5-7-2019

From Crewed to Single-Pilot Operations: Pilot Performance and Workload Management

Anja K. Faulhaber

FROM CREWED TO SINGLE-PILOT OPERATIONS: PILOT PERFORMANCE AND WORKLOAD MANAGEMENT

Anja K. Faulhaber
TU Braunschweig, Institute of Flight Guidance
Braunschweig, Germany

Higher levels of automation have come to replace human roles in the cockpit. Therefore, a further reduction of the crew size from two pilots to one has become an option. Such single-pilot operations (SPO) need to provide at least the same safety standards as today's two-crew operations (TCO). The present study aims at identifying potential issues in pilot performance and workload during SPO as opposed to TCO. Fourteen pilots flew short ILS approach and landing scenarios in a fixed-base A320 flight simulator. A 2x3 factorial design was used with the factors crew configuration (TCO and SPO) and scenario (baseline, turbulence and abnormal). Performance data and subjective workload ratings were collected. The results suggest that workload might be problematic mostly during abnormal situations. The design of adequate support solutions for such situations will be a major challenge for the implementation of SPO.

Commercial aircraft are commonly operated by two pilots – the pilot flying (PF) and the pilot monitoring (PM). This crew configuration could change in the future, considering the current discussion about a possible reduction of the crew size to one pilot. Economic factors are the main drivers for the transition toward these so-called reduced-crew or single-pilot operations (SPO). Airlines want to save costs, gain more operational flexibility and prepare for an expected pilot shortage due to the growing demand for commercial aviation (Bilimoria, Johnson, & Schutte, 2014; Comerford et al., 2013). The reduction of crew size has, in fact, a historical background in commercial aviation. During the past decades, cockpit crews have gradually been reduced from initially five crew members to today's two-crew operations (TCO). So far this 'de-crewing' has not led to any safety issues when it was accompanied by adequate technological support (Harris, 2007). In light of this historical trend and taking into account the ongoing technological progress, a transition to SPO seems like the logical next step.

However, the implementation of SPO will be more complex than the previous transitions from five to two crew members. Eliminating the second pilot means eliminating a part of the redundancy in the cockpit which has been a foundation for safe operations in aviation. Additional support through automation might not be enough anymore to ensure safe flight conditions either (Bilimoria et al., 2014). It has even been proposed that we need a revolutionary approach entailing a complete rethinking of the pilot's role and hence of the allocation of tasks between human and machine (Boy, 2014; Sprengart, Neis, & Schiefele, 2018). Further research is required in this context to form a profound basis for a possible reconfiguration of the flight deck for SPO.

From a human-centered perspective, one of the major challenges in the introduction of SPO is workload (Koltz et al., 2015). Especially during abnormal scenarios, workload can reach critical levels in SPO (Bailey, Kramer, Kennedy, Stephens, & Etherington, 2017; Etherington, Kramer, Bailey, Kennedy, & Stephens, 2016). For normal scenarios, results from previous studies are ambiguous. Bailey et al. (2017) investigated workload during normal TCO and SPO conditions and found that workload ratings for TCO were higher than expected, almost at the same level with the SPO workload ratings. An additional post-test questionnaire did, however, reveal a significant result for the effect of crew configuration on workload ratings. The authors made limitations in their study design responsible for biased workload ratings, leaving open questions regarding the general validity of the study results. As human workload does play an important role in the conceptualization of SPO, these open issues require further investigation.

The present study aims at tackling these open issues and providing a better understanding of workload and performance in SPO. Therefore, a flight simulator study was conducted. The study design was loosely based on Etherington et al. (2016) and Bailey et al. (2017) but a focus on the approach and landing phases of flight was chosen. These are particularly demanding phases for pilots and can be expected to reach critical levels in SPO (Koltz et al., 2015). Additionally, a within-subject design was used to avoid effects of individual differences. The complete design of the study will be explained in more detail subsequently.

Material and Methods

Participants

Fourteen pilots (1 female) participated in the study. They were aged between 26 and 56 years (M = 41.14, SD = 9.44) and their flying experience ranged from 300 to 22000 flight hours (M = 6204, SD = 6271). Five of them were captains and first officers each and one of them was senior first officer. The remaining three participants didn't report a rank because they were not working for an airline at the time. Participation was voluntary and unpaid. The study was performed according to institutional and national standards for the protection of human subjects.

Experiment Design

The study was conducted in a fixed-base A320 flight simulator. A 2x3 factorial within-subject design was chosen. The factors were crew configuration (TCO and SPO) and scenario (baseline, turbulence and abnormal). Participants flew one trial per condition resulting in a total of six experimental trials. The task for each trial was to manually fly short ILS approach and landing scenarios at Frankfurt Airport, runway 25 left. Each trial lasted about 2.5 minutes. The initial situation was always the same: The scenario started 8 nm from the runway at an altitude of about 2600 ft. Airspeed was set to 180 kt, the landing gear was still retracted and the flaps were already extended to 15° (indication 2). The view was clear and there were no clouds. The wind was calm; only in the turbulence scenario moderate turbulence was simulated. In the abnormal scenario, an engine fire was induced when the participants reached an altitude of 1800 ft.

The NASA Task Load Index (TLX) (Hart & Staveland, 1988) was used to assess subjective workload ratings directly after each scenario. It consists of six workload subscales – mental demand, physical demand, temporal demand, performance, effort, and frustration – which are all rated on a scale from 0-100. Qualitative data in the form of video and audio recordings from each session as well as debriefing interviews were collected. Additionally, eye tracking data and simulator parameter were recorded. Only data of the PF sitting in the left seat were collected. The present paper will focus on the analysis of the TLX scores and observed behavior patterns related to performance and workload management.

Procedure

Pilots participated in teams of two. Upon arrival, they were briefed on the experiment, received the material (checklists, charts and a Quick Reference Handbook) and gave informed consent. Afterward, each participant was allowed one or two training trials as PF depending on their prior experience with the A320. If they felt confident after the first training trial, the second one was skipped. The experiment started with the first participant as PF in the SPO condition. The second participant was waiting in the briefing room. After finishing the three scenarios in the SPO condition, the second pilot joined in as PM and together they flew the same scenarios in the TCO condition. Then participants changed roles and seats – the first participant became PM and the second one PF. Now the second pilot flew all three scenarios first in the TCO condition with the PM and afterward alone in the SPO condition. Hence, half the participants started with the SPO condition while the other half started with the TCO condition. After each scenario, the PF completed the NASA TLX. The order of the scenarios was balanced; each participant was assigned a different order. When the experimental trials were completed, a short debriefing interview with both participants concluded the session. The total duration was about two hours.

Results

In order to investigate the effects of the factors crew configuration and scenario on the perceived workload, a two-way repeated measures ANOVA was performed on the NASA TLX data. The level of significance was $p \leq 0.05$. Matlab was used for all analyses.

Workload

The results showed that workload was at the same level for SPO and TCO baseline conditions but trended higher for the turbulence and abnormal conditions in SPO (Figure 1). The baseline condition yielded nearly the same mean values for TCO (M = 37.74, SD = 20.74) and SPO (M = 37.62, SD = 15.11). In the turbulence condition, there was a small difference with means of 37.68 (SD = 13.81) for TCO and 43.15 (SD = 13.39) for SPO. As expected, the abnormal condition received the highest workload scores and the most prominent difference in ratings with mean scores of 50.3 (SD = 14.36) for TCO and 56.9 (SD = 17.88) for SPO.

However, the results from the ANOVA revealed that the effect of the factor crew configuration did not reach significance ($F_{1,13} = 2.54$, $p = 0.135$, $\eta^2_p = 0.163$) and neither did the interaction effect ($F_{2,26} = 1.15$, $p = 0.331$, $\eta^2_p = 0.082$). A significant main effect was found though for the factor scenario ($F_{2,26} = 8.02$, $p = 0.002$, $\eta^2_p = 0.382$). A bonferroni post-hoc test showed that this effect applied only to the abnormal scenario compared to both the baseline ($p = 0.021$) and turbulence scenarios ($p = 0.026$). The post-hoc comparison of baseline and turbulence scenarios did not reach significance.

The subscales of the NASA TLX were additionally analyzed separately to understand which of them were affected most by the factor crew configuration and which contributed most to the overall workload rating. A look at the unweighted mean scores showed that the subscales for mental demand and effort received the highest mean workload scores in general (Figure 2). With the exception of the performance subscale, scores were consistently higher in SPO as opposed to TCO conditions, even though the difference remains relatively small.

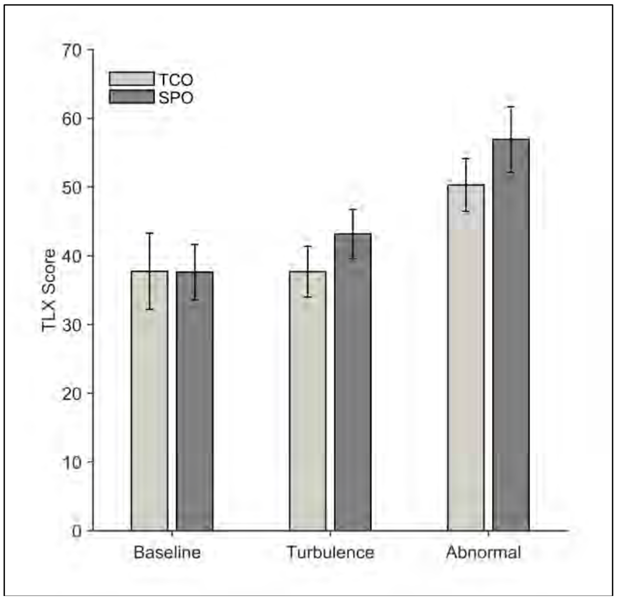

Figure 1. Bar graphs showing NASA TLX unweighted mean composite workload scores representative of the 2x3 factorial design. Error bars show standard errors of the mean.

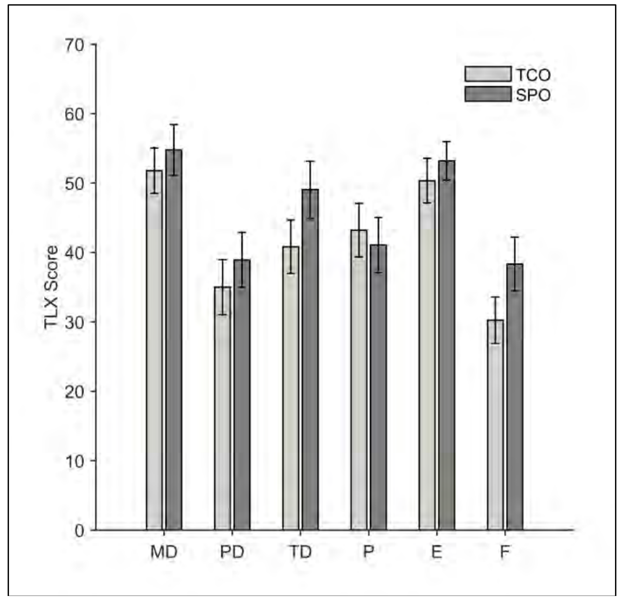

Figure 2. Bar graphs showing the NASA TLX unweighted mean workload scores for each subscale for the factor crew configuration. Error bars show standard errors of the mean.

The NASA TLX mean subscale workload scores for the factor crew configuration are summarized with the respective ANOVA results in Table 1. The subscales for temporal demand and frustration show the highest difference in mean scores. These were also the only subscales for which a significant effect of the factor crew configuration was found. In conclusion, temporal demand and frustration seem to be the subscales affected most by the crew configuration while mental demand and effort contribute most to the overall workload ratings.

Table 1.
Statistics for the NASA TLX unweighted workload scores for the factor crew configuration.

	Mean (SD)		ANOVA		η^2_p
	TCO	SPO	$F_{1,13}$	p	
Mental	51.79 (21.24)	54.76 (23.84)	0.62	0.445	0.046
Physical	35 (25.71)	38.93 (25.72)	0.99	0.337	0.071
Temporal	40.83 (24.86)	49.05 (26.65	8.17	**0.013**	0.386
Performance	43.21 (24.96)	41.07 (25.7)	0.22	0.650	0.016
Effort	50.36 (20.76)	53.21 (18)	0.40	0.536	0.030
Frustration	30.24 (21.75)	38.33 (25.15)	7.23	**0.019**	0.358
Composite	41.91 (17.27)	45.89 (17.26)	2.54	0.135	0.163

Note. Significant effects are highlighted in boldface.

Performance

Qualitative analysis of pilot's behavior patterns during the experiment revealed that participants developed different strategies to manage workload in the SPO condition. The majority of participants (9 out of 14) talked to themselves or called out each step while following the landing checklist. Some of them even made exactly the same calls they were used to from TCO. Thinking aloud was however never mentioned nor asked for during the briefing session and this could hence be interpreted as a way to handle workload. Further analyses of pilot performance showed that checklist usage was more consistent in TCO. When distracted by other tasks, the PF was generally more prone to forget the completion of the landing checklist. An interesting case of this type happened during the abnormal scenario in the SPO condition and led to a crashed landing because the landing gear had not been extended. The participant confused the warning sound indicating that the gear was still retracted with the alarm triggered by the engine fire and became aware of this mistake shortly before touchdown when it was too late. There were similar situations in the TCO condition, where the PF did not actively demand to check the status of the landing checklist because he or she was distracted by other tasks. In these cases, the PM reminded the PF of the checklist and suggested further steps if necessary.

In general, there was no consensus on whether the abnormal procedure for the engine fire should be performed at all during the approach and landing phases of flight. In fact, only four participants performed the procedure in both TCO and SPO abnormal conditions consistently. Six of the participants decided that it would always be best to concentrate on the landing and to disregard the warning completely. They only informed ATC about the situation and cleared the warning. Interestingly, the remaining four participants performed the procedure thoroughly in the TCO condition. In the SPO condition, however, they either decided to disregard the warning or they started the abnormal procedure checklist and aborted before they could complete it. One of the participants even commented that it would be risky to perform the procedure without someone else watching over it.

Discussion and Conclusions

The present study aimed at investigating workload and performance in SPO compared to TCO conditions. Results revealed that workload was not perceived as higher in baseline SPO conditions but only in scenarios involving turbulence or abnormal procedures. This is to a certain extent in line with the results from previous studies (Bailey et al., 2017; Etherington et al., 2016). However, the differences in workload ratings between SPO and TCO were small and several participants even reported after the experiment, that they did not perceive workload as a major issue for SPO. Comparing the NASA TLX ratings to results from other studies shows, though, that the highest mean of 56.9 from the SPO abnormal condition is higher than 75% of all TLX scores obtained from aircraft piloting tasks (Grier, 2016). It can, therefore, be considered as relatively high. In particular the temporal and frustration dimensions of workload were found to be affected most by the SPO condition. Observation of pilot performance also indicated that higher workload did lead to more errors and less accuracy in the completion of tasks, especially during the abnormal SPO condition. Further challenges for the implementation of SPO are therefore to design adequate support for such high workload situations. Several concepts have already been proposed such as a ground operator (Lachter et al., 2014) or a harbor pilot (Koltz et al., 2015). However, further research in this area is required.

Acknowledgements

This research was conducted within the graduate program "Gendered configurations of humans and machines. Interdisciplinary analyses of technology" (KoMMa.G) funded by the federal state of Lower Saxony, Germany. The author would like to thank Prof. Dr. Peter Hecker and Dr. Thomas Feuerle. Special thanks go to the Institute of Flight Guidance at the German Aerospace Center (DLR) Braunschweig, in particular to Tatjana Kapol and Maik Friedrich.

References

Bailey, R. E., Kramer, L. J., Kennedy, K. D., Stephens, C. L., & Etherington, T. J. (2017). An assessment of reduced crew and single pilot operations in commercial transport aircraft operation. *IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), St. Petersburg, FL, USA*. https://doi.org/10.1109/DASC.2017.8101988

Bilimoria, K. D., Johnson, W. W., & Schutte, P. C. (2014). Conceptual framework for single pilot operations. *International Conference on Human-Computer Interaction in Aerospace (HCI-Aero 2014), Santa Clara, CA*. https://doi.org/10.1145/2669592.2669647

Boy, G. A. (2014). Requirements for single pilot operations in commercial aviation: A first high-level cognitive function analysis. *Complex Systems Design and Management International Conference, Paris, France*.

Comerford, D., Brandt, S. L., Lachter, J., Wu, S.-C., Mogford, R., Battiste, V., & Johnson, W. W. (2013). *NASA's single-pilot operations technical interchange meeting: Proceedings and findings*. Moffett Field, CA.

Etherington, T. J., Kramer, L. J., Bailey, R. E., Kennedy, K. D., & Stephens, C. L. (2016). Quantifying pilot contribution to flight safety for normal and non-normal airline operations. *IEEE/AIAA 35th Digital Avionics Systems Conference, Sacramento, CA*. https://doi.org/10.1109/DASC.2016.7778094

Grier, R. A. (2016). How high is high? A meta-analysis of NASA-TLX global workload scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59*(1), 1727–1731. https://doi.org/10.1177/1541931215591373

Harris, D. (2007). A human-centred design agenda for the development of single crew operated commercial aircraft. *Aircraft Engineering and Aerospace Technology, 79*(5), 518–526. https://doi.org/10.1108/00022660710780650

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). Amsterdam: North Holland Press.

Koltz, M. T., Roberts, Z. S., Sweet, J., Battiste, H., Cunningham, J., Battiste, V., et al. (2015). An investigation of the harbor pilot concept for single pilot operations. *Procedia Manufacturing, 3*, 2937–2944. https://doi.org/10.1016/j.promfg.2015.07.948

Lachter, J., Brandt, S. L., Battiste, V., Ligda, S. V., Matessa, M., & Johnson, W. W. (2014). Toward single pilot operations: Developing a ground station. *International Conference on Human-Computer Interaction in Aerospace (HCI-Aero 2014), Santa Clara, CA*. https://doi.org/10.1145/2669592.2669685

Sprengart, S. M., Neis, S. M., & Schiefele, J. (2018). Role of the human operator in future commercial reduced crew operations. *2018 IEEE/AIAA 37th Digital Avionics Systems Conference, London*. https://doi.org/10.1109/DASC.2018.8569803