

III

“Correlating Automated and Flight Instructor Assessments of Straight-In Landing Approaches by Novice Pilots on a Flight Simulator” by Bruce E. Heath, M. Javed Khan, Marcia Rossi, Syed Firasat Ali, Proceedings, Huntsville Simulation Conference, Huntsville, Alabama, October 2005, 8 pages.

Blank Page

Correlating Computed and Flight Instructor Assessments of Straight-In Landing Approaches by Novice Pilots on a Flight Simulator

Bruce E. Heath (bheath35@hotmail.com), M. Javed Khan (mjkhan@tuskegee.edu),
Marcia Rossi (rossi@tuskegee.edu), Syed Firasat Ali (sfali_tusk@yahoo.com)
Tuskegee University, Tuskegee, AL 36088

The rising cost of flight training and the low cost of powerful computers have resulted in increasing use of PC-based flight simulators. This has prompted FAA standards regulating such use and allowing aspects of training on simulators meeting these standards to be substituted for flight time. However, the FAA regulations require an authorized flight instructor as part of the training environment. Thus, while costs associated with flight time have been reduced, the cost associated with the need for a flight instructor still remains. The obvious area of research, therefore, has been to develop intelligent simulators. However, the two main challenges of such attempts have been training strategies and assessment. The research reported in this paper was conducted to evaluate various performance metrics of a straight-in landing approach by 33 novice pilots flying a light single engine aircraft simulation. These metrics were compared to assessments of these flights by two flight instructors to establish a correlation between the two techniques in an attempt to determine a composite performance metric for this flight maneuver.

This research was supported by NASA Ames Award CTS-0320305 and NASA Dryden Grant NAG4-226.

INTRODUCTION

With the advent of faster computers, it has become possible and cost effective to use personal computer-based aviation training devices (PCATD) to provide supplemental training for pilots. Such devices are regulated by the Federal Aviation Administration (FAA). The airplane or flight training device (FTD) flight-hour training time required for an instrument rating may be reduced by using PCATDs that have been determined to meet acceptable FAA standards. For the student pilot in a Part 61 or Part 141 flight school, 10 of the total allowable flight-hours in a FTD can be flown in a PCATD. However, the regulations state that a certified flight instructor (CFI) must also be present (FAA 1997). Vruels and Obermayer (1985)

emphasize the need of automated performance measures on modern simulators as a substitute for direct evaluation of performance by an instructor. One strategy to address this goal is to determine how closely a computer can emulate a CFI's evaluation of a student pilot. Therefore, an investigation of different computed evaluation techniques is needed to compare against instructor evaluations.

Rantanen and Talleur (2001) have suggested the following five primary measures for pilot performance from the flight data recorder for nine flight parameters: (1) Standard deviation. A small standard deviation is indicative of good performance. (2) Root mean square error (RMSE) which summarizes the overall error. (3) Number of deviations, which is a measure of

occurrences of the aircraft staying outside the prescribed tolerances. (4) Time outside tolerance is the cumulative time the aircraft spends outside prescribed tolerances. (5) Mean time to exceed tolerance is computed from the rate of change between successive data points and the aircraft's position relative to a given tolerance. Heath and Crier (2003) used computerized scoring and certified flight instructors (CFI) to attempt to devise a computed performance measure for a level turn. A performance index based on deviations from the required heading, altitude and airspeed was compared with the instructors' letter grades. One of the challenges of such comparisons is inter-rater reliability of the flight instructors.

The objective of the current study was to explore the possibility of determining computed measures for a landing task by novice pilots on a flight simulator which would realistically depict their performance. The approach there has been that if these performance measures correlate with assessments of the task by CFIs that these measures then can be used routinely to inform student pilots of their performance. This study chose to use the standard deviations of the errors in flight parameters to represent pilot performance. A straight-in landing approach was chosen as the maneuver to be studied.

METHOD

Equipment

The experiment was performed at the Aerospace Science Engineering Department, Tuskegee University. The flight simulation environment consisted of four computers that were all running Microsoft Flight Simulator 2002 (MSFS 2002). The computers were connected by LAN using a Netgear Fast Ethernet Switch. Three of the

computers were used for out of the window (OTW) views. One computer was used as the main computer. Communication between the computers was made using FSUIPC; a third party software that runs over IPX/SPX. The OTW view was created using Wideview software; also third party. Data was collected into Excel spreadsheets using a third party flight data recorder. All of the computers run Windows XP Professional. The OTW view computers are all Gateway E series computers with the following specifications: Pentium 4 2.60GHz processor Hyper-Threading Technology with 512KB cache, 512 MB ram, bus speed 800 MHz and memory speed of 333 MHz. An NVIDIA GeForce FX5200 is the video card used for the OTW view computers. The master computer has the following specifications: Pentium 4 2.4GHz processor with 512 KB cache, 512 MB ram, bus speed 133 MHz. The video card for the master computer is a NVIDIA GeForce Ti 4600 with 128MB memory. A Thrustmaster HOTAS joystick, a Cougar throttle and CH Products rudder pedals were used.

Procedure

Three CFIs were asked to individually view and evaluate 33 video recordings of straight-in landing approaches made by novice pilots on the flight simulator. The pilots were students at Tuskegee University who had had little or no experience flying an airplane or simulator. Two of the flight instructors also had airline transport pilot (ATP) certification with one CFI having 7100+ flying hours with 580+ instructional hours while the other CFI had 5800+ flying hours with 4000+ instructional hours. The third CFI had extensive experience in certification requirements and flight testing and flight simulation.

The task given to the pilots was to maintain 75 knots airspeed, descend at 500 feet per minute, and head towards and line up with the centerline in a straight-in approach to the Bremerton International Airport in the Seattle, WA area. The aircraft being flown was a Cessna 172. The flight scenario was a modified version of a landing lesson from the “flight school” portion of Microsoft Flight Simulator, MSFS 2002. The approaches made by the pilots were recorded and sent to CFIs for evaluation. The instructors were asked only to grade the approach, and not the flare and touchdown. Further, they were told to turn the sound off of the recording, and to disregard any text messages that may have appeared on the screen. They were asked to evaluate the flights after the ‘autopilot’ light was turned off. The CFIs were asked to assign each flight a grade based on a 100 point scale. They were also asked to monitor three flight parameters namely, airspeed, rate of descent and runway alignment and record the best and worst controlled flight parameters and the flight parameter most affecting their grading.

The various flight parameters including airspeed, heading and altitude were written to an Excel file with a 1HZ frequency for each flight. Also certain other events such as ‘autopilot status’, ‘aircraft on ground’ were also written to the file. The data for each flight was then processed to determine the error in airspeed, rate of descent and runway alignment. The standard deviations for these errors were then calculated.

RESULTS & DISCUSSION

The performance measures investigated were the standard deviation of the errors in runway alignment, rate of descent, airspeed and linear combinations of these deviations.

Pearson Product Moment Correlations were calculated between these measures and the individual instructor scores. Weak correlations ($r^2 = 0.263, 0.09, 0.12$ for rate of descent error, $r^2 = 0.08, 0.1, 0.05$ for runway alignment error, $r^2 = 0.14, 0.19, 0.1$ for airspeed error) prompted a detailed analysis of the instructors’ grading.

Correlations between the pairs of instructors grades were calculated and found to be only moderate ($r^2 = 0.4334, 0.5054, 0.4334$). Therefore, the instructor responses were analyzed to determine whether there was consistency in scoring procedures not captured in the grades. The frequencies of citing the best and worst controlled parameters are shown in Figs. 1(a) and 1(b). As can be seen, all three instructors most frequently cited runway alignment as the best controlled parameter while citing rate of descent as the worst controlled parameter. In addition, one instructor also cited airspeed almost as frequently as the other two parameters as the worst controlled parameter. This suggested that a composite of the standard deviations of runway alignment error, rate of descent error and airspeed error to be a candidate error metric.

Seven flights were identified for which all three instructors agreed on both best and worst controlled parameters. Correlations between the instructor grades and the composite error metric (a linear combination of the standard deviations of errors in runway alignment, rate of descent and airspeed) were calculated for these flights (Fig. 2). A strong and significant correlation was found for all three instructors ($r^2 = 0.8153, t = 4.6878, p < 0.01, r^2 = 0.7749, t = 4.14877, p < 0.01, r^2 = 0.628, t = 2.91, p < 0.05$). Correlations between pairs of instructor grades were also calculated for these seven flights (Fig. 3). It can be seen that despite their agreement on the best and

worst controlled parameters, one of the flight instructor's grades was weak to moderately correlated with the other two ($r^2 = 0.8896, 0.44, 0.2517$).

In conclusion the following observations are noted:

- (a) Inconsistency in grading between instructors is a probable cause for weak correlations between instructor grading and the error metric.
- (b) Despite agreement on best and worst controlled parameters, the weak to moderate correlation of one instructor's grades with the others suggests that the instructors' grades may be influenced by other factors.

FUTURE RESEARCH

- Study the effect of student pilot experience on inter-CFI correlation.
- Evaluate the influence of student pilot experience on correlation of computed performance measures with CFI-grades.
- Investigate other performance measures to determine most robust measure.

REFERENCES

1. Brannick, M. T., Prince, C. and Sala, E. (2002). The Reliability of Instructor Evaluations of Crew Performance: Good News and Not So Good News, *The International Journal of Aviation Psychology*, 12(3), 241–261.
2. Heath, B. and Crier, T. (2003). Comparison of different methods of grading a level turn task on a flight simulator, *Proceedings of Huntsville Simulation Conference*, October 2003.
3. Rantanen, E. M., & Talleur, D. A. (2001). Measurement of pilot performance during instrument flight using flight data recorders. *International Journal of Aviation Research and Development*, 1(2), 89-102.
4. Vruels, D. and Obermayer, R.W. (1985). Human-system performance measurement in training simulators. *Human Factors*, 27(3), 241-250.

Worst Controlled Parameter

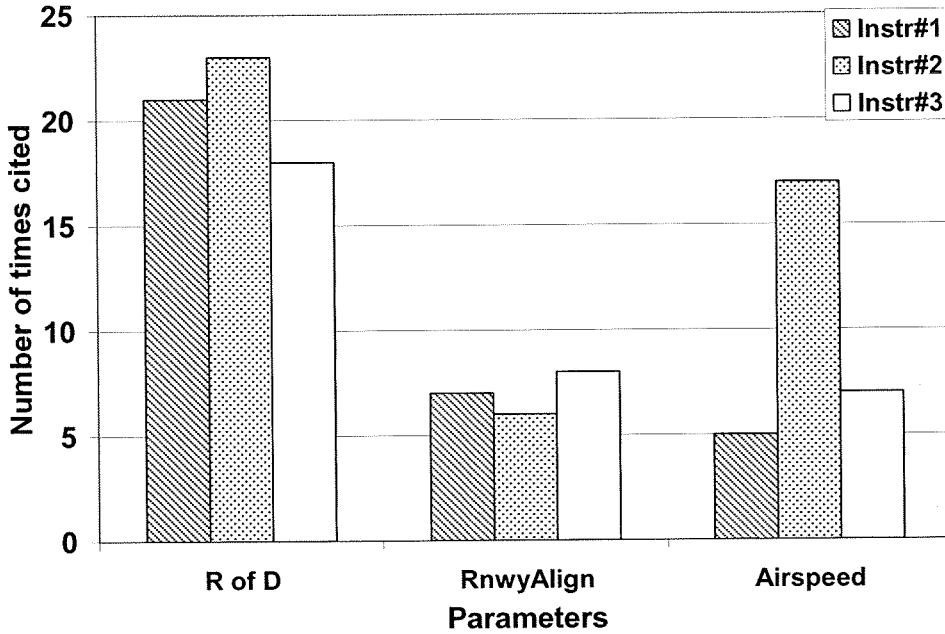


Figure 1(a): Number of Citations of Worst Controlled Parameter

Best Controlled Parameter

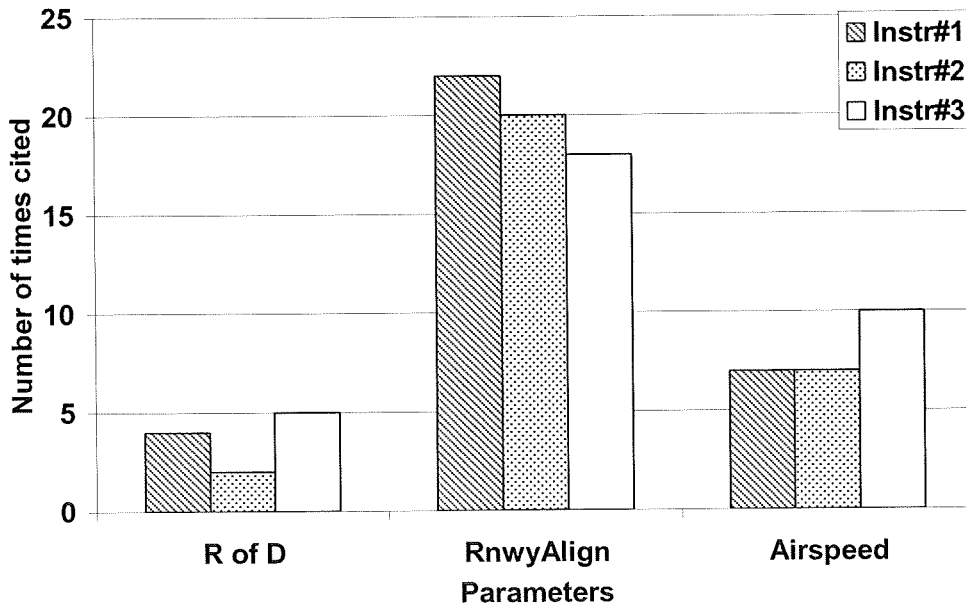


Figure 1(b): Number of Citations of Best Controlled Parameter

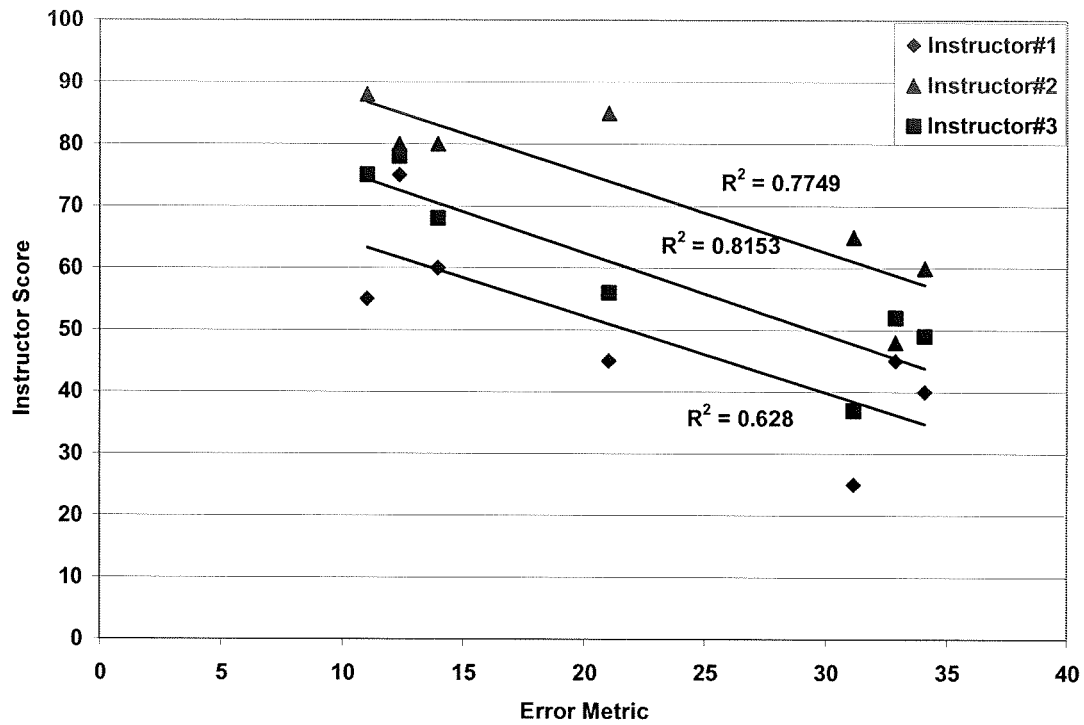


Figure 2: Correlation between the Composite Error Metric and Instructors' Scores

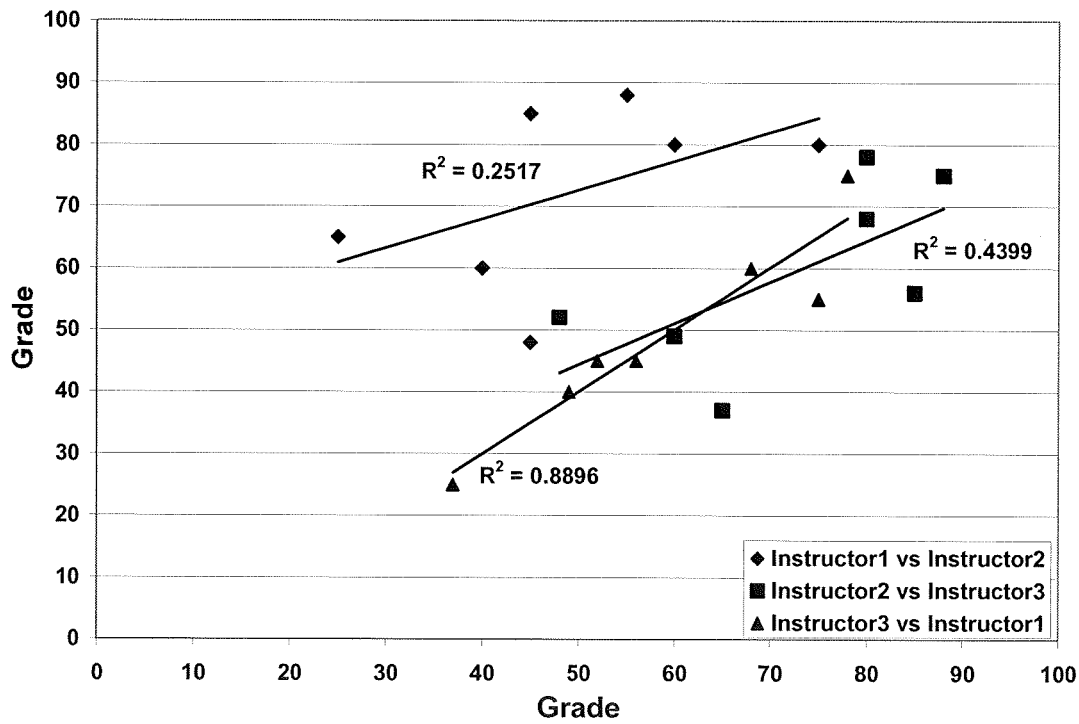


Figure 3: Correlation between Flight Instructors' Scores