous studies at transfer, but the effect sizes in the previous studies were determined by experts to be small and operationally irrelevant. We leave it up to the readers to interpret the effect sizes in the current study, but regardless of the sizes, the current results are mostly in-line with previous results in that there was a general trend for there to be no differences between the training success of FFS-

trained and FFT-trained pilots. Most importantly, there were no group differences at transfer-testing in this and, for the most part, previous studies. Perhaps the most interesting motion-related difference between this study and previous studies, however, is the lack of a difference in pedal reaction time to the  $V_1$  cut in the current study. This indicates that the one-degree-freedom seat-motion system, although essentially different from the hexapod platform motion system of an FFS, may provide motion cues to the same *qualitative* level as an FFS, based on the flight performance data presented here.

### Acknowledgments

The authors thank Dr. Eleana Edens from the FAA Air Traffic Organization Human Factors Research and Engineering Division and Dr. Thomas Longridge from the FAA Flight Standards Service Voluntary Safety Programs for support of this work. We thank the NAA and the airplane and the simulator manufacturers for their vision in developing the FFT and the training center personnel for their considerable efforts and expertise in conducting this study. We also thank Dr. Mary Stearns, Chief of the Volpe Center's Behavioral Safety Research and Demonstration Division, for her comments. Also from the Volpe Center, we thank Dr. Michael Zuschlag, a senior researcher in the Human Factors/Behavioral Safety Research and Demonstration Division, for sharing his statistical expertise; Miles Bellman, Aerospace Engineer in the Navigation and Surveillance Division, and Laura Birmingham, also of the Behavioral Safety R&D Division, for their technical support.

### References

- <sup>1</sup>Bürki-Cohen, J., Sparko, A. L., and Go, T. H., "Training Value of a Fixed-base Flight Simulator with a Dynamic Seat," *AIAA Modeling and Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics, AIAA-2007-6564, 2007.
- <sup>2</sup>Bürki-Cohen, J. and Go, T.H., "The Effect of Simulator Motion Cues on Initial Training of Airline Pilots," *AIAA Modeling and Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics, AIAA-2005-6109, 2005.
- <sup>3</sup>Go, T. H., Bürki-Cohen, J., Chung, W. W., Schroeder, J., Saillant, G., Jacobs, S., and Longridge, T., "The Effects of Enhanced Hexapod Motion on Airline Pilot Recurrent Training and Evaluation," *AIAA Modeling and Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics, AIAA-2003-5678, 2003.
- <sup>4</sup>Bürki-Cohen, J., Go, and T.H., Longridge, T., "Flight Simulator Fidelity Considerations for Total Airline Training and Evaluation," *AIAA Modeling and Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics, AIAA-2001-4425, 2001.
- <sup>5</sup>White, A. D., "The Impact of Cue Fidelity on Pilot Behavior and Performance," FDS Dept., Defence Research Agency, Bedford, UK, 1994.
- <sup>6</sup>Schroeder, J., Chung, W., Tran, D., and Laforce, S., "Pilot-Induced Oscillation Prediction with Three Levels of Simulation Motion Displacement," *AIAA Atmospheric Flight Mechanics Conference*, American Institute of Aeronautics and Astronautics, AIAA-98-4333, 1998.
- <sup>7</sup>Bjorkman, E. A., "Flight Test Evaluation of Techniques to Predict Pilot Induced Oscillations," Masters Thesis, AFIT/GAE/AA/86J-1, Air Force Institute of Technology, WPAFB, OH, 1986.