How Can Procedures and Checklists Help Pilots in Abnormal Flight Situations?

Andreas Haslbeck[1]*, Patrick Gontar*[1]*, and Ekkehart Schubert*[2]

[1]*Institute of Ergonomics*
Technische Universität München
Boltzmannstrasse 15, 85747 Garching, Germany

[2]*Section Flight Guidance and Air Transportation*
Technische Universität Berlin
Marchstrasse 12-14, 10587 Berlin, Germany

ABSTRACT

This paper introduces a flight simulator study to evaluate procedures and checklists for use in abnormal situations. These utilities are meant to support pilots in hazardous situations, but for extreme situations, they might be inappropriate. Sixty crews (A340, A320) were required to fly in an abnormal flight scenario with a technical defect (loss of one main hydraulic system), which leads to several subsequent events. The aircrews had to mitigate these situations in order to safely land. The decreasing remaining fuel intensified time pressure while overloaded procedures had to be performed. In this study, the provided procedures were tested in this simulated crisis scenario.

Keywords: flight simulator scenario, abnormals, standard operation procedures

INTRODUCTION

Many checklists (digital and paper-based) and also standard operation procedures (SOPs) have been established to help pilots act in abnormal, often critical situations (simply called abnormals in aviation). These tools have to be as accurate and comprehensive as possible but also short and easy to handle. Thus, aircraft manufacturers and airlines have to manage a tradeoff between these opposing goals. As procedures are typically developed by expert pilots and technicians, and also by lawyers, a crucial question is whether these checklist tools are partially polluted by items whose sole purpose is to relieve the manufacturers and airlines of legal responsibility in case of an accident. Another issue is the context a procedure was made. It might happen that these checklists are not suitable for all practical situations and contexts, for example during fuel emergencies or adverse weather conditions. Furthermore, recommendations stated by accident reports led to stretched procedures. These additional guidelines are important to avoid future incidents, however, these items frequently lengthen the affected checklists.

An abnormal situation in aviation is frequently defined as the opposite of normal operation and more serious than non-normal situations. When analyzing such abnormals, they frequently turn out as highly complex and unfamiliar to the pilots. These circumstances normally lead to frequent human/operator errors (VDI, 2003) and sometimes time pressure often emerges as an additional hazard, especially in the approach and landing phase. A well-known accident is US Airways flight 1549, which ditched into the Hudson River. Cpt. Sullenberger and his crew did not have much time and were losing thrust from both engines. After the birdstrike, which resulted in the rapid loss of thrust, the crew had about three minutes to reach the ground – in this case, the Hudson River off midtown Manhattan. In these short moments, they were not able to properly complete procedures and checklists, because the flight crew also had to communicate with air traffic control (ATC), control the airplane, try to restart the engines and to find a spot to land. This situation can be described as an abnormal with high time pressure. Another abnormal with very different

https://openaccess.cms-conferences.org/#/publications/book/978-1-4951-2098-5

circumstances was an uncontained engine failure of an Airbus A380 known as Qantas Flight 32 on the 4^{th} November 2010. In this incident, a second aircrew was present and aided the primary aircrew. Additionally, the aircraft had a sufficient amount of fuel remaining and therefore no time pressure was involved. Despite these factors, the pilots also decided to not follow through with all the required procedures.

This leads to one problem of recent recurrent training in aviation: There is a strong focus on scenarios which are partly known to pilots beforehand and which offer a rather standardized way to accomplish (Casner, Geven, & Williams, 2013). This approach enables one to learn how to solve such a scenario and delivers objective and standardized evaluation criteria for instructors and check-pilots. In contrast, this kind of scenario-based training does not promote the ability to handle unknown hazardous situations. Therefore, this paper introduces a study to evaluate the use of abnormal checklists and procedures in an unknown and time critical scenario.

PREVIOUS STUDIES EVALUATING PILOTS' BEHAVIOR IN ABNORMAL SITUATIONS

Nikolic and Sarter (2007) conducted a flight simulator study to observe and evaluate pilots' strategies for diagnosing and recovering from abnormals. Their twelve Boeing 747 pilots were required to handle three different abnormal events: an unexpected and difficult climbing clearance, and two different issues of mode awareness. The authors did not find any significant differences in the captains' performance compared to the results of the first officers. Nevertheless, all pilots struggled with the given events at some point in the scenario. Further results showed a broad variety of recovery methods and nearly all pilots failed to diagnose the situation properly. Ebbatson, Harris, Huddlestone, and Sears (2010) combined a manual flying task with an engine failure (single engine ILS with backing crosswind) to evaluate pilots' performance. Some might not consider flying an aircraft manually as an abnormal, but recent accidents (e.g. Colgan Air 3407, Air France 447, Asiana 214) have powered a debate about the influence of automation on manual flying skills, which led to official attention (Federal Aviation Administration, 2013). Under manual flying circumstances, which are also associated with higher workload, an evaluation of manual handling skills is a valid approach for abnormals. Ebbatson et al. (2010) found an influence of short-term flight experience on manual flying skills and control input strategies. Casner et al. (2013) faced 18 Boeing 747 pilots with different abnormal situations (stall, wind shear and engine failure) under two different experimental conditions: under known conditions like airline training and under unexpected conditions. The results seen in this study imply that pilots can cope with known and frequently trained abnormals, but they have difficulty in unknown or unforeseen events. The outcome of the flown scenarios showed a wider spread of actions, decisions, and flight performance than someone might anticipate in recurrent training. Their conclusions on this study represent argue for improving pilots' training with surprising abnormal scenarios.

These studies describe the different problems today's pilots are faced with. In sum, aircrews show weaknesses in diagnosing unexpected events and such surprising scenarios are very rare in pilot's training. Another issue is the use of automation and consequential skill degradation. One important guideline for operators in safety critical domains is that under normal conditions they must adhere to the given SOPs. They should only deviate from SOPs when this behavior seems to provide a safer consequence to their actions. This principle is also valid for pilots. Thus, an experiment evaluating procedures and pilots should consider both situations: the need to adhere to SOPs and the need to dismiss SOPs. One approach to create a situation where a pilot needs to deviate from SOPs is to significantly reduce the time budget. In aviation, the time budget and the remaining fuel strongly correlate with each other.

METHOD

Research questions

The main research question (1) of this study is whether the provided procedures and checklists adequately support pilots in abnormal situations. This question addresses two different issues: whether these procedures are appropriate for safe flight conduction in relevant situations and whether the provided material (paper-based or electronic checklists) is adequately usable for use in abnormal situations? The first aspect is addressed in this study and the latter aspect is not directly addressed, but corresponding observations are documented. The authors hypothesize that the provided checklists and procedures are not suitable for time-critical situations with abnormal events like technical problems of an aircraft.

A corresponding research question (2) is whether the level of practice and training or the experience of pilots has an effect on pilots' capability to successfully manage a time-critical scenario during an abnormal event. It can be assumed that experienced pilots make better decisions than pilots with less experience. The opposite may be true for manual flying skills where pilots with a high level of practice and training are assumed to perform better than pilots with low levels. Pilots' level of practice and training is negatively correlated with their experience (Haslbeck, Schubert et al., 2012). While first officers on the short-haul only have little experience, they are well practiced and trained because of their frequent flight operations. On the other hand, captains on the long-haul are very experienced, but since they conduct less flight operations in the same time period, they suffer from a rather low level of practice and training.

Operationalization and measurement

In the following section, this paper introduces an experimental design to evaluate the usefulness of some abnormal procedures and checklists. This experiment was performed in cooperation with a major European airline within two qualified full-flight simulators (JAR-FSTD A) and professional airline pilots as participants.

According to the stated research question addressing the handling of checklists (1), eye-tracking data are expected to show the direct interaction with these lists and represent the data acquisition of the relevant information. After perceiving and processing environmental information, communication is the next step in pilots' information processing to maintain a shared or team mental model. Helmreich and Sexton (2004) found differences in the use of first person plural between captains and first officers. Other indicators are the amount of spoken words and density of spoken words of the two crew members and the distribution between them. Furthermore, the content of the spoken speech acts are of concern. Frequently asked questions or speech acts of dissent for instance may indicate an ambiguous implementation of checklists in cockpit. Therefore, all audio channels have to be recorded separately and synchronously. This allows a separate analyzing for both pilots later on. Non-verbal communication has to be recorded by video cameras as well. Also here, when analyzing shared team mental models, the gaze behavior of both pilots has to be recorded. When performing the different checklist actions, pilots on one hand interact with each other but just as well with the aircraft. This interaction can be observed on video data basis and even more precisely when recording the aircraft parameter like certain switches' position or control inputs – to name only a few. The flight parameters additionally help to answer the further research questions such as manual flying qualities, which might become a concern later on.

According to the described operationalization, the following measurement should be incorporated: both pilots' gaze behavior is tracked by two synchronously recording eye tracking devices, three audio channels are recorded (CPT, FO, ATC) simultaneously, and flight parameters are recorded by the flight simulator data recorder. To ease the understanding of the pilots' interaction later on, cameras record from two different view angles. The level of practice and training can be easily assessed by the help of questionnaires.

Further requirements

To measure operators' performance in professional domains, Haslbeck, Schubert et al. (2012) have stated some requirements to be taken into account for finding variance in the experimental results. An important recommendation is the need for non-voluntary (randomized sample) participants to avoid a self-selection bias. One lesson learned from the abovementioned study is a distinct self-selection bias that occurs when selecting only volunteer participants, especially when there is a chance to fail the experiment. In the first announcement of this earlier experiment (Haslbeck, Schubert et al., 2012), pilots were asked to participate on a voluntary base. Afterwards, additional pilots were randomly chosen and mandated to participate by the respective airline. Some of the participants had also answered to the announcement for the volunteer participation and after a review of the results, it turned out that the original volunteer pilots were amongst the best rated participants. In addition, a realistic setting with pilots in uniform, real recorded communication from ATC and foreign aircraft were created for this study. This was done to maintain the appearance and feel of a real flight to evoke a pilot's typical behavior in flight.

Technical flight scenario

To incorporate the aspect of different levels of practice and training, the experiment has to be conducted on two different fleets: long and short haul. As usual for the fleets but also a foreign destination, the long haul crews had to perform a non-precision landing scenario at New York's John F. Kennedy (KJFK) airport in the USA in an Airbus A340-600 aircraft (with four engines); the short-haul crews (Airbus A320-200, with two engines) had to perform a similar non-precision approach at Nice Côte d'Azur (LFMN) airport in France (see Figure 1).

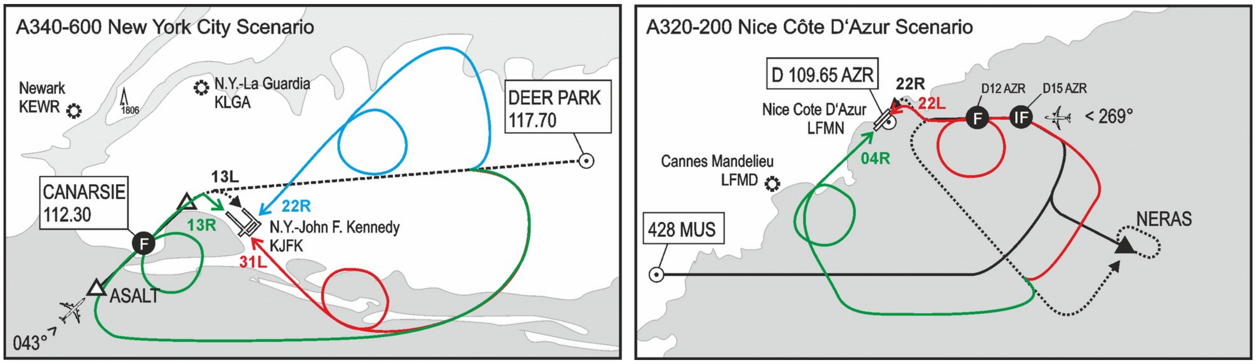

Figure 1. New York City (left) and Nice Côte D'Azur (right) flight scenario (with exemplary flight tracks)

The weather conditions in both scenarios (A320, A340) were moderate with steady wind from south (180°/10kt) and light rain (runways *slippery when wet*). The crew entered the simulator, which was already prepared for the approach (flight management system programmed for a non-precision approach, autopilot setup in *managed* mode and *altitude hold* mode, auto thrust in *speed* mode, navigational and radio setup prepared). The aircraft was already flying with activated position freeze (like a helicopter mode) and realistic ATC in the background. When the crew had taken place and completed their individual settings and adjustments, the scenario began with position freeze released. Fuel on board lead to a remaining flight time of approximately one hour (A340=7,500kg and A320=2,500kg); both scenarios started in 3,000ft altitude with the flaps in position two and appropriate speed (A340=190kt and A320=170kt). The crew then reduced speed for the approach and, with the gear lever selected down by the pilot monitoring, a malfunction of the green hydraulic system was invoked. The green hydraulic fluid was lost within ten seconds and the nose gear remains unlocked as a first consequence. The crew receives cautions from the affected hydraulic system and a master warning because of the unsafe landing gear (main gear deployed and nose gear retracted) after 30 seconds. A go-around had to be performed and the gear remained down because green hydraulic system was needed for retraction. In New York the A340 crew followed the standard missed approach procedure climbing 4,000ft to DEER PARK (see Figure 1 dotted line on left side); in Nice the A320 crew has turned left climbing 2,000ft for a hold at NERAS (see Figure 1 dotted line on right side). With gear down, the fuel consumption was two times higher than normal, which halves the remaining flight time (see Table. 1) and caused additional time pressure for the crew.

Table 1: Approximate time schedule for both scenarios (A340/A320)

Event	**Time**	**Remaining Flight Time (from Fuel)**
start of the scenario	0 min.	60 min. left
gear down, loss of green hydraulic	3 - 10 min.	57 - 50 min., with gear down shorten to 25 min. (<50%)
status cleared, ready for second approach	15 min.	20 - 15 min. left
slats/flaps jammed	17 - 19 min.	12 - 10 min. left
touch down, end of the scenario	30 min.	approx.. 5 min. left

With the well-known aviation principle *Aviate, Navigate, Communicate than Manage the Systems* the electronic centralized aircraft monitoring system (ECAM) for managing the systems came now into the crews' focus. They had to execute the necessary ECAM actions but also get deeper into the time consuming emergency checklist in the onboard *Quick Reference Handbook*. Under time pressure, because of the unsafe and also noisy gear with high fuel consumption, the crew had to decide which airport is appropriate for the upcoming emergency landing. Therefore, the inflight landing distance procedure (ILD) had to be performed. With only 20 minutes of fuel remaining, the crew had to declare an emergency at the latest with minimum diversion fuel with a commitment to stay (in New York a diversion to Newark is possible but also time consuming with no further improvement, same situation in Nice with Marseille or Genua). With four runways in KJFK (13R/31L, 13L/31R, 04L/22R, 04R/22L) and two in LFMN (04L/22R, 04R/22L) the crew had also to choose a safe kind of approach (precision with tail wind or non-precision with possible second go-around) and proper runway length, which fits best with the current situation (see Figure 1 for possible flight tracks and Figure 2 for airport map and available runway lengths). ATC assisted the crew with available information about weather and runways. The cabin crew asked for an announcement from the cockpit to the passengers to improve the feeling of reality in such a situation. According to their airport and runway decision, a second approach was planned. Consider automatic terminal information service (ATIS), 13L and 22R in KJFK were available runways under the current wind and airport situation. Because of a displaced threshold on 22R, only 2,638m runway was available for landing but additional 821m could have been used if the crew would have required full runway length (they had to undershoot the glideslope; see Figure 2 left side). Runway 31L with 3,428m (also with a displaced threshold and additional 994m) was offered by ATC with a slight tailwind component as well. Runway 13R with 3,800m was also offered with a second longer non-precision approach (CANARSIE). For Nice Cote D'Azur airport the crew could choose between a second non-precision approach on runway 22L (which is longer than 22R; see Figure 2) or a longer but precision approach on 04R with also a slight tailwind component.

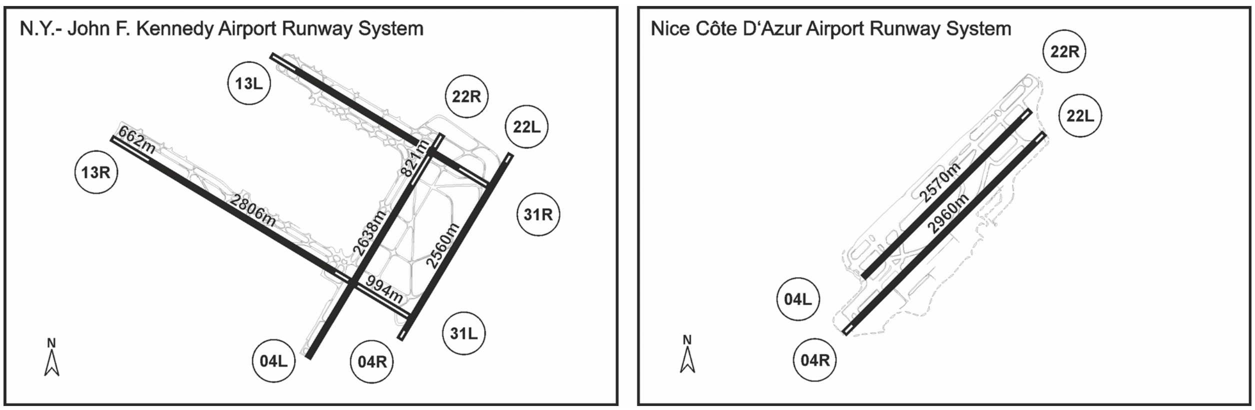

Figure 2. New York City (left) and Nice Côte D'Azur (right) airport map with runway system

A second sequenced malfunction was invoked when the crew was setting slats (or flaps if not in clean configuration): The missing green hydraulic system also slowed down slats and flaps movement so that an additional subsequent error occurred. With dissymmetrical slats and flaps movement, the wing tip brake came active and stopped the slats or flaps in a jammed position. This subsequent error with an occurrence probability of about 10^{-7} is irrevocable and can only be resolved by the maintenance on the ground. A second missed approach or, under increasing time pressure from the remaining fuel, a circling (as shown in Figure 1) had to be flown by the crew. After informing ATC, the crew had again to complete (clear) the ECAM system displays. In this situation, the crews decided whether to skip some steps from the QRH procedure and to shorten the second ILD checking or to try to sequence these procedures fully. The scenario ended after approximately 30 minutes (about 5 min remaining flight time; see Table 1) with touching down and aircraft full stop on the runway. If the crew ran out of fuel, the instructor interrupted the scenario shortly before, in order to prevent the participants leaving the experiment with a negative experience.

DISCUSSION AND OUTLOOK

The introduced experimental design is able to show the appropriateness of currently used checklists and potential improving possibilities. One question to address is whether the A340 and A320 scenarios can be compared to each other. When considering the usefulness of abnormal procedures and checklists on both fleets, the study can be regarded as two separate experiments, delivering results for both aircraft types. Both types have different technical procedures in part, but the level of difficulty of the scenario is very equal for both. The crews of both fleets are faced with the same problems: loss of green hydraulic and go-around, ECAM actions and ILD procedure under time pressure, slats/flaps problem during second approach, ECAM actions and ILD procedure under severe time pressure and landing with an abnormal high speed. The only significant difference between both scenarios is given by the two airports: Nice Cote D'Azur has only two runways and there is no additional difficulty by displayed thresholds. Therefore, the results from both fleets can also be compared to each other when thinking about pilots' performance (e.g. in communication) under different levels of practice and training (A340 crews vs. A320 crews). The authors of this paper believe that abnormal procedures and checklist provided will emerge as inappropriate in critical situations with time pressure. Either these utilities cannot be accomplished on time or won't be even started. The provided ILD procedure, supplied as a paper-copy will also lead to errors in use because of font sizes, attention errors or misinterpretations.

Further research questions refer to the influence of communication on the crews' mental models and overall performance. Regarding communication, the Crew Resource Management can be observed and evaluated by the instructor or the pilots themselves (Gontar, Hoermann, Deischl, & Haslbeck, 2014). Besides the above mentioned main research questions, some further aspects of this scenario are to be recorded and analyzed afterwards. One important field of pilots' performance is manual flying. In the presented scenario pilots will be faced with manual flying in the normal manner at the end of the approach. Under abnormal conditions, it's the captain's duty to take over control and serve as the pilot flying. This can lead to a comparison of captains' manual flying skills between both fleets with some limitation, considering the different airports and aircraft types.

ACKNOWLEDGEMENTS

This work was funded by the Federal Ministry of Economics and Technology via the Project Management Agency for Aeronautics Research within the Federal Aeronautical Research Program. The authors want to thank CPT Manfred Müller, the retired Training CPTs Manfred Binder and Peter Cröniger as well as CPT Carsten Schmidt-Moll and Verena Porstner for their contribution to the project and their support during the experimental conduction.

REFERENCES

Casner, S. M., Geven, R. W., & Williams, K. T. (2013). The Effectiveness of Airline Pilot Training for Abnormal Events. Human Factors, 55(3), 477–485. doi:10.1177/0018720812466893

Ebbatson, M., Harris, D., Huddlestone, J., & Sears, R. (2010). The relationship between manual handling performance and recent flying experience in air transport pilots. Ergonomics, 53(2), 268–277.

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

Gontar, P., Hoermann, H.-J., Deischl, J., Haslbeck, A. (2014, in press). How pilots assess their non-technical performance – a flight simulator study. In *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics*.

Haslbeck, A., Schubert, E., Onnasch, L., Hüttig, G., Bubb, H., & Bengler, K. (2012). Manual flying skills under the influence of performance shaping factors. Work: A Journal of Prevention, Assessment and Rehabilitation, 41(Supplement 1/2012), 178–183. Retrieved from http://dx.doi.org/10.3233/WOR-2012-0153-178

Helmreich, R. L., & Sexton, J. B. (2004). Group interaction under threat and high workload. In T. M. Childress & R. Dietrich (Eds.), Group interaction in high risk environments (pp. 9–23). Aldershot: Ashgate.

Nikolic, M. I., & Sarter, N. B. (2007). Flight Deck Disturbance Management: A Simulator Study of Diagnosis and Recovery From Breakdowns in Pilot-Automation Coordination. Human Factors, 49(4), 553–563. doi:10.1518/001872007X215647

VDI guideline VDI 4006-2 (2003). Human reliability - Methods for quantitative assessment of human reliability. Berlin: Beuth Verlag.