Effects of performance pressure on learning during recurrent pilot training

Sepulcri, A.; van Paassen, M.M.; Landman, H.M.; Stroosma, O.; Mulder, Max

Publication date
2025

Document Version
Final published version

Published in
Procedings of the 23rd International Symposium on Aviation Psychology

Citation (APA)
Sepulcri, A., van Paassen, M. M., Landman, H. M., Stroosma, O., & Mulder, M. (2025). Effects of performance pressure on learning during recurrent pilot training. In *Procedings of the 23rd International Symposium on Aviation Psychology* (pp. 120-125)

23rd International Symposium on Aviation Psychology

May 27–30, 2025

doi.org/10.5399/osu/1188

Cover image courtesy of Wright
State University Libraries' Special
Collections and Archives

Description:

A view of the Wright Model A Flyer
in flight with a passenger onboard.

Proceedings of the 23rd International Symposium on Aviation Psychology

Online

May 27–30, 2025

Hosted by Oregon State University

EFFECTS OF PERFORMANCE PRESSURE ON LEARNING DURING RECURRENT PILOT TRAINING

Andrea Sepulcri, M. M. (René) van Paassen, Annemarie Landman, Olaf Stroosma, Max Mulder
Aerospace Engineering – Delft University of Technology
Delft, The Netherlands

Recurrent training helps pilots maintain or upgrade their flying skills, however, performance pressure during training may reduce the effectiveness of recurrent training. To study the effects of performance pressure on pilot behaviour and learning, pilots (n=17) holding an airline transport pilot licence and type rating for large multi engine aircraft were invited to a simulator experiment where they would learn a new autopilot system. The low-pressure group flew a series of training scenarios under no performance pressure whereas the high-pressure group flew the same training scenarios with prompts designed to induce performance pressure and raise their anxiety. The high-pressure group did in fact undergo training with a significantly higher state anxiety, This had no significant effect on their behaviour during training or their performance in a test scenario.

Introduction

It was found that learning autopilot logic and behaviour is more efficient when pilots are allowed to learn in an exploratory fashion (van Leeuwen et al., 2024). However, traditional recurrent training has utilised a task-based approach (Burratto & Graef, 2021) emphasising the use of standard operating procedures (SOPs), which, while effective for situations with objectively correct answers, ultimately may limit pilots' problem-solving abilities in unknown situations.

Pilots regularly undergo recurrent training to maintain or upgrade their skills (EASA, 2016). Monitoring pilots during training and imposing consequences on them based on their performance induces performance pressure on pilots which may encourage them to adhere to SOPs rather than utilise genuine problem-solving and diagnosis skills reducing the effectiveness of exploratory training. This indicates the need to investigate the effects that performance pressure has on a pilot's ability to learn during automation training.

The literature on the effects of performance pressure on performance and learning are mixed. As found by Eisenberger and Aselage (2009), expecting a higher reward for high performance promotes a compulsion to perform well, which subsequently leads to increased perceived self-determination and increased intrinsic job interest. In a repeated manual landing task, performance pressure was found to increase anxiety and reduce attention but was ultimately inconsequential to performance (Allsop & Gray, 2014). However, in a stimulus-response task any change in performance pressure was found to have a detrimental effect to performance (Cassell et al., 2018). In an experiment studying learning scientific content under pressure by Hinze and Rapp (2014), performing retrieval practice whilst under performance pressure was detrimental to long-term retention and understanding of scientific content. However, participants were able to overcome performance pressure in a test (Hinze & Rapp, 2014).

We hypothesize both differences in behaviour and training quality due to performance pressure; pilots who underwent training with increased performance pressure will be less inclined to explore and utilise different functions of the autopilot in the training scenarios, and have effectively a lower quality of training, reflected by poorer performance in the test scenarios.

Method

Design, Participants, Equiment

The experiment was conducted as a between-subjects experiment with participants being divided into a low-pressure group and a high-pressure group, who during training received prompts intended to elevate their anxiety into a high pressure state. All participants underwent the test with the same prompts and instructions. Participants (n=17) possessed an ATPL and were type rated for commercial aircraft, and had no experience operating the Garmin G1000 with the GFC700 autopilot which the simulation was based on. The participants were divided into two groups balanced over three factors: age, total flight hours and trait anxiety (STAI-T) measured using the State-Trait Anxiety Inventory from Spielberger et al. (1983). All measures were reported by the pilots ahead of their experiment sessions.

Table 1: *Overview of autopilot modes, ranked from lowest to highest complexity*

(a) *Vertical modes*

Autopilot mode	Description
Pitch hold	Hold the set pitch angle
Altitude hold	Hold the current altitude
Vertical speed	Climb or descend at the set vertical speed
Flight level change	Climb or descend at the set airspeed
Vertical navigation	Follow the programmed vertical profile

(b) *Horizontal modes*

Autopilot mode	Description
Roll hold	Hold the set roll angle
Heading select	Fly at the selected heading
Navigation (VOR)	Intercept the VOR beacon at the set radial
Navigation (GPS)	Follow the programmed flight plan

The experiment was performed in the SIMONA Research Simulator (SRS) at the aerospace engineering faculty at the Delft University of Technology (Stroosma et al., 2003). The aircraft modelled in the simulator was a Piper PA-34 Seneca III which was previously used in other experiments testing startle and surprise, and automation training. The avionics used were based on the Garmin G1000 Primary Flight Display (PFD) and Multi Function Display (MFD) simulation utilised in previous automation training experiments (van Leeuwen et al., 2024). The avionics were extended with features from the Garmin G1000 pilot's guide as needed for the experiment.

General Experiment Procedure

Participants underwent a fifteen minute ground school presentation on the basic operation of the autopilot system, explaining which modes were available and how to activate them, further interactions within the autopilot were left up to the participants to discover. This was followed by the training scenarios in the simulator. The low-pressure group received instructions intended to keep their state anxiety low and the high-pressure group received instructions designed to raise their state anxiety (Appendix). Before each scenario, the participants received a flight plan with waypoints, courses, altitudes and distances that they were instructed to fly to the best of their abilities. The training phase was further split up into two phases. In the first phase, the aircraft flew without any instrument failures to help the participants learn and familiarise themselves with the operation of the autopilot. In the second phase, the participants encountered instrument failures where they were asked to continue flying using the next highest mode of automation available. After each training scenario, pilots were then asked to rate the mental effort required for the scenario using the rating scale mental effort (RSME) from Zijlstra and van Doorn (1985). At the end of the training scenarios, both groups then rated their state anxiety throughout the training phase using the state-trait anxiety inventory state form (STAI-S) from Spielberger et al. (1983) to ensure the pressure manipulation was successful. In the test phase both groups flew the same testing scenarios, after each scenario pilots answered the RSME, and at the end of these scenarios the STAI-S was again completed. A ten-question multiple choice test was used to test automation knowledge after training and testing.

In the ground school, the pilots were told the available autopilot modes in the experiment ranked in terms of complexity as shown in Tables 1a and 1b. The level of complexity was determined based on how much workload and pilot input was needed to operate the mode. To promote learning to interact with the autopilot, participants were instructed to always use the highest functioning mode of automation.

Scenario Design

The design of training scenarios is crucial in order to enable effective evidence-based training (Dahlstrom et al., 2022). Moreover, in this experiment the training scenarios would be the only opportunities for pilots to learn how to operate the autopilot ahead of the testing scenarios. To ensure effective scenarios, no scenarios would have an optimal solution of manual flight. Since pilots are trained to initially disconnect autopilot and fly manually when encountering automation failures, it was decided that scenarios with manual flight as a solution would not provide any meaningful results or differences between the two groups.

Furthermore, all training scenarios had to be linked to one of the eventual testing scenarios. This would allow for knowledge transfer from the training to the testing scenario, therefore rewarding pilots who learned more in the training scenarios and consequently had a better overall knowledge of the operation of the autopilot. Moreover, all scenarios would contain periods of low workload where the primary focus was on monitoring the autopilot followed by a period of high workload where pilots would be required to diagnose and resolve failures. Finally,

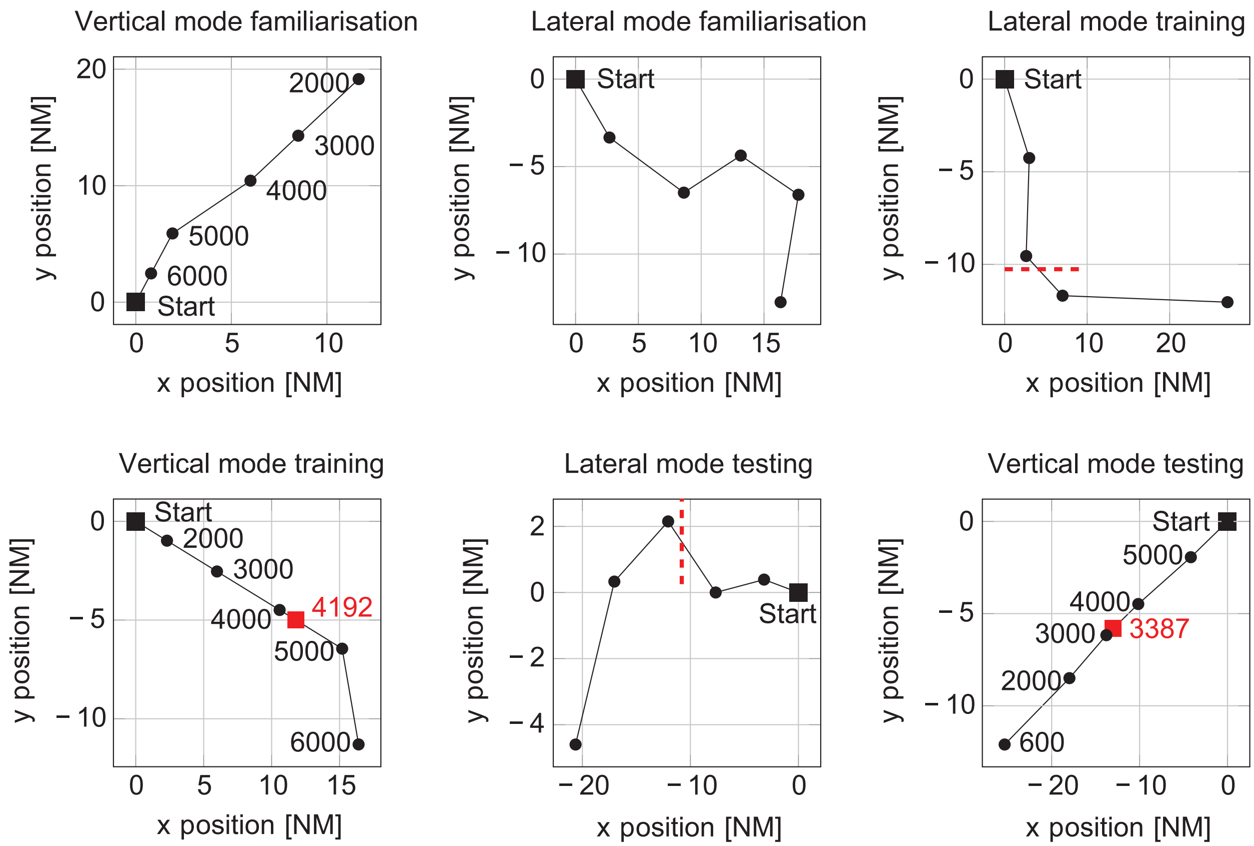

Figure 1: All the training and testing scenarios used in the experiment as viewed from above. Altitude for each node is indicated if the waypoint altitudes were included in the flight plan. The square nodes indicate the beginning of each scenario and the red nodes/lines indicate locations where the instrument failures occur

where possible the scenarios should have multiple possible solutions to the failures where pilots could use different autopilot modes to fly the rest of the scenario, again rewarding pilots who had more effective training enabling them to select the more optimal autopilot mode. Using these guidelines, six autopilot training scenarios were designed. Three scenarios focused on using the vertical modes of the autopilot and three scenarios focused on using the lateral modes. Moreover, the highest available mode of automation at any point in the scenarios was not necessarily the highest autopilot mode overall, therefore encouraging pilots to explore and utilise the different modes. The flight plans for these scenarios can be seen in Figure 1.

Scenarios started with a manual flight for familiarization with the model, controls and displays, and then (1) a vertical mode familiarization, starting in VNAV after which pilots could try the other vertical modes, and a similar horizontal familiarization scenario. Training scenarios promoted interaction with and learning about the autopilot modes. In the lateral mode training scenario, pilots navigated several waypoints using VOR beacons. After the second VOR waypoint the VOR receiver would fail rendering it unusable. Pilots who navigated the waypoints well could then see that a GPS route was also available and navigate the rest of the scenario using GPS, unless by this time drift from the LNAV track was too large, and heading select would be possible.In the vertical mode training scenario, pilots flew a climbing scenario from 2,000 feet to 6,000 feet where at 4,192 feet of altitude the pitot tube would clog causing the airspeed indication to become unreliable and rapidly increase as the altitude increased. If the aircraft was in flight level change, the flight director would begin indicating an increasingly higher pitch angle threatening to stall the aircraft if appropriate action was not taken. Since the altitude tape was still functioning as expected, pilots could then switch modes and complete the rest of the scenario climbing in vertical speed mode.

In the first testing scenario, the pilots flew a route navigating using GPS. Just before the third waypoint, the GPS would silently fail causing the map on the MFD, and the ETA and distance to the next waypoint to freeze. If no

Table 2: *Training and testing scenarios and failures, and where applicable the highest possible mode and corresponding solution after failure*

Scenario	Failure	Highest Mode Available	Highest Mode After Failure
Introductory manual flight	Nominal flight		
Vertical mode familiarisation	Nominal flight		
Lateral mode familiarisation	Nominal flight		
Lateral mode training	VOR receiver failure	Navigation (VOR)	Navigation (GPS)
Vertical mode training	Blocked pitot tube	Flight level change	Vertical speed
Lateral mode testing	GPS failure	Navigation (GPS)	Navigation (VOR)
Vertical mode testing	Blocked static port	Vertical navigation	Pitch hold

action was taken the aircraft would then hold the heading it was flying at and overshoot the waypoint. Pilots would therefore have to recognise that the GPS had failed and take action. This could be seen from the fact that the MFD would be frozen along with the estimated time of arrival and distance from the waypoint never updating, contradicting the information from the distance measuring equipment (DME). Learning to read information about the GPS routes and using the DME was possible in the lateral familiarisation and training scenarios. The two standby VORs were tuned to the final two waypoints of the scenario and therefore the optimal action was to change the navigation source to VOR and use the DME information to ascertain their relative position.

In the second testing scenario, pilots were tasked with a descent from 5,000 feet to 600 feet. At 3,387 feet the static port would clog causing the altitude tape and vertical speed indicator to freeze and the airspeed indicator to increase as the altitude decreased. In vertical navigation mode, this would cause the aircraft to continue descending, overshooting the altitude of the next waypoint. The optimal solution was therefore to disregard the instruments on the PFD and instead rely on the altitude and airspeed indication given in the standby instruments and complete the scenario using pitch hold mode and manually changing the pitch reference to complete the descent. In the vertical training scenario, pilots would have learnt that the flight level change mode would not work with an unreliable airspeed indication and that the vertical speed mode relied purely on the altitude indication making both of those modes unviable.

Measures and Results

Manipulation and Realism

In order to test whether the pressure manipulation was successful, at the end of training all participants were asked to rate their state anxiety using the state-trait anxiety inventory (Spielberger et al., 1983). A Mann-Whitney U Test indicated that the high-pressure group (mean=24.6, SD=3.64) did have a significantly higher state anxiety during training compared to the low-pressure group (mean=30.2, SD=5.83) ($U = 15.0$, $p < 0.0469$).

Many pilots successfully diagnosed the instrument failures in the testing scenarios and stated that similar instrument failures are utilised in recurrent pilot training. Moreover, when asked to motivate their actions in testing, some pilots cited the training scenarios as having provided insight into how certain autopilot modes functioned, influencing their choices in the testing scenarios. This indicated that for some participants, knowledge transfer from the training scenarios to the testing scenarios did occur from the exploratory training.

Exploration

To determine the difference between the two groups in terms of exploratory behaviour, the number of mode changes performed by each group in the familiarisation scenarios and the number of mode changes before failure in the training scenarios were logged. In the scenarios focusing on the lateral modes, only the lateral mode changes were logged and vice versa with the vertical modes. In the vertical and lateral familiarisation scenarios, the low-pressure group switched modes more frequently compared to the high-pressure group. Whereas in the lateral and vertical training scenarios the high-pressure group switched modes more frequently before failures. A Mann-Whitney U test showed that the difference was not significant in the vertical mode familiarisation scenario ($U = 42.5$, $p < 0.562$) or

	Low-pressure	High-pressure	
	Mean (SD)	Mean (SD)	p
Lateral mode testing	86.8 (71.2)	78.9 (60.9)	0.815
Vertical mode testing	16.2 (5.37)	13.5 (9.60)	0.167

Table 3: *Time taken to first switch modes after failure in the testing scenarios*

the lateral mode familiarisation scenario ($U = 40.5$, $p < 0.699$). The difference was not significant either in the lateral mode training scenario ($U = 31.0$, $p < 0.660$) or the vertical mode training scenario ($U = 20.0$, $p < 0.132$).

Problem-solving and diagnosis

To determine the pilots' diagnosis skills, the time to first change autopilot modes after instrument failure was logged. In both the lateral and vertical mode testing scenarios, the high-pressure group switched modes faster than the low-pressure group. A Mann-Whitney U test revealed no differences between the groups in the lateral mode testing scenario ($U = 39.0$, $p < 0.815$) or the vertical mode testing scenario ($U = 51.0$, $p < 0.167$). The summary for these measures is given in Table 3.

To quantify the participants' problem-solving skills, two measures were used. First, how long it took the pilot to select their final mode of automation after the instrument failure and the number of mode changes after the instrument failure. The high-pressure group selected the final mode of automation earlier in the lateral mode testing scenario whereas in the vertical mode testing scenario the low-pressure group selected the final automation mode earlier. A Mann-Whitney U test revealed that the difference between the two groups was not significant in the lateral mode testing scenario ($U = 46.0$, $p < 0.370$) or the vertical mode testing scenario ($U = 32.0$, $p < 0.743$).

When looking at the number of mode changes after failure; the low-pressure group switched between fewer modes in the lateral and vertical mode testing scenarios. A Mann-Whitney U test showed no significant differences however in the lateral mode testing scenario ($U = 34.0$, $p < 0.884$) or the vertical mode testing scenario ($U = 30.0$, $p < 0.593$).

Performance

In the horizontal mode testing scenario, the low-pressure and high-pressure group selected the optimal mode with the exact same frequency. In the vertical mode testing scenario, the high-pressure group selected the optimal mode more frequently. When analysed using Pearson's Chi-Square test. In both the lateral mode testing scenario ($\chi^2(1) = 0$, $p < 1$) and the vertical mode testing scenario ($\chi^2(1) = 0$, $p < 1$) the selection of the final modes was found to be completely independent of the type of training that the participants received.

System knowledge

In the paper test that the participants completed after the experiment, the low-pressure group (mean=5.13, SD=2.03) scored lower than the high-pressure group (mean=5.44, SD=2.00) on average. However, a Mann-Whitney U test showed that the difference was not significant ($U = 33.0$, $p < 0.808$).

Conclusion

The two hypotheses proposed for this study were that pilots who underwent training under a high-pressure manipulation would be discouraged from exhibiting exploratory behaviour during training scenarios and that pilots who underwent training under a high-pressure manipulation would receive a lower quality of training followed by worse performance in test scenarios. Ultimately, the results of the experiment demonstrate that in the context of the experiment; training under a high-pressure manipulation ultimately did not have any significant effect on the pilots' behaviour in training nor their performance in a test or system knowledge. However, it cannot be concluded that performance pressure does not have any effect on learning rather that other factors such as individual differences between pilots and the structure of the training scenarios may have had a large contribution to the variation in the data. The recommendation is therefore to continue the research into the effects of performance pressure on learning.

Acknowledgment

The author would like thank all of the experiment participants especially those sorted into the high-pressure group as well as acknowledge support from technicians from the TU Delft faculty of Aerospace Engineering.

Appendix: Pressure Manipulation

The only difference between the low-pressure and high-pressure group was the performance pressure manipulation that was applied to them during the briefing, ground school and training scenarios. The pressure manipulation was based on interviews with active pilots and from experiences in previous experiments. This pressure manipulation was applied in two phases of the experiment, firstly during the briefing and the ground school, and then again in the training phase.

During the briefing, participants in the high-pressure group were told that they would be monitored and their data would be recorded for analysis throughout the whole experiment, and that the number of training scenarios they would undergo would depend on their performance during the training scenarios. During the training phase, participants in the high-pressure group were once again reminded before each scenario that their data were being recorded for analysis and that the number of training scenarios they underwent was dependent on their performance. Moreover, in the low-pressure group the participants were given all the training scenario flight plans at once whereas in the high-pressure group participants first received the two familiarisation flight plans in a folder with "Training" written on it. After completing those, they were told that due to their performance they would be required to undergo some more training scenarios and received a new folder with "Extra" written on it containing four more flight plans.

At the end of the experiment, all participants were told the real title of the study and the high-pressure group were told that they were sorted into the high-pressure group and the prompts heard in the experiment were unrelated to how they may have actually performed.

References

Allsop, J., & Gray, R. (2014). Flying under pressure: Effects of anxiety on attention and gaze behavior in aviation. *Journal of Applied Research in Memory and Cognition*, *3*, 63–71.

Burratto, F., & Graef, R. (2021). Training pilots for resilience. *The Airbus Safety Magazine*.

Cassell, V. E., Beattie, S. J., & Lawrence, G. P. (2018). Changing performance pressure between training and competition influences action planning because of a reduction in the efficiency of action execution. *Anxiety, Stress and Coping*, *31*, 107–120.

Dahlstrom, N., Cameron, M., & Kennedy, R. J. (2022). Simulation for experiential training (set) as an enabler for evidence based training (ebt).

EASA. (2016). *Part-fcl*.

Eisenberger, R., & Aselage, J. (2009). Incremental effects of reward on experienced performance pressure: Positive outcomes for intrinsic interest and creativity. *Journal of Organizational Behavior*, *30*, 95–117.

Hinze, S. R., & Rapp, D. N. (2014). Retrieval (sometimes) enhances learning: Performance pressure reduces the benefits of retrieval practice. *Applied Cognitive Psychology*, *28*, 597–606.

Spielberger, C. D., Gorssuch, R. L., Lushene, P. R., Vagg, P. R., & Jacobs, G. A. (1983). Manual for the state-trait anxiety inventory. *Consulting Psychologists Press*.

Stroosma, O., van Paassen, M. M., & Mulder, M. (2003). Using the simona research simulator for human-machine interaction research. *AIAA Modeling and Simulation Technologies Conference and Exhibit*.

van Leeuwen, J. K., Landman, A., Groen, E. L., Mumaw, R. J., Stroosma, O., van Paassen, M. M., & Mulder, M. (2024). Using problem-based exploratory training to improve pilot understanding of autopilot functions. *Cognition, Technology & Work*.

Zijlstra, F. R. H., & van Doorn, L. (1985). *The construction of a scale to measure perceived effort*.

International Symposium on Aviation Psychology