Flying the Needles: Flight Deck Automation Erodes Fine-Motor Flying Skills among Airline Pilots

Andreas Haslbeck, Technische Universität München, Munich, Germany, **Hans-Juergen Hoermann**, German Aerospace Center, Hamburg, Germany

Précis: This study evaluated the influence of practice and training on fine-motor flying skills. 126 randomly selected airline pilots had to perform a manual raw data precision approach. All results indicate that flight practice is a significantly stronger predictor for manual flying performance than the time since flight school or flight experience.

Running Head: Flying the Needles

Empirical article.

Exact word count text: 4.653.

Exact word count references: 977.

Acknowledgements

This work was funded by the German Federal Ministry of Economics and Technology via the Project Management Agency for Aeronautics Research (DLR) within the Federal Aeronautical Research Program (LuFo IV-2). The authors acknowledge the support of Patrick Gontar and Ekkehart Schubert in both flight simulator experiments.

Address correspondence to Andreas Haslbeck, e-mail: haslbeck@tum.de

Objective: The aim of this study was to evaluate the influence of practice and training on fine-motor flying skills during a manual ILS approach.

Background: There is an ongoing debate that manual flying skills of long-haul crews suffer from a lack of flight practice due to conducting only a few flights per month and the intensive use of automation. However, objective evidence is rare.

Method: 126 randomly selected airline pilots had to perform a manual flight scenario with a raw data precision approach. Pilots were assigned to four equal groups according to their level of practice and training, by Fleet (short-haul, long-haul) and Rank (first officer, captain).

Results: Average ILS deviation scores differed significantly in relation to the group assignments. The strongest predictor variable was fleet, indicating degraded performance among long-haul pilots.

Conclusion: Manual flying skills are subject to erosion due to a lack of practice on long-haul fleets: all results support the conclusion that recent flight practice is a significantly stronger predictor for fine-motor flying performance than the time period since flight school or even the total or type-specific flight experience.

Application: Long-haul crews have to be supported in a timely manner by adequate training tailored to address manual skills or by operational provisions like mixed-fleet flying or more frequent transitions between short-haul and long-haul operation.

Keywords: skilled performance, automation, perceptual-motor performance, manual controls, information processing

In his classical book about pilots' stick and rudder skills Wolfgang Langewiesche (1944) explained that for learning the *art of flying an aircraft* the pilot sometimes needs to withstand his or her natural responses. For example, in a stall situation at low altitude the correct recovery requires to push the stick forward and to point the aircraft's nose to the ground. A very strong skill is required to hold back powerful instinctive behaviors of pulling the stick backwards in this situation. Hard and continuous drill is indispensable for pilots to acquire and maintain the adequate touch and feel which is essential to manually control the aircraft in any conceivable maneuver. However, in todays advanced technology aircraft pilots are often lacking sufficient opportunities to practice when they are relieved too often from manual flying tasks by using automated systems (cf. SKYbrary, 2016a).

Manual control implies lateral (roll, heading), vertical (pitch, altitude, vertical speed), and longitudinal (airspeed) control of an aircraft (Puentes, 2011) mainly through adequate fine-motor inputs by the human pilot to a control yoke (Boeing types) or a sidestick (Airbus types) governing an aircraft's pitch and roll, and the thrust levers. Yaw control by rudder pedals is a minor task performed in normal operation only momentarily during takeoff, the landing flare, and the deceleration after touch down. In other words, manual control means hand flying by reference to raw data without highly automated systems like flight director, autopilot, autothrust, or other flight management systems (SKYbrary, 2016a). *Raw data* flying specifies the absence of the flight director (Casner, Geven, Recker, & Schooler, 2014). Under this basic but challenging condition the pilot performs a compensatory tracking task and in parallel cognitively processes information about speed, altitude, and the flightpath. This task requires adequate knowledge and skills to control the dynamics of the aircraft to actively follow the intended trajectory. In addition the localizer and glideslope indicators are amongst the most important information in the case of an instrument landing. For Airbus aircraft, manual flying is still supported by *envelope protections* in *normal law* mode.

Previous Work investigating Pilots' Manual Flying Skills

Increased flight deck automation could reduce the opportunity for flight deck crews to practice their manual flying skills, and therefore, could degrade their levels of performance. Early warnings were raised by Wiener and Curry (1980) even prior to having broader data-sets to examine this anticipated threat. Childs and Spears (1986) addressed perceptual and cognitive aspects of manual flying and its degradation, while Sarter and Woods (1994) reported few deficits in pilots' proficiency in standard tasks like aborting a takeoff or disengaging the approach mode. Veillette (1995) showed a significant influence of automation on manual flying skills. More recent experimental studies were performed by Gillen (2008), showing that pilots performed below licensing standards, and Ebbatson (2009) evaluated the effect of degradation on manual skills due to a lack of opportunities to practice among short-haul crews. In one of the latest studies, Casner et al. (2014) observed pilots having difficulties in cognitive tasks corresponding to manual flight. Quite recently though, aviation regulatory authorities have raised common concerns about the deterioration of basic manual flying skills among pilots flying highly automated aircraft and recommended some preventive actions (Federal Aviation Administration, 2013a; European Aviation Safety Agency, 2013; Civil Aviation Authority, 2014). On one side, there is a high level of agreement among pilot and training communities as well as manufacturers that manual and cognitive flying skills tend to decline because of a lack of practice due to increased use of automated systems. On the other side, as Civil Aviation Authority (2014, p150) criticizes, scientific findings are still inconclusive as to which degree such decline occurs because evidence is often based on pilots' opinions and experiences or an analysis of narratives.

Evidence from Accident Statistics

With respect to long-haul crews there is only anecdotal evidence that they suffer from an absence of practice opportunities, resulting in lower manual skills (Drappier, 2008; Learmount, 2011; Civil Aviation Authority, 2014). Aviation accidents with clear indications of a lack of manual flying skills, like the prominent Air France flight 447 (Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, 2012) and Asiana Airlines flight 218 (National Transportation Safety Board, 2014), are too rare to provide statistical evidence. Dismukes, Berman, and Loukopoulos (2007) analyzed 19 predominantly short-haul aviation accidents in the U.S. and found in eight cases clear evidence for insufficient manual flying skills at least as a contributing factor. Lacking manual skills are also involved in many upset and loss-of-control accidents (Lambregts, Nesemeier, Wilborn, & Newman, 2008; Newman, 2012). The International Air Transport Association (IATA) published 2015 an in-depth analysis of 415 accidents in commercial aviation, which occurred worldwide between 2010 and 2014. In 26% of these accidents (mostly landing accidents, but also loss of control in flight) IATA found tangible evidence that manual handling flight crew errors were involved (p.29). As one of the recommendations to operators IATA concludes that “Stable approaches are the first defense against runway excursions. The final, more important, defense is landing in the touchdown zone” (International Air Transport Association, 2015, p.77).

Degradation of Fine-Motor Flying Skills

While manual flying could be considered a rather simple tracking task in theory, pilots need regular practice and training to maintain this distinct set of fine-motor skills. Short-haul and long-haul operations support the maintenance of manual flying skills differently: for the former, eight to twelve duty cycles per month with up to four legs each is typical and for the latter, three to four long-range flights are performed monthly due to legal rest periods. Thus, both types of operation lead to different levels of practice in pilots’ manual skills. None of the abovementioned studies directly addressed the influence of practice and training on fine-motor skills. Based on the reviewed literature, we identified a research gap concerning a valid, holistic, and comparative evaluation of pilots’ manual flight proficiency: (1) regarding different types of operations – long-haul and short-haul, (2) different levels of experience, responsibility, and tasks – captain (CPT) and first officer (FO), (3) under the recent amount of exposure to automation in today’s advanced technology aircraft, and (4) in a realistic flying scenario familiar to the pilots. In our work we are dealing with pilots who have different levels of practice, training, and experience while facing the same flying tasks within identical limits of licensing standards (European Union, 2011).

Research Questions and Hypotheses

The main research question of this paper is: what influence does the *level of practice and training* have on fine-motor flying skills? The *level of practice and training* is not a single and measurable metric or unit, but rather a concept concerning flight proficiency that includes several influences: (1) elapsed time since initial flight training, addressing the long-term skill degradation (c.f. Ebbatson, 2009; Franks, Hay, & Mavin, 2014); (2) daily flight practice, addressing on-the-job training (Fleishman, 1966; Savion-Lemieux & Penhune, 2005); and (3) the influence of flight simulator sessions, when periodically selected flying tasks and maneuvers are to be practiced (recurrent training) and tested (proficiency checks) under the supervision of trainers and examiners (Buckley & Caple, 2009). It was hypothesized that with a higher level of practice and training, pilots show better fine-motor flight performance (Haslbeck, Kirchner, Schubert, & Bengler, 2014). Referring to the *level of practice and training*, a secondary research question arises: whether the time dated back to flight school and expertise (determined by Rank) or the daily flight practice (determined by Fleet) has a stronger influence on manual flight proficiency. If the former aspect prevails, first officers would perform better, because their elapsed time since flight school is shorter; in the latter case, short-haul crews would perform better, because they have more daily flight practice. Ebbatson (2009) and Franks et al. (2014) have argued that (initial) training long ago cannot sufficiently support recent manual flying skills. Addressing these research questions, we expect to see a stronger effect from the daily flight practice and a slightly weaker effect from the time period since flight school, which consequently leads to the expected order of manual flight performance: FO short-haul > CPT short-haul > FO long-haul > CPT long-haul.

METHOD

This paper reports on the analysis of the fine-motor flight performance of airline pilots derived from two consecutive flight simulator studies. Both studies were funded by a German research program in cooperation with a major European airline. Experiment A took place in 2011, comparing manual flight performance of two groups of pilots: FOs scheduled on short-haul service, representing a high level of practice and training, as well as CPTs scheduled for long-haul operation, representing a low level of practice and training. To complement experiment A with the two missing groups (i.e. CPTs on short-haul and FOs on long-haul), experiment B was conducted in 2013 with CPTs scheduled for short-haul service and FOs scheduled for long-haul service.

Apparatus

Airbus types were selected for two reasons: first, the fly-by-wire technology designed under the commonality principle (Vadrot & Aubry, 1994) ensures very similar handling characteristics and second, being equipped with second generation electrical flight control systems they expose pilots to high levels of automation (Brière and Traverse, 1993). Thus, we conducted both experiments in a southern German flight simulator training center equipped with two Airbus-type full-flight simulators (FFS Level D): one Airbus A320 device and another one in an Airbus A340-600 configuration.

Scenario and Instruction

Prior to both experiments, all participants completed another simulated flight scenario concerning operational problems (35 min) and simultaneously warmed up for the manual flying task. The flight scenario (10 min) for our study was a manual approach to Munich Airport (26R EDDM) and was the same for all participants. It started shortly before a defect in the autopilot and the flight director occurred; thus, lateral and vertical control of the aircraft in an instrument landing system (ILS) approach had to be performed manually based on raw data. Pilots were, however, allowed to use autothrust for the longitudinal control, but very few did; auto-trim was engaged. The weather was set to the following parameters: visibility 1,200 m, wind 220°/17-22 kts gusty, ceiling 270 ft, light rain.

All pilots were instructed to perform a landing as accurately as possible according to company standard operating procedures and ATP licensing standards. Participants had the role of pilot flying (PF) and were supported by a pilot monitoring who was either a confederate pilot (experiment A) or the second participating crewmember (experiment B).

Participants

All participants were randomly selected by the crew scheduling department of the cooperating airline, and occupied the same seat as well as the same aircraft type for which they were rated. Four groups (stratified random sample) of ATP licensed airline pilots, about 30 pilots per group, were scheduled as the PF in this experiment: FOs and CPTs on Airbus A320 as well as Airbus A340. All participating pilots experience routinely four annual 4-hours flight simulator training sessions: two recurrent training sessions and two legal licensing checks. Table 1 shows relevant demographical data for the 126 participants. The overall flight hours are a general measure of flying experience. Landings within the last 30 days prior to the experiment account for short-term practice, and time period since flight school indicates long-term skill retention.

Table 1: Demographic data for participants

Rank	Fleet	N	Age		Flight hours: overall/on type		Landings in past 30 days		Years since flight school	
			mean	SD	mean	SD	mean	SD	mean	SD
FO	A320	39	30.1	2.8	3,438/2,415	1,848/1,266	16.1	6.3	5.8	2.8
FO	A340	28	36.4	3.3	7,204/3,469	1,987/1,812	2.4	1.5	12.2	2.9
CPT	A320	30	43.0	4.3	11,276/3,847	1,931/2,355	16.6	10.2	18.1	3.3
CPT	A340	29	49.8	3.6	14,969/2,909	2,951/1,818	3.5	2.1	24.4	4.1

Note. ILS deviation data evaluation is based on these 126 participants; TD points were calculated for these and four more CPTs on A340.

Dependent Measures

This is an Accepted Manuscript of an article published by Sage Publishing in Human Factors on 13 Apr 2016, available online:http://hfs.sagepub.com/cgi/pmidlookup?view=long&pmid=27076096.

Deviations from ideal flight performance and landing parameters (cf. International Air Transport Association, 2015, p77) were recorded and analyzed. ILS flightpath deflections were recorded in two different dimensions: for horizontal deviations from the localizer (LOC), and for vertical deviations from the 3°-glideslope (GS), maximum values were considered and the root mean square errors (RMSE) were calculated. The latter is a frequently used measure for fine-motor flight performance evaluations (Rantanen, Johnson, & Talleur, 2004; McClernon, Miller, & Christensen, 2012), even if it does not deliver the position information (Hubbard, 1987), which is of no interest for this study. Flight crew licensing standards require adhering deflections no larger than one dot for precision approaches (European Union, 2011). This unit corresponds to a deviation of ± 0.8° on the LOC and ± 0.4° on the GS for Airbus types and is indicated on two scales in the primary flight display. These flightpath deviations were measured for three different altitude segments. The upper segment, 3,000-1,000 ft above ground level (AGL), represents the initial instrument approach phase with medium difficulty, preparing the *stabilized approach* (SKYbrary, 2016b). The next segment, 1,000-270 ft AGL, stands for the increasingly demanding instrument approach phase within the limits for a stabilized approach, not exceeding deviations larger than one dot. The last segment, 270-50 ft AGL, represents the transition to the visual approach shortly before the landing flare. Apparently, GS data becomes somewhat unreliable in the last segment because some pilots seem to have commenced the flare above 50 ft AGL. The datasets were only analyzed if the participant had completely finished the approach without aborting it. A further measure for manual flight performance with high practical relevance is the first point of touching down (TD) upon landing. We measured these TD points in two dimensions: absolute longitudinal distances to the threshold of the runway and absolute lateral distances to the centerline. The ideal TD point is the boldly marked aiming point located about 400 m (1,312 ft) behind the threshold. For this evaluation all datasets were included where the aircraft touched down, with or without a preceding go-around.

RESULTS

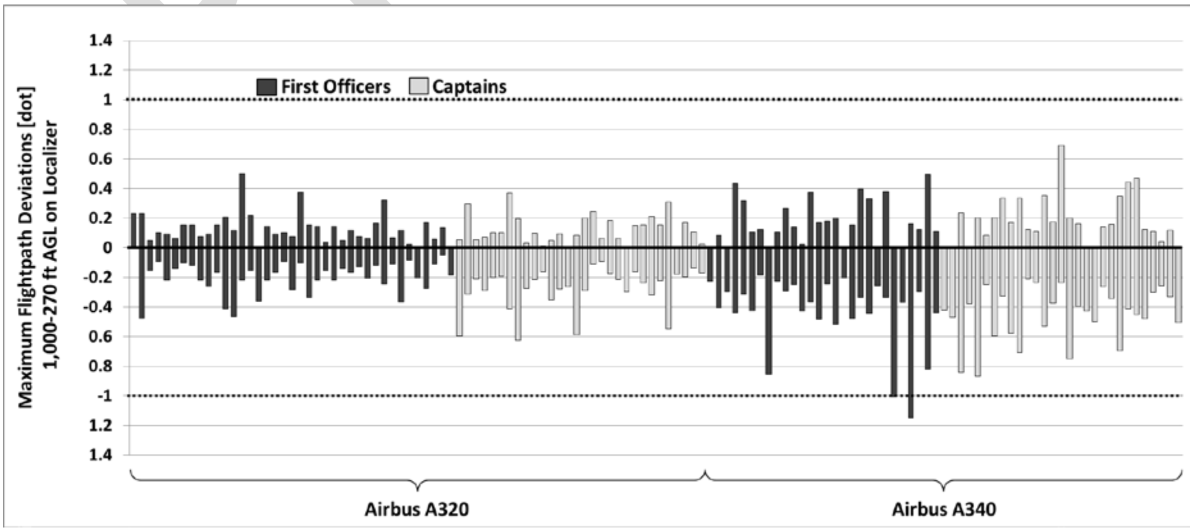
All effects will be reported as significant at $p < .05$ and η_p^2 is given as effect size. If the assumption of sphericity is violated, Greenhouse-Geisser estimates of sphericity are reported by ε. The mean values of all groups are presented in Table 2 for both measures: the flightpath deviations and the touchdown points.

Table 2: Results of pilot groups for ILS flightpath and touchdown point deviations

	RMSE ILS flightpath deviations [dot]												TD point deviations [m]			
	3,000-1,000 ft AGL				1,000-270 ft AGL				270-50 ft AGL				LONG		LAT	
	LOC		GS		LOC		GS		LOC		GS					
Group	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
FO A320	.10	.04	.14	.05	.12	.06	.19	.08	.14	.08	.74	.55	418	115	2.0	1.5
FO A340	.20	.08	.25	.10	.25	.14	.30	.13	.23	.18	.88	.67	501	133	2.1	2.0
CPT A320	.11	.04	.13	.06	.14	.06	.17	.07	.18	.12	.58	.28	428	104	1.2	1.1
CPT A340	.31	.22	.38	.20	.26	.10	.38	.18	.25	.11	1.2	.72	510	182	5.3	5.0

Deviations from Localizer and Glideslope

The maximum deviations on localizer (Figure 1) and glideslope (Figure 2) in the segment between 1,000 and 270 ft AGL indicate that ten (18%) out of 57 approaches (all on A340) exceeded with at least one flight parameter the limits of a stabilized approach or of the licensing standards of a precision approach.



This is an Accepted Manuscript of an article published by Sage Publishing in Human Factors on 13 Apr 2016, available online:http://hfs.sagepub.com/cgi/pmidlookup?view=long&pmid=27076096.

Figure 1: Individual maximum flightpath deviations on localizer. Dotted lines indicate limits not to be exceeded for stabilized approach and license checks.

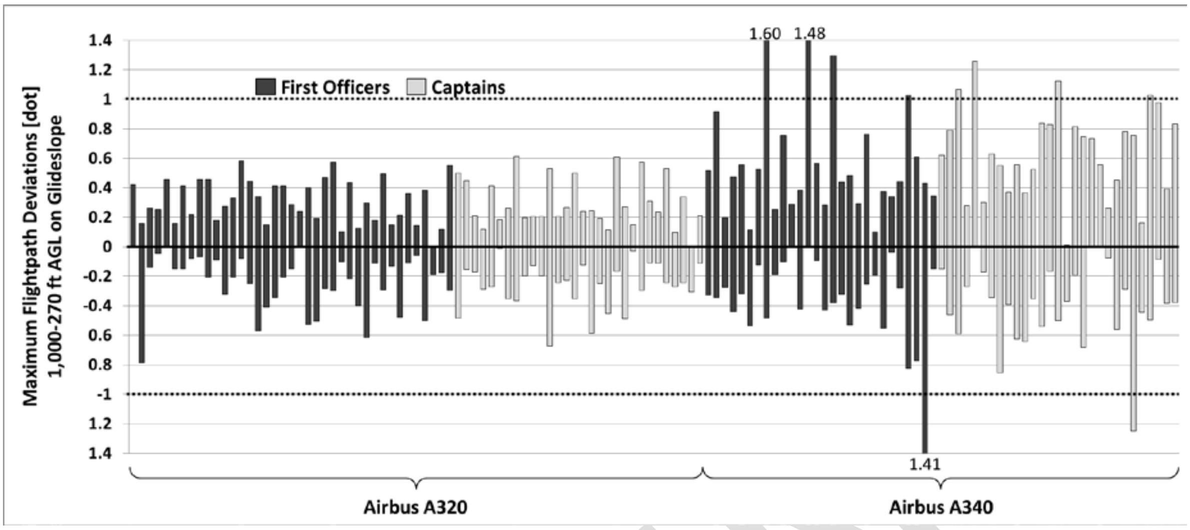


Figure 2: Individual maximum flightpath deviations on glideslope. Dotted lines indicate limits not to be exceeded for a stabilized approach and license checks. Values larger than ± 1.4 dot are depicted by numbers.

All individual flightpath deviations averaged by the RMSE in the most relevant segment between 1,000 and 270 ft AGL are displayed in Figure 3. These flight performance data were analyzed with a 2 × 2 × 3 (between-subjects Fleet [A320, A340] × between-subjects Rank [FO, CPT] × within-subjects Altitude [3,000-1,000, 1,000-270, 270-50]) multivariate analysis of variance with two dependent variables: deviations on localizer (LOC) and glideslope (GS). In most cases RMSE and absolute deviation scores are not normally distributed, but positively skewed. For that reason all flightpath deviation data have been log-transformed for the further statistical analysis. The results of ILS deviations, which contain in a number of significant between- and within-subjects effects, are depicted in Table 3. Average differences between groups are visualized in Figure 4 and Figure 5.

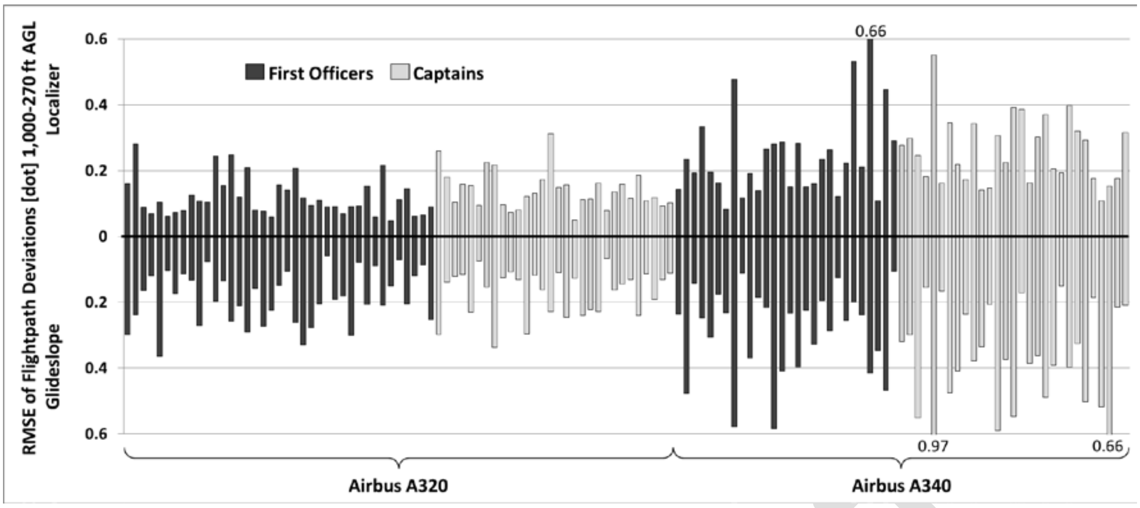


Figure 3: RMSE of all individual flightpath deviations on localizer (upper half) and glideslope (lower half). Values larger than ± 0.6 dot are depicted by numbers.

Table 3: Statistical analysis of log-transformed ILS flightpath deviations

Source	Multivariate Tests	Univariate Tests
between-subjects		
Fleet	V = .45, F(2, 121) = 48.90,	LOC: F(1, 122) = 76.84, p < .001, η_p^2 = .39
	p < .001, η_p^2 = .45	GS: F(1, 122) = 75.77, p < .001, η_p^2 = .38
Rank	V = .08, F(2, 121) = 5.50,	LOC: F(1, 122) = 11.09, p = .001, η_p^2 = .08
	p = .005, η_p^2 = .08	GS: F(1, 122) = 3.24, p = .074, η_p^2 = .03, n.s.
Fleet * Rank	V = .09, F(2, 121) = 5.60,	LOC: F(1, 122) = .20, p = .655, η_p^2 = .00, n.s.
	p = .005, η_p^2 = .08	GS: F(1, 122) = 9.20, p = .003, η_p^2 = .07
within-subjects		
Altitude	V = .80, F(4, 119) = 118.53,	LOC: ε = .87, F(1.75, 213.18) = 1.78, p = .175, η_p^2 = .01, n.s.
	p < .001, η_p^2 = .80	GS: ε = .88, F(1.76, 215.03) = 318.66, p < .001, η_p^2 = .72
Altitude * Fleet	V = .14, F(4, 119) = 4.73,	LOC: ε = .87, F(1.75, 213.18) = 8.00, p = .001, η_p^2 = .06.
	p = .001, η_p^2 = .14	GS: ε = .88, F(1.76, 215.03) = 7.00, p = .002, η_p^2 = .05
Altitude * Rank	V = .02, F(4, 119) = .70,	
	p = .595, η_p^2 = .02, n.s.	
Altitude * Fleet * Rank	V = .03, F(4, 119) = .81,	
	p = .524, η_p^2 = .03, n.s.	

Note. These analyses are based on log10-transformed flightpath deviation data. The assumption of normality was violated for only one out of 24 datasets: glideslope, 3,000-1,000 ft AGL, CPTs on A320. The assumption of equality of covariance matrices and error variances was violated; however, by having only two dependent variables, we assume this violation as minor (Tabachnick & Fidell, 2007, p. 252). For multivariate tests, Pillai's statistic was reported. When the assumption of sphericity was violated (within-subjects tests), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε). The assumption of equality of error variances was violated for one out of six datasets: localizer, 270-50 ft AGL.

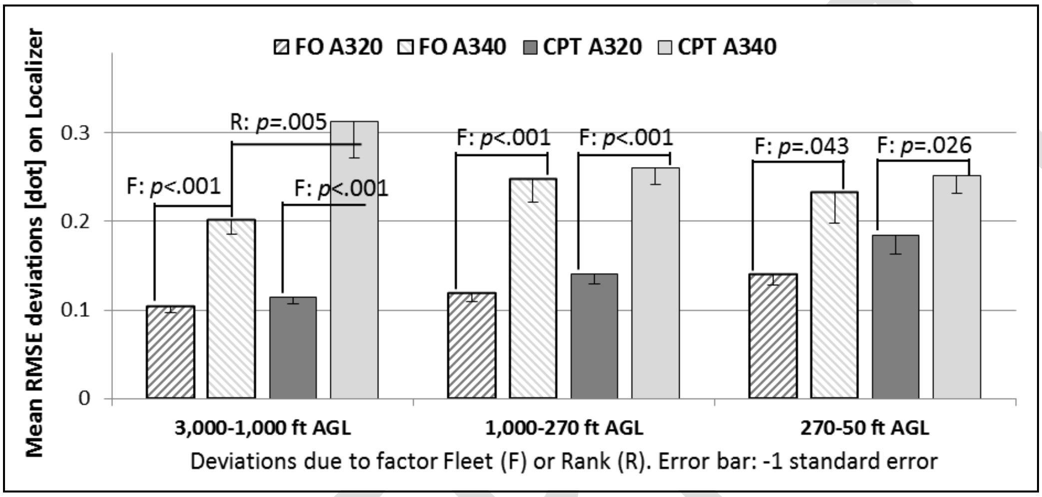


Figure 4: Visualization of localizer deviations based on non-transformed data indicating Bonferroni post hoc comparisons based on log-transformed data.

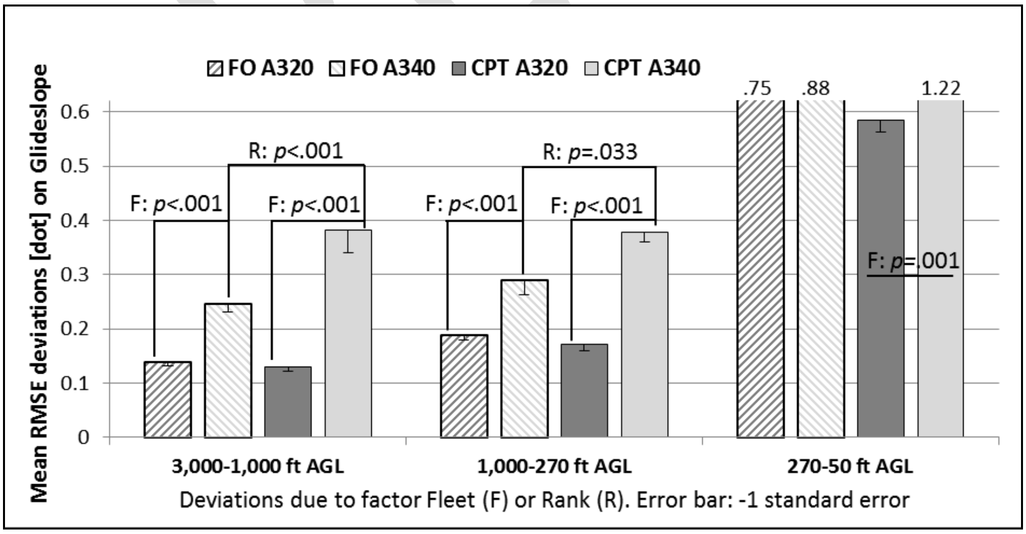


Figure 5: Visualization of glideslope deviations based on non-transformed data indicating Bonferroni post hoc comparisons based on log-transformed data. Values larger than 0.6 dot are depicted by numbers.

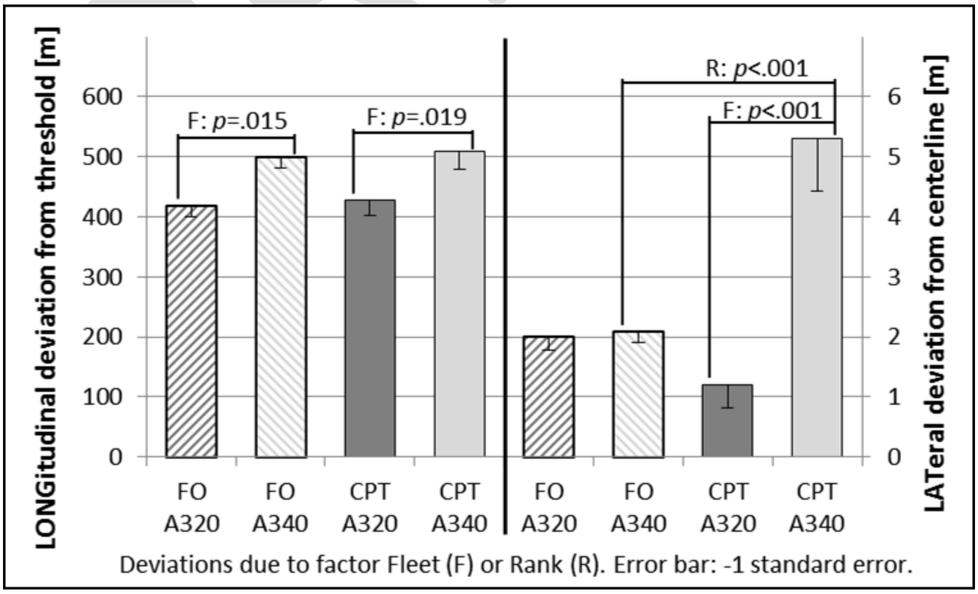
Analysis of Touchdown Points

A two-way MANOVA was conducted, including the aforementioned factors Fleet and Rank, and two dependent variables: longitudinal (LONG) and lateral (LAT) absolute distance to the threshold and to the centerline of the runway, respectively. Furthermore, Pillai's trace was chosen as a rather robust multivariate test statistic showing significant differences between groups. Statistical analysis of TD points is given by Table 4, and differences between groups are illustrated in Figure 6.

Table 4: Statistical analysis of TD-point deviations

Source	Multivariate Tests	Univariate Tests
Fleet	$V = .18$, $F(2, 125) = 13.43$, $p < .001$, $\eta_p^2 = .18$	LONG: $F(1, 126) = 11.67$, $p = .001$, $\eta_p^2 = .09$ LAT: $F(1, 122) = 17.25$, $p < .001$, $\eta_p^2 = .12$
Rank	$V = .04$, $F(2, 121) = 2.54$, $p = .083$, $\eta_p^2 = .04$, n.s.	°LONG: $F(1, 126) = .15$, $p = .699$, $\eta_p^2 = .00$, n.s. °LAT: $F(1, 126) = 5.06$, $p = .026$, $\eta_p^2 = .04$
Fleet * Rank	$V = .11$, $F(2, 125) = 7.90$, $p = .001$, $\eta_p^2 = .11$	LONG: $F(1, 126) = .00$, $p = .980$, $\eta_p^2 = .00$, n.s. LAT: $F(1, 126) = 15.84$, $p < .001$, $\eta_p^2 = .11$

Note. For multivariate tests, Pillai's statistic was reported. The assumption of equality of covariance matrices and error variances was violated. °For the factor Rank, only a statistical trend was found in the multivariate test. Univariate results are reported to complete the picture (Tabachnick & Fidell, 2007, p. 269).



This is an Accepted Manuscript of an article published by Sage Publishing in Human Factors on 13 Apr 2016, available online:http://hfs.sagepub.com/cgi/pmidlookup?view=long&pmid=27076096.

Figure 6: TD-point deviations indicating univariate test results.

Effect of Age and Time Since Initial Training

As shown in Table 1, the between-subjects effects of Rank and Fleet are correlated with the elapsed time since initial training and also with age. Based on the assumption that basic flight training is essential for building manual flying skills and that these skills are prone to decay without regular practice in advanced-technology aircraft, the reported significant effects of Fleet and Rank could be confounded with differences in time since initial flight training or with age. In order to rule out the possibility that we simply found time-related effects, we included age and time since flight training as covariates in separate MANCOVAs, corresponding to the abovementioned MANOVAs in Table 3 and Table 4. In fact, age and time since flight training are intercorrelated, r_s = .95, *p* (one-tailed) < .001. Consequently the results are almost identical, and we only report the findings for the covariate time since initial training.

In the MANCOVA of the ILS deviations, with time since initial training as a covariate, only two between-subjects effects are significant. While the effect for Rank disappears, Fleet (*V* = .31, *F*(2, 120) = 26.72, *p* < .001, η_p^2 = .31) and Fleet * Rank (*V* = .09, *F*(2, 120) = 5.98, *p* = .003, η_p^2 = .09) are significant. The effect of time since initial training itself is not significant. The findings for the TD-point deviations are similar. Time since initial training has no significant direct effect on the absolute deviation scores, and Rank is also not a significant between-subjects factor. Nevertheless, the differences between long-haul and short-haul pilots are still statistically significant (*V* = .08, *F*(2, 123) = 5.13, *p* = .007, η_p^2 = .08) as well as the Fleet * Rank (*V* = .11, *F*(2, 123) = 7.71, *p* = .001, η_p^2 = .11) interaction. Through these MANCOVAs we can provide evidence that the differences in flying performance between long-haul and short-haul pilots cannot be interpreted as simple age or time effects. This confirms our second research question that daily flight practice has a stronger influence on manual flight proficiency than the time dated back to flight school.

DISCUSSION

This is an Accepted Manuscript of an article published by Sage Publishing in Human Factors on 13 Apr 2016, available online:http://hfs.sagepub.com/cgi/pmidlookup?view=long&pmid=27076096.

General Findings According to Fine-Motor Flying Skills

With the comparison of flight performance of pilots on long-haul versus short-haul fleets, this study offers a quasi-experimental approach to the analysis of practice and training effects on the level of manual flying skills. The reported results clearly confirm that the level of practice and training as measured by daily flying practice and elapsed time since initial flight training does have significant influences on fine-motor flying skills of airline pilots. In summary, we found the following rank order for fine-motor flight performance: CPT A320 > FO A320 > FO A340 > CPT A340. According to Table 1, the A320 CPTs had at least two advantageous factors: (1) they performed the highest number of landings in the past month, and (2) they had more flight hours on type compared to all other groups. The A320 FOs had an equal amount of practice in the past month but less total flying experience and less time on type. When looking at the long-haul data it seems that the total flight time and the time on type beyond 2,000 or 3,000 hours are less important factors for the level of manual skills than the daily practice and the time period since flight school. Therefore, the A340 FOs generally had more difficulties than the A320 FOs. Moreover, the A340 CPTs could not use their enormous flying experience as an advantage for the manual flying tasks. These senior long-haul pilots perform on average less than a quarter of the number of takeoffs and landings compared to short-haul pilots (Table 1). Hence, they have substantially less opportunity to practice their manual flying skills. If the level of skill is directly related to the amount of daily practice, long-haul pilots should show inferior performance in a manual flying task. Besides the type of operation (Fleet), another factor is suspected as being responsible for reduced manual flying skills: the pilot generation (called *time since initial training* in this study). According to the opinion of the European Aviation Safety Agency (2013, p1), senior pilots may be less comfortable with automation, while younger pilots may lack basic flying skills because they normally have less flying time on non-glass cockpit aircraft types. But then, the time interval since basic flight training and hence, the time for skill decay is shorter for younger pilots. Consequently as expected we found an interaction of the main effect for Fleet with the factor Rank.

A limiting factor of our previous research (Haslbeck et al., 2014) was that the different sources of variance (e.g. level of practice, flying experience, and type of aircraft) could not be separated because only A320 FOs had been compared to A340 CPTs. To reach a more conclusive comparison, we included two additional groups, A320 CPTs and A340 FOs, in this study. The findings concerning the importance of the amount of current practice for the level of manual flying skills can generally be confirmed. After all, in this study we identified additional factors that are related to time since initial training, age, and experience.

All results in this study have clearly shown substantial influences of Fleet on all manual flight performance scores. Many long-haul pilots have demonstrated consistently larger deviations from the ideal ILS flightpath, which can be explained by the lower level of practice. While the mean RMSE deviations from the localizer tend to remain constant across the three altitude segments, the deviations from the glideslope increase sharply for the final segment – the transition from instrument to visual flying. The first order interaction effect (Altitude * Fleet) illustrates that the differences between the fleets become somewhat smaller when the aircraft approaches the ground, with the exception of the glideslope deviations during the visual segment. Long-haul crews flew higher above the glideslope, obviously aiming at a TD point wider into the runway. For the sake of completeness it must be said that Munich (EDDM) has 4,000 m long runways which significantly reduce the potential consequences of longer landings.

As a second performance parameter, we analyzed the absolute distances of the TD points from the threshold and from the centerline, respectively. Again, a strong effect for the factor Fleet was found. Short-haul pilots landed closer to the centerline and about 400 m down the runway, while long-haul pilots performed longer landings (about 500 m beyond the threshold) with larger deviations from the centerline. The interaction Fleet * Rank is due to the lower performance of long-haul CPTs.

The nature of effects for the between-subjects factor Rank is more complex because CPTs were on average 13 to 14 years older than the FOs. Additionally, CPTs have accumulated about 8,000 hours more flight time, and the time since initial flight training was 12 to 13 years longer. Age, flight time, and time since flight school are highly correlated. In order to neutralize these confounding variables we executed MANCOVAs with time since initial training as a covariate. In these analyses all between-subjects effects of Rank were insignificant, while Fleet still explained 31 % of the variance for the ILS deviation measures and 8 % of the variance for the TD points. Obviously, the level of practice measured by the number of executed landings per month contributed significantly to the decrement of manual flying skills in both CPTs and FOs of the long-haul fleet. This effect remained significant regardless of age or other time-related factors. The first-order interaction effects Fleet * Rank explained further variance in the analyses of ILS deviations and of the TD points. While CPTs showed better performance than FOs on the short-haul fleet, the long-haul FOs performed slightly better than the long-haul CPTs.

Besides the significance of current practice for fine-motor flying performance, our findings do not confirm recent concerns about a general lack of basic flying skills among the younger generation of pilots (European Aviation Safety Agency, 2013; Civil Aviation Authority, 2014). As the youngest group, the A320 FOs with an average age of 30 and a little less than six years of airline experience performed second best on the manual ILS and landing. At this stage of their career, they clearly had sufficient opportunity in practice and training to develop the necessary level of flying skills. In summary, pilots with little recurrent practice and extensive use of automation seem to be running the risk of losing Langenwiesche's (1944) touch and feel of how to *fly* an aircraft. Especially, when considering Figure 3 the concern arises, what happens to pilots with even higher automated aircraft and longer working lifetime possibly spending on long-haul operation?

Limitations

Our analysis was carried out with pilots from one airline only. Findings could be different in other airlines with other training schemes and other pilot career models for their flight crews. We are also aware that *manual flying skills* in fact cover more tasks than a manual ILS approach with a precise landing within the touch-down zone. However, our aim was to complement existing research with objective performance data. By using scale deflections as the unit of accuracy we assured that flightpath deviations are not weighed disproportionally against distance to touchdown.

One latent confound in this study deals with the different aircraft types. Both aircraft types are equipped with a *fly-by-wire* flight control system which has been designed to provide the same flying and handling qualities and to maintain the highest applicable extent of commonality (Brière & Traverse, 1993; Favre, 1994; Joint Aviation Authorities, 2004; Bissonnette & Culet, 2013) providing a similar look and feel for the pilot. Nevertheless, differences in the dynamics between both types exist. However, the question is not whether both types can be controlled identically, but whether both types can be controlled identically precise when sufficient pilot training accounting for specific peculiarities of each type has been completed. This second question can be confirmed by the fact that runways, precision approaches, and certification standards for these types as well as licensing standards for the pilots are the same. According to the manufacturers homepage “a large majority of pilots praising the handling qualities of Airbus aircraft and their commonality across the complete range of products” (Airbus, 2016). Additional evidence for the commonality of the two aircraft types with respect to achievable precision in manual control comes from several A340 pilots in our sample who performed nearly on the same high level as the A320 crews did (Figure 1-3). However, none of the A340 pilots had that high level of daily flight practice as the A320 pilots did.

Recommendations

Until fail-proof automation outperforms human performance in all situations, the pilots remain the last line of defense in the cockpit. Under the described circumstances these pilots need even stronger manual flying skills as proposed by Langewiesche (1944) in the earlier days of aviation. Based on our findings we suggest a number of organizational and design recommendations for how to prevent a significant deterioration of manual flying skills. First of all, our findings indicate that the recent amount of regular simulator training is not sufficient to maintain manual flying skills of long-

This is an Accepted Manuscript of an article published by Sage Publishing in Human Factors on 13 Apr 2016, available online:http://hfs.sagepub.com/cgi/pmidlookup?view=long&pmid=27076096.

haul crews. Pilots with part-time schedules or reduced flight duties, like management pilots and pilots on parental leave, need special attention even when operating on short-haul service. Specific flight simulator training with a focus on manual flying tasks is one possible intervention. Mixed-fleet flying could be another powerful approach to increase a pilot’s practice if negative transfer effects can be kept under control (Lyall & Wickens, 2005). In this case pilots would perform short-haul and long-haul flights with type ratings for both types of aircraft in an alternating scheme. Another measure could be the operation of highly frequented short-haul connections with long-haul aircraft, like several flights within Japan. More manual flight practice could also be derived by changing companies’ automation policies to encourage pilots to fly manually if the situation permits (Federal Aviation Administration, 2013b; European Aviation Safety Agency, 2013). Such interventions have to be applied in the earlier stages of a pilot’s career before degradation can take place. Otherwise, avoidance behaviors and a feeling of discomfort according to manual flying could lead into a negative spiral of permanently less manual flight conduction. From a design perspective, intelligent (Geiselman, Johnson, & Buck, 2013) or adaptive automation (Parasuraman, 2000) could charge a pilot with several tasks to maintain his attention and situation awareness, thus keeping the pilot *in the loop*. Short-term effects can be avoiding automation surprises, which can lead to severe accidents, while a long-term effect can be the preservation of skills. However, all recommendations have to be considered for potential tradeoffs at the expense of safety by possible undesired side effects.

KEY POINTS

- Commercial airline pilots showed different levels of practice and training according to their scheduled type of operation, short-haul or long-haul.
- Fleet (distinction between short-haul and long-haul) showed large significant effects on all analyzed manual flight performance indicators.

- Rank (distinction between captain and first officer) only showed little effects on manual flight performance.
- All results supported the conclusion that recent flight practice is a significantly stronger predictor for manual flying performance than the time period since flight school or even the total or type-specific flight experience.

REFERENCES

Airbus SAS. (2016). Fly-by-Wire. Retrieved from http://www.airbus.com/innovation/proven-concepts/in-design/fly-by-wire/

Bureau d’Enquêtes et d’Analyses pour la sécurité de l’aviation civile. (2012). Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris. Le Bourget.

Bissonnette, N. & Culet, J. P. (2013). *Airbus A318/A319/A320/A321/A330/A340* (Flight Standardization Board (FSB) Report). Renton.

Brière, D., & Traverse, P. (1993). AIRBUS A320/A330/A340 Electrical Flight Controls - A family of fault-tolerant systems: A Family of Fault-Tolerant Systems. *Proceedings of The Twenty-Third International Symposium on Fault-Tolerant Computing*, 616–623. doi:10.1109/FTCS.1993.627364

Buckley, R., & Caple, J. (2009). The theory & practice of training (6th ed.). London: Kogan Page.

Casner, S. M., Geven, R. W., Recker, M. P., & Schooler, J. W. (2014). The Retention of Manual Flying Skills in the Automated Cockpit. Human Factors, 56(8), 1506–1516. doi:10.1177/0018720814535628

Childs, J. M., & Spears, W. D. (1986). Flight-skill decay and recurrent training. Perceptual and Motor Skills, 62(1), 235–242. doi:10.2466/pms.1986.62.1.235

Civil Aviation Authority. (2014). Flight-crew human factors handbook: CAP 737. Gatwick Airport South.

This is an Accepted Manuscript of an article published by Sage Publishing in Human Factors on 13 Apr 2016, available online:http://hfs.sagepub.com/cgi/pmidlookup?view=long&pmid=27076096.

Dismukes, K., Berman, B. A., & Loukopoulos, L. D. (2007). The limits of expertise: Rethinking pilot error and the causes of airline accidents. Ashgate studies in human factors for flight operations. Aldershot: Ashgate.

Drappier, J. (2008). The erosion of manual flying skills in highly automated aircraft. Flight Comment, 2008(2), 33–36.

Ebbatson, M. (2009). The Loss of Manual Flying Skills in Pilots of Highly Automated Airliners (PhD Thesis). Cranfield University, Cranfield.

European Aviation Safety Agency. (2013). EASA Automation Policy: Bridging Design and Training Principles. Cologne.

Commission Regulation (EU) No 1178/2011 of 3 November 2011, Volume 54 Official Journal of the European Union (European Union 2011).

Favre, C. (1994). Fly-by-wire for commercial aircraft: The Airbus experience. *International Journal of Control*, *59*(1), 139–157. doi:10.1080/00207179408923072

Federal Aviation Administration. (2013a). Operational Use of Flight Path Magamenet Systems: Final Report of the Performance-based operations Aviation Rulemaking Committee.

Federal Aviation Administration. (2013b). *Safety Alert for Operators* (No. SAFO 13002). Washington, DC.

Fleishman, E. A. (1966). Human Abilities and the Acquisition of Skill: Comments on Professor Jones' Paper. In E. A. Bilodeau (Ed.), Acquisition of Skill (pp. 147–167). New York: Academic Press.

Franks, P., Hay, S., & Mavin, T. J. (2014). Can competency-based training fly?: An overview of key issues for ab initio pilot training. *International Journal of Training Research*, *12*(2), 132–147. doi:10.5172/ijtr.2014.12.2.132

Geiselman, E. E., Johnson, C. M., & Buck, D. R. (2013). Flight Deck Automation: Invaluable Collaborator or Insidious Enabler? *Ergonomics in Design: The Quarterly of Human Factors Applications*, *21*(3), 22–26. doi:10.1177/1064804613491268

Gillen, M. (2008). Degradation of Piloting Skills (Master's Thesis). University of North Dakota, Grand Forks.

Haslbeck, A., Kirchner, P., Schubert, E., & Bengler, K. (2014). A Flight Simulator Study to Evaluate Manual Flying Skills of Airline Pilots. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 58, pp. 11–15). SAGE Journals.

Hubbard, D. C. (1987). Inadequacy of root mean square error as a performance measure. In International Symposium on Aviation Psychology, 4th (pp. 698–704).

International Air Transport Association. (2015). *Safety Report 2014*. Issued April 2015. Montréal.

Joint Aviation Authorities. (2004). Airbus A320 - A330 - A340 Cross Crew Qualification & Mixed Fleet Flying.

Lambregts, A. A., Nesemeier, G., Wilborn, J. E., & Newman, R. L. (2008). *Airplane Upsets: Old Problem, New Issues* (AIAA Modeling and Simulation Technologies Conference and Exhibit No. AIAA 2008-6867). Honolulu, Hawaii.

Langewiesche, W. (1944). *Stick and rudder: An explanation of the art of flying*. New York: McGraw-Hill.

Learmount, D. (2011). Industry sounds warnings on airline pilot skills. Retrieved from https://www.flightglobal.com/news/articles/industry-sounds-warnings-on-airline-pilot-skills-352727/

Lyall, B., & Wickens, C. D. (2005). Mixed Fleet Flying Between Two Commercial Aircraft Types: An Empirical Evaluation of the Role of Negative Transfer. In Human Factors and Ergonomics Society (Ed.), Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (Vol. 49, pp. 45–48).

McClernon, C. K., Miller, J. C., & Christensen, J. C. (2012). Variance as a Method for Objectively Assessing Pilot Performance. In Human Factors and Ergonomics Society (Ed.), Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting (pp. 85–89).

Newman, R. L. (2012). *Thirty Years of Airline Loss-of-Control Mishaps* (AIAA Modeling and Simulation Technologies Conference and Exhibit No. AIAA 2012-4495). Minneapolis, Minnesota.

National Transportation Safety Board. (2014). Crash of Asiana Flight 214 Accident Report Summary: Descent Below Visual Glidepath and Impact With Seawall. Public Meeting of June 24, 2014. Washington, DC. Retrieved from http://www.ntsb.gov/news/events/2014/asiana214/abstract.html

Parasuraman, R. (2000). Designing automation for human use: empirical studies and quantitative models. Ergonomics, 43(7), 931–951. doi:10.1080/001401300409125

Puentes, A. (2011). The Manual Flight Skill of Airline Pilots (Master's Thesis). San Jose State University, San Jose. Retrieved from http://scholarworks.sjsu.edu/etd_theses/4109

Rantanen, E. M., Johnson, N. R., & Talleur, D. A. (2004). The Effectiveness of Personal Computer Aviation Training Device, a Flight Training Device, and an Airplane in Conducting Instrument Proficiency Checks: Volume 2: Objective Pilot Performance Measures. Final Technical Report AHFD-04-16/FAA-04-6. Oklahoma City.

Sarter, N. B., & Woods, D. D. (1994). Pilot Interaction With Cockpit Automation II: An Experimental Study of Pilots' Model and Awareness of the Flight Management System. The International Journal of Aviation Psychology, 4(1), 1–28. doi:10.1207/s15327108ijap0401_1

Savion-Lemieux, T., & Penhune, V. B. (2005). The effects of practice and delay on motor skill learning and retention. Experimental brain research, 161(4), 423–431. doi:10.1007/s00221-004-2085-9

SKYbrary. (2016a). Pilot Handling Skills. Retrieved from http://www.skybrary.aero/index.php/Pilot_Handling_Skills

SKYbrary. (2016b). Stabilised Approach. Retrieved from http://www.skybrary.aero/index.php/Stabilised_Approach

Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th ed). Boston: Pearson/Allyn & Bacon.

This is an Accepted Manuscript of an article published by Sage Publishing in Human Factors on 13 Apr 2016, available online:http://hfs.sagepub.com/cgi/pmidlookup?view=long&pmid=27076096.

Vadrot, R., & Aubry, C. (1994). Cross Crew Qualification and Mixed Fleet Flying: The Airbus Familiy Concept. FAST, (17), 21–26.

Veillette, P. R. (1995). Differences in aircrew manual skills in automated and conventional flight decks. Transportation Research Record, (1480), 43–50.

Wiener, E. L., & Curry, R. E. (1980). Flight-Deck Automation: Promises and Problems. Ergonomics, 23(10), 995–1011. doi:10.1080/00140138008924809

Andreas Haslbeck received a Diploma degree in Mechanical Engineering from Technische Universität München in 2007. He is currently pursuing a Doctorate in Ergonomics. At the Institute of Ergonomics (TUM), where he was involved in experimental research using eye tracking and simulation, he conducted a large flight simulator study evaluating manual flying skills under the influence of practice and training. Besides eye tracking data, this experiment also recorded and analyzed workload, fatigue, and complacency effects.

Hans Juergen Hoermann is an accredited Aviation Psychologist at the German Aerospace Center (DLR) since almost 30 years. He holds a PhD in Applied Psychology (1987) from the Free University in Berlin. His research focuses on cockpit human factors and crew resource management training. As a Ground Instructor he is involved in pilot training.

(Digital) Appendix:

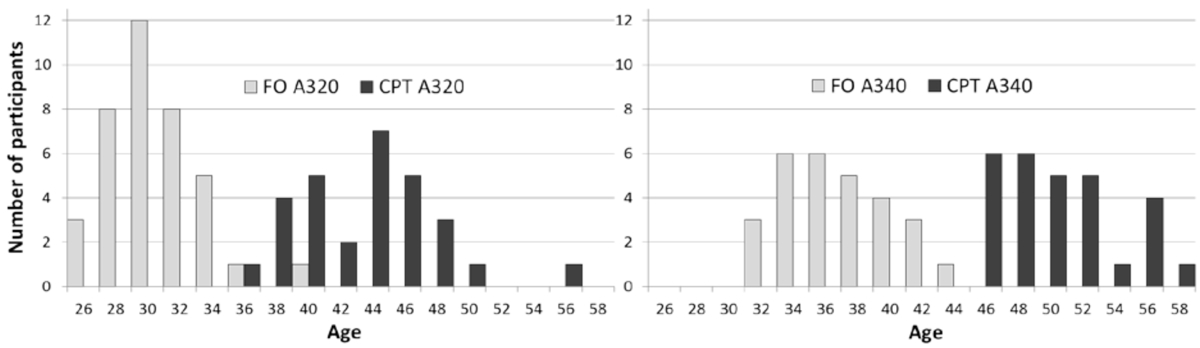


Figure 7: Distribution of participants' age

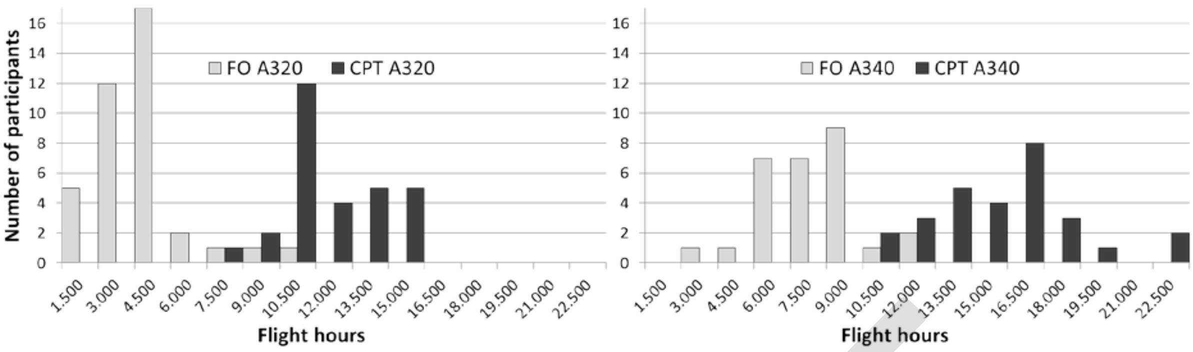


Figure 8: Distribution of participants' flight hours

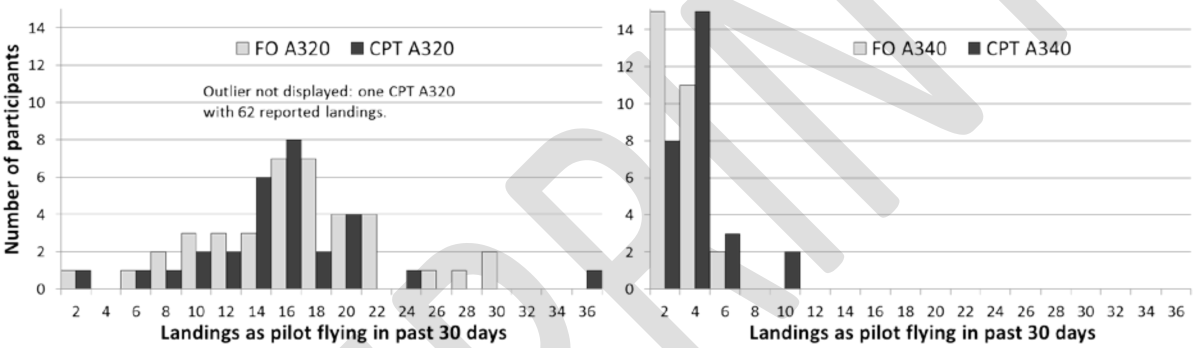


Figure 9: Distribution of participants' landings as pilot flying in past 30 days

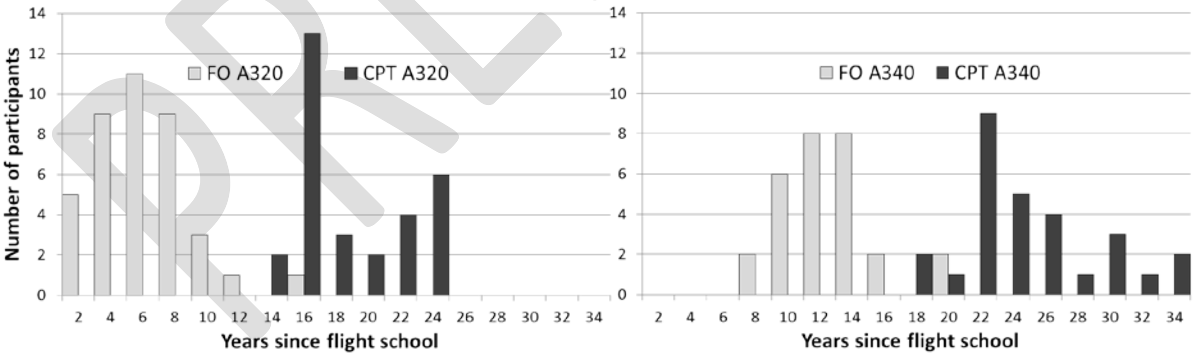


Figure 10: Distribution of participants' years since flight school

This is an Accepted Manuscript of an article published by Sage Publishing in Human Factors on 13 Apr 2016, available online:http://hfs.sagepub.com/cgi/pmidlookup?view=long&pmid=27076096.