

Manual flying skills under the influence of performance shaping factors

Andreas Haslbeck^{a,*}, Ekkehart Schubert^b, Linda Onnasch^c, Gerhard Hüttig^b, Heiner Bubb^a and Klaus Bengler^a

^a*Institute of Ergonomics, Technische Universität München, Boltzmannstrasse 15, 85747 Garching, Germany*

^b*Institute of Aeronautics and Astronautics, Section Flight Guidance and Air Transportation, Technische Universität Berlin, Marchstrasse 12-14, 10587 Berlin, Germany*

^c*Work, Engineering & Organizational Psychology, Department of Psychology and Ergonomics, Technische Universität Berlin, Marchstrasse 12, 10587 Berlin, Germany*

Abstract. This paper describes an experimental study investigating pilots' manual flying skills. In today's line oriented flight training, basic flying skills are neglected frequently. So, the study examines the manual flying skills of commercial airline pilots under the influence of several performance shaping factors like training, practice or fatigue in a landing scenario. The landing phase shows a disproportionate high percentage of aircraft accidents and it is typically flown by hand. The study is to be undertaken with randomly selected pilots in a full motion flight simulator to ensure a high validity of the results.

Keywords: manual flying skills, training, practice, performance shaping factors

1. Introduction

Basic flying skills are taught in the first chapters of a young pilot's basic training. They start with a small aircraft like Beechcraft Bonanza or similar types. The aim of this phase is to develop manual flying skills and an understanding of what "flying" is. Young pilots do their initial training from knowledge based understanding of the aircraft to a skill based handling of the aircraft. This process follows the skills-rules-knowledge model by Rasmussen [21].

Initial training is likely to be the phase with most intensive training of basic flying skills. The next steps of education are handling larger aircraft, managing the automation, crew resource management, human factors, safety issues – are all also very important matters. But manual flying is never again trained so intensively afterwards.

Therefore a potential problem for today's pilots is that there is not enough capacity to maintain a high amount of training opportunities for manual flying in line operation. So their manual flying capabilities

tend to degrade from the moment of their start of flying profession.

In addition, pilots' performance is not only affected by their manual flying skills, but also by other performance shaping factors (PSFs) like environmental conditions (mostly weather), social factors and team-interaction in the cockpit, fatigue or perceived workload. In this situation, many pilots also show a rather complacent behavior towards automation and their own skills; they overly rely on skills learned in the past but meanwhile these skills have eroded.

2. Present situation

There are only few recent studies on manual flying skills but an ever increasing number of accidents, which are traced back to lacking basic flying skills like Colgan Air flight 3407 [19]. This accident is a prime example of eroded manual flying skills in combination with a broad variety of performance shaping factors. Although there seems to be a substantial amount of belief that manual flying capabili-

* Corresponding author. E-Mail: haslbeck@tum.de, Phone +49 89 289 15406, Fax +49 89 289 15389

ties are decreasing, there still remains little scientific evidence that proves this point [9].

Under these circumstances, complacent behavior is a wide spread phenomenon. A study by Gillen [13, 14] has shown a gap between pilots' perception of subjective and objective risk. On the one hand, this experiment indicates that basic instrument skills are significantly deteriorating over time which means that the 30 tested pilots performed "well below" the FAA standards for ATP certification, despite the fact that they once had to pass this certificate. On the other hand, the participants responded as having more trust in their own skills.

Ebbatson [10] mentioned the "loss of manual flying skills" in his thesis. In a full flight simulator experiment and with different manual flying tasks, he found out that neither the amount of overall flight experience nor the flight experience on a specific type of aircraft, are sufficient predictors for manual flying skills. He additionally stated that these skills strongly degrade over time even in a study which has been conducted with volunteering pilots. These findings imply little benefit gained from long dated back practice and training lessons. Consequently, this requires frequent trainings for pilots on manual flying.

The training vice president of Airbus Industries [8] gives different reasons for diminishing skills. Besides economic pressure to avoid hand flying, he complains about the infrequent opportunities for long-haul pilots to practice manual flying. For example, it could happen that pilots on very long distance flights have no opportunity to manually land for one month or longer.

3. Experimental design

This paper now describes an experimental design to evaluate pilots' manual flying skills under the influence of some specific performance shaping factors. This experiment is to be done in a full flight simulator in co-operation with a major European airline and with professional pilots as participants. By using realistic type elements, such as a qualified flight simulator (JAR-STD 1A Level D) and professional flight crew, a reasonable scientific evaluation can be made.

3.1. Research questions

The aim of the current research is to investigate the effects of lack of practice and training on the quality of manual flying. It is expected that the less time that has passed between the initial training and the present,

and, the more flight operations that are done, the better the manual flying skills will be. The expectation is that training is overcompensated by practice and both are assumed to have a stronger effect on manual flying than "dry" experience; like long-standing pilots such as long-haul captains have.

Besides the main goal, there are more research questions to be explored: Is there a correlation between fatigue and manual flying skills? What effect does the level of practice and training have on workload perceived by pilots? Is there a lack of mode awareness or a kind of mode confusion in high workload situations, what will affect manual flying skills (e.g. in a high energy approach)? Is there a tendency to complacent behavior in high workload situations?

3.2. General requirements

To specify the plan of this very resource intensive experiment about the pilots' capability to hand fly an aircraft, some specific requirements should be taken into account:

- participants of the experiment should not be acquired on voluntary basis – a randomized sample is necessary to avoid a self-selection or volunteer bias [22], which means that pilots with rather high pronounced skills would volunteer,
- a power analysis for this experiment recommends at least 26 subjects under the assumption of having strong effects,
- participants need to be commercial airline pilots, because specific tasks to be considered are only trained and practiced by them,
- flight tasks and scenario have to be in a realistic setting, meaning that such should be conducted in a full flight simulator with real air traffic control (ATC) instructions because real radio communication raises pilots' workload if they have to distinguish between distant and own messages,
- also the cabin crew has to be involved to keep the passengers in mind for decisions,
- the tasks need to be able to be fulfilled but also difficult enough that there is a chance to fail,
- the simulation scenario procedure has to be highly standardized to ensure that all participants face in general the same technical, environmental and organizational conditions,
- relevant influences on pilots' performance need to be collected in the experimental environment.

These requirements ensure that variance can be found in the results, and also that gathered data can be analyzed and considered under the resilience engi-

neering approach [15]. Here, the pilots' actions can be evaluated according to whether they are successful and show highly developed manual flying skills or not. The approach at this point is to learn what successful (concerning manual flying tasks) pilots do differently in comparison to others. One important aspect would be, whether some elements of their aircraft handling can be identified as especially well performed and taken as a good standard. Examples would be the stick handling [18] or the instrument panel scan shown by different subjects.

3.3. Resulting experimental design

Based on the above mentioned requirements, the following experimental design was created.

Approach and landings are the flight phases with the highest number of aircraft accidents [4, 16]. So, for the experiment, a hand flown landing has to be performed by all participants. While on short-haul flights, only two pilots work together in the cockpit per flight, on long-haul routes, at many carriers three pilots are on one shift. That means, not each of these three have the opportunity to conduct a landing. Furthermore when bad weather conditions occur (low visibility) a landing is performed via auto-pilot ("hands-off landing").

As the independent variable, the level of training and practice – also one main PSF – is taken. A between groups design will be used with two different groups of pilots regarding the independent variable. For the first group, elder long-haul captains are taken with an assumed low level of practice and training. They have only a small number of flights per month [9]. Working on a typical long-haul schedule, they encounter about three to four landings by their own per month. In addition, their initial flight training dates back the longest possible period of time among pilots. There is also some evidence, reported by flight training instructors that captains tend to have lower manual flying skills.

The contrasting group should show a high level of practice and training: younger short-haul first officers are designated to participate. Their initial training dates back to the recent past and in short-haul schedule, they have to perform a significantly higher number of landings per month, up to 25 and more.

The dependant variable is the manual flying performance. Specifically, variations from ideal glide slope, localizer and approach airspeed are taken into consideration.

The following variables mediating the effect of training on performance are considered: fatigue, perceived workload and complacency towards automation [20].

Pilots normally work in a team of two, but in this experiment, only one of them is subject to the experiment. Here, the participant has the role of flying the airplane, referred to as the pilot flying (PF), while the other pilot's role consists of doing preliminary work for the PF. The pilot not flying (PNF) – better called pilot monitoring (PM) does surveillance, radio traffic, and, operates auxiliary systems like flaps or the landing gear. As both pilots are accustomed to working together, the PM is provided by the experimental team. This person must also be a commercial airline pilot and his behavior needs to remain constant for all participating PFs; thus, avoiding variance in the experimental results is caused by the PM.

To collect comparable data within and between both groups, the same type of aircraft would be best. But the requirement of having pilots with a minimum and a maximum of practice stands in contradiction to taking only one type of aircraft. To adhere to both demands, two different couples of aircraft are supposable: from Boeing the B737 and the B747 or B777 and from Airbus the A320 and the A330/340 or A380. In this experiment, both groups of pilots should fly models of the same manufacturer, because within all models of one company the man-machine-interface and automation concepts are rather similar. The tendency for the experiment is towards the Airbus fleet, because of its highly standardized cockpit design and similar aircraft handling (commonality) which enables easy changes between the A320 and A330/A340 family. Another reason is the technocentered higher automation level which needs a pilot to understand complex interrelations between the technical systems – "he monitors the monitor"[2]. Today's pilot is more and more distanced from a direct influence on the task of flight guidance (fly-by-wire system) and flight management [3]. So, pilots are getting pushed "out-of-the-loop" [28]. If a pilot detects a difference between the expected and observed results of his input, then Sarter [26] refers this to "automation surprises".

3.4. Flight scenario

To provide a realistic scenario for manual flying tasks, the following situation was set up (see also Table 1). Because of practical reasons, for the planned experiment, a full motion flight simulator

was made available during night hours. The intended landing was placed in the early morning hours, typical for long-haul overnight flights from Asia (Shanghai) and a short-haul flight from Russia (St. Petersburg). Before the experiment, all participants (in this case three per night) had a late dinner together with the experimental team. In this setting, the participants get to know the whole experimental team.

In the first part of the scenario, much foreign ATC communication (“party line”) between the airport and other approaching aircraft can be heard. These messages represent other airliners which reach the Munich Airport at morning hours – the same as they do in reality. The simulation scenario is preceded by an uneventful flight, which is briefed to all pilots. A following missed approach, which is triggered by too much tailwind in the final approach, can show a pilot’s decision for a safe go around instead of an unsafe landing. This situation stays inexplicit and uncertain until the very end because the wind information given by the airport shows an acceptable level of tailwind but the measured wind speed in the aircraft stays constantly too high. Furthermore the preceding aircraft had a safe landing on the 08L Munich runway without reporting too much tailwind. So the PF is confronted with a trade-off between the abort of the approach for a safe second try or a fuel-saving and economic landing with a noticeable higher risk. If the go around is not processed by the PF, the PM is instructed to call out (and perform, if necessary) “go around” at a height of 70 feet (ceiling here at 500 ft.). Preceding a second approach without tailwind because of a runway change (turning the landing direction on runway) an autopilot approach mode failure is evoked to gather information concerning mode awareness. Shortly before intercepting the localizer, the PF has to engage the approach mode. In this setting, the aircraft will automatically intercept and afterwards follow the localizer. With the above described evoked error, the aircraft will not engage the approach mode and stays in heading mode. Only an aware pilot will register that situation and recover the mode (e.g. turn the heading knob) or fly the localizer manually. A complacent PF will over trust the automation and register the defect very late, when the aircraft already overflies the localizer. After completion of this mode awareness check and before reaching the final approach fix, the simulator instructor deactivates the complete auto flight system. This invokes a malfunction disengaging the flight director and auto-pilot modes. Dealing that situation, the PF has at any case to proceed with a manually flown landing (raw data approach, ceiling here 270 ft.). The

aircraft has enough fuel to perform a second go around in this situation, but with the commitment to land (because of fuel limitation) PF must fly the aircraft manually under current weather conditions at Munich Airport. If a second go around would be performed, the weather conditions would be improved by the simulator instructor and if the PF asks for help, the PM could also take over control, because the participants should leave the experiment without being unmasked among the simulator crew.

Table 1. Timeline of flight scenario

00 min: start of scenario with fixed weight; PF comes into cockpit after last shift changeover at FL 200
20 min: missed approach due to strong tailwind (08L) and runway change
22 min: commitment to stay at Munich Airport
30 min: approach mode and autopilot malfunction
40 min: hand-flown landing on runway 26R

3.5. Measurements and method

For the assessment of pilots’ practice and training (independent variable), several different measures can be used. Flight hours in general and for the recent aircraft type and the amount of own hand flown landings within the last 30 days give an impression on the level of practice. The time prior to the last two simulator training lessons and its content give information on the training level of a pilot.

Flight performance (dependent variable) is to be objectively measured by the flight simulator data. Here, a multiplicity of technical parameters can be recorded. Along with the variations from ideal glide slope, localizer and approach airspeed, all operators’ input and displayed information, can be stored. In addition, a subjective self-assessment of the manual flight performance is also to be done as well as the same evaluation by the simulator instructor.

As vigilance, sleepiness, fatigue, and, the sleep history, have an influence on a pilot’s performance, these covariates are recorded. Some of the measurements concerning fatigue in the broader sense are done three times: after the dinner (about one up to two hours before the experiment), immediately before and after the experiment. These measures are conducted several times – a psychomotor vigilance test [7], the Karolinska Sleepiness Scale [1], one visual analog scale [cf. 12], different standardized fatigue scores [25, 24] and the critical flicker fusion frequency (CFF). For the assessment of the CFF, an

apparatus built by the Institute of Ergonomics is taken [cf. 27]. A sleep log [23] gathers data concerning the last three days prior to the experiment. Individual perceived workload is surveyed by the NASA-TLX questionnaire in the raw-version without weighting factors [5]. To assess the tendency of pilots to overly rely on automation, a questionnaire related to complacency [11] is taken. Complacency is to be measured in the situation when the approach mode will not arm before the second approach without auto flight system.

In addition to these measures, the participants' behavior can be very important information especially in certain situations. To gain an impression of both pilots' actions, all spoken communication is recorded as well as all observable actions are video recorded. For an advanced analysis of the operation, the PF is equipped with a head-mounted eye-tracking system. So, specific visual behavior and patterns can be reviewed afterwards. Furthermore the panel scan, i.e. fixations at cockpit instruments, like the primary flight display, can be measured.

4. Discussion

4.1. Discussion on anticipated results

The most important results of this study will be the information concerning the quality of manual flying. For a qualitative evaluation, these data have to be compared to pilots' certification guidelines [17] if they would have passed at the moment. The authors of this paper do not believe that all participants will perform above an acceptable level. A quantitative analysis will show variance within and between both groups of pilots. It is expected that the FOs group would show better results than the captains' group because of different amounts of every day practice. Here, three different measures are to be analyzed:

- information concerning the aircraft's state on which pilots have to react,
- frequencies and amplitudes of pilots steering inputs, and
- resulting path flown by the aircraft.

Eye tracking data will give detailed information about a pilot's individual panel scan. On the one hand, it can be compared if these procedures match to what was learned in flight school. On the other hand, it can be seen if decisions and actions performed by the PF based on the right perceived information. This can

also evaluate an airline's initial and recurrent training and operation.

4.2. General discussion

This experiment aims to measure the manual flying skills of pilots which might degrade over time [10] and in general because of spreading automation in today's cockpits. This suggested method is rather novel in comparison to many recent simulator studies because there is no specific sense of emergency the pilots must cope with. The scenario is typical for an approach at morning hours. For the purpose of presenting a realistic situation where hand flying is necessary, an autopilot defect is evoked. The tailwind situation is very typical for pilots and so the runway change is. Hence, the scenario is rather uneventful and offers mid-level difficulty. This kind of experiment also adheres to the idea of "drift into failure" [6], which means that the everyday operation can hassle pilots, and, make flight operation more error prone without any noticeable signs. This especially lets a socio-technical system drift into danger and failure. If in such a scenario a non sufficient performance is shown by pilots, there is no evidence that these pilots would be able to perform better in a more dangerous situation.

If the influence of the here deployed performance shaping factors significantly lowers the quality of manual flying, there would be no doubt that countermeasures need to be implemented into today's air traffic operation.

On criticism of this study, it has to be mentioned that there is only one manual flying task. A larger variety of such tasks would demand a significantly longer flight phase if embedded into a realistic scenario or drastic breaks between different tasks. However, with these further tasks, a more holistic image of a pilot's manual flying tasks could be recorded. This is a trade-off between an operable experiment and a comprehensive scientific approach. In addition, the measurement of fatigue and sleepiness is always difficult and often dissatisfying. Therefore several methods have been implemented into this study.

5. Summary and outlook

This paper has shown an experimental design to analyze pilots' manual flying skills, and, it is assumed that some deficiencies will be found there. The more interesting of questions is, of course, what

shall be done in light of these results. Should deficient pilots be taken out of operation? Here, the answer has to be “no”. This wouldn’t make sense because of economic reasons and also other social aspects. The more sustainable answer could be to analyze training programs – where does this erosion arise; to reconsider flight operation – is it necessary, that some pilots only perform on long-haul schedules; or could some more ergonomically designed cockpit instruments better support manual skills?

Acknowledgements

This work was conducted in co-operation with Deutsche Lufthansa AG, funded by the Federal Ministry of Economics and Technology via the Project Management Agency for Aeronautics Research within the Federal Aeronautical Research Program (LuFo IV-2). The authors thank all participants of the experiment for their contribution to the project.

References

- [1] Åkerstedt, T., & Gillberg, M. (1990). Subjective and objective Sleepiness in the Active Individual. *Intern. J. of Neuroscience*, 52(1-2), 29–37.
- [2] Bainbridge, L. (1987). Ironies of Automation. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.), *New technology and human error* (pp. 271–283). Chichester; New York: J. Wiley.
- [3] Billings, C. E. (1991). *Toward a Human-Centered Aircraft Automation Philosophy*. *The International Journal of Aviation Psychology*, 1(4), 267–270.
- [4] Boeing Commercial Airplanes (2011). *Statistical Summary of Commercial Jets Airplane Accidents: Worldwide Operations 1959 - 2010*. Seattle.
- [5] Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and raw task load index (TLX) correlations: are paired comparisons necessary? In A. Mital (Ed.), *Advances in industrial ergonomics and safety*. Proceedings of the Annual International Industrial Ergonomics and Safety Conference held in Cincinnati, Ohio, U.S.A., 5-9 June 1989 (pp. 481–485). New York: Taylor & Francis.
- [6] Dekker, S. W. A. (2004). Why we need new accident models. *Human Factors and Aerospace Safety*, 4(1), 1–18.
- [7] Dorrian, J., Rogers, N., & Dinges, D. (2005). Psychomotor Vigilance Performance: Neurocognitive Assay Sensitive to Sleep Loss. In C. A. Kushida (Ed.), *Sleep deprivation. Clinical issues, pharmacology, and sleep loss effects* (pp. 39–70). New York: Marcel Dekker.
- [8] Drappier, J. (2008). The erosion of manual flying skills in highly automated aircraft. *Flight Comment*, 2008(2), 33–36.
- [9] Drappier, J. (2009). *Training Challenges for the A350*. Asia Pacific Airline Training Symposium, Hong Kong.
- [10] Ebbatson, M. (2009). *The Loss of Manual Flying Skills in Pilots of Highly Automated Airlines* (PhD Thesis). Cranfield University, Cranfield.
- [11] Feuerberg, B., Bahner, J., & Manzey, D. (2005). Interindividuelle Unterschiede im Umgang mit Automation – Entwicklung eines Fragebogens zur Erfassung des Complacency-Potentials. In L. Urbas & C. Steffens (Eds.), *Zustandserkennung und Systemgestaltung*. 6. Berliner Werkstatt Mensch-Maschine-Systeme 13. bis 15. Oktober 2005 (ZMMS Spektrum Band 19) (pp. 199–202). Düsseldorf: VDI-Verl.
- [12] Folstein, M., & Luria, R. (1973). Reliability, validity, and clinical application of the visual analogue mood scale. *Psychological Medicine*, 3(4), 479–486.
- [13] Gillen, M. (2008). *Degradation of Piloting Skills* (Master's Thesis). University of North Dakota, Grand Forks.
- [14] Gillen, M. (2010). *Diminishing Skills?: An examination of basic instrument flying by airline pilots reveals performance below ATP standards*. *AeroSafety World Magazine*, 2010(July), 30–34.
- [15] Hollnagel, E., Woods, D. D., & Leveson, N. (2006). *Resilience engineering: Concepts and precepts*. Aldershot, England ; Burlington, VT: Ashgate.
- [16] International Air Transport Association (2011). *Safety Report 2010*. Issued April 2011 (No. 9049-11). Montréal.
- [17] Joint Aviation Authorities (2006). *Joint Aviation Requirements: JAR-FCL 1. Flight Crew Licensing* (No. Amendment 7). Englewood, Colorado.
- [18] Mayer, J. & Cox, T. H. (2003). *Evaluation of Two Unique Side Stick Controllers in a Fixed-Base Flight Simulator* (No. NASA/TM-2003-212042). Dryden Flight Research Center, Edwards, California.
- [19] National Transportation Safety Board (2010). *Loss of Control on Approach Colgan Air, Inc.: Operating as Continental Connection Flight 3407 Bombardier DHC-8-400, N200WQ Clarence Center, New York February 12, 2009*. Accident Aircraft Report (NTSB/AAR No. 10/01). Washington.
- [20] Parasuraman, R., Molloy, R., & Singh, I. (1993). Performance consequences of automation induced complacency. *International Journal of Aviation Psychology*, 3(1), 1–23.
- [21] Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3), 257–266.
- [22] Rosenthal, R., & Rosnow, R. L. (2007). *Essentials of behavioral research: Methods and data analysis* (3rd). Boston: McGraw-Hill.
- [23] Samel, A., Diedrich, A., Drescher, J., Lorenz, B., Plath, G., Vejvoda, M., & Wenzel, J. (1997). *Langzeitmonitoring psychophysiologischer Größen in der Flugphysiologie*. *Der Internist*, (38), 755–769.
- [24] Samel, A., Wegmann, H.-M., & Vejvoda, M. (1997). *Aircrew Fatigue in Long-Haul Operations*. *Accident Analysis and Prevention*, 29(4).
- [25] Samn, S. & Perelli, L. (1982). *Estimating aircrew fatigue: a technique with application to airlift operations* (SAM-TR-82-21). Brooks Air Force Base, Texas.
- [26] Sarter Nadine B. (1998). In D. D. Woods & Sarter Nadine B. (Eds.), *Learning from Automation Surprises and "Going Sour" Accidents: Progress on Human-Centered Automation*. Final Report (NASA/CR-1998-207061). NASA Ames Research Center.
- [27] Schmidtke, H. (1951). Über die Messung der psychischen Ermüdung mit Hilfe des Flimmertests. *Psychologische Forschung*, 23, 409–463.
- [28] Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd). Upper Saddle River, NJ: Prentice Hall.