A TRAINING STUDY TO IMPROVE MONITORING DURING A GO-AROUND

Randall J. Mumaw
San Jose State University
San José, CA

Amber Villa
NASA Ames Research Center
Moffett Field, CA

Dorrit Billman
NASA Ames Research Center
Moffett Field, CA

Lucas Cusano
NASA Ames Research Center
Moffett Field, CA

The go-around (GA) maneuver varies greatly in line operations from how it is typically trained. To support improved maneuver performance, we developed a training module based on a description of skilled monitoring performance. We had US airline pilots fly GAs in the Pilot Monitoring (PM) role both before and after training. The Pilot Flying (PF), a confederate, flew some GAs poorly to see whether the PM could detect and correct the poorly flown maneuvers. Our data showed that training that focuses on developing clear expectations about system performance can lead to significant improvements in monitoring performance. The PMs trained in the experimental condition outperformed the control group on several important operational measures while flying realistic GA scenarios.

A go-around is a maneuver used to abandon an approach with the goal of trying again or diverting to another airport. In US commercial flight operations in the late 2010s, roughly 20,000 go-arounds (GA) were performed every year, which means, on average, about four GAs occurred per every 1000 operations (ASIAS, 2021).

Although the description of a GA is straightforward, in line operations, the situations where GAs occur can vary greatly. A recent study by Lombaerts et al. (2024) found that 40% of GAs were initiated above 1000 feet (AGL). However, during flightcrew simulator training, GAs are predominantly initiated in the 50-100 feet AGL range, often with one engine, a very different version of the maneuver. Because the GA maneuver is a rare occurrence for any particular pilot, can occur unexpectedly, can impose a significant workload, and is narrowly trained, effective maneuver performance is a concern. Indeed, operational data have revealed a range of undesired outcomes, and a few GAs have ended in fatal accidents (BEA, 2013).

One approach for reducing GA maneuver risk, especially for GAs with higher initiation altitudes, is to improve monitoring to ensure that poor maneuver performance is detected and corrected. Discussions with several airlines did not reveal a prescription for effective monitoring during this high-workload maneuver. Therefore, we developed a description of skilled monitoring performance, targeting the Pilot Monitoring (PM) during maneuver performance.

Our general approach to skilled monitoring emphasizes sense-making activities. Monitoring is cast as a cognitive activity that strives to establish the airplane's current situation in the operational environment to generate appropriate expectations about relevant indications and their values and to identify potential threats to flight path management (see Mumaw et al., 2020 for a broader discussion). Applying this description of skilled monitoring to the GA maneuver, we proposed that skilled monitoring for a GA should be informed by:

- a description of how the airplane indications and the autoflight system will behave during the maneuver, including the time-course of those changes, to establish strong expectations for monitoring flight deck indications,
- the indications that are the most relevant for guiding control inputs as the maneuver progresses, and
- thresholds for triggering the need for control interventions.

Further, a detailed analysis of the GA maneuver identified three different maneuver phases that require a shift in which information should be monitored:

- Pre-TOGA engagement – The period prior to engaging the Takeoff/Go-around (TOGA) mode in which the flightcrew can compare the current airplane state to GA targets.
- TOGA and clean-up – The period where the PM should be monitoring the progress of the GA maneuver, including pitch attitude, power, vertical speed, airspeed, and the necessary changes to the autoflight modes.
- Post-TOGA transition – The period where the airplane has been reconfigured as much as possible, and there is a shift to operating after the GA. Monitoring during this phase is about ensuring that the airplane is configured for this next phase of operations.

From this description of the maneuver and of skilled monitoring performance, we developed a training module for the PM. The study's primary objective was to determine whether training based on our description of skilled monitoring could improve the performance of the PM during a GA maneuver.

Method and Materials

Design and Participants

The study design used a between- and a within-participant factor. Two conditions were created for PM training: *737-specific* training and *Control* training, and pilots were randomly assigned to a training condition. As a within-participant design element, each participant served as his/her own control by performing two matched sets of eight GAs: Set I prior to training and Set II after. The participants were 16 First Officers recruited from US commercial airlines. All were type-rated and current on the 737NG. This study was approved by the NASA IRB.

Materials: Training modules

737-specific training module. The objective of this training module was to provide strong expectations about how the GA maneuver progressed to facilitate more-effective monitoring. After a general introduction, the training addressed the three GA phases (Pre-TOGA engagement, TOGA and clean-up, and Post-TOGA transition), the importance of a "rapid assessment" prior to initiating a GA, and the four questions to guide the rapid assessment. Also, a set of guided video illustrations—using the Primary Flight Display (PFD) and other flight deck displays—were used to show the expected changes to the autoflight modes, airspeed, pitch attitude, and power during the maneuver, along with guidelines on thresholds for calling out performance degradations to the Pilot Flying (PF).

Control training module. The objective of the *Control* training module was to replicate content that is typical of airline training for the GA maneuver. It began with the same general introduction but then focused on standard GA procedure call-outs and actions with several videos of flightcrews performing a GA. Although this training also offered illustrations of displays, it removed all instruction regarding where to look during the maneuver, what performance to expect, and the criteria for calling out performance degradations.

Materials: GA scenarios

Two sets of eight GA scenarios were created. Each scenario started six miles from the runway threshold on an ILS approach to one of four airports. The GA maneuver was triggered by a clearance. Each scenario added PM tasks with the objective of forcing the PM to balance realistic operational demands with monitoring the maneuver.

For four scenarios in each set, the PF, following a script, flew the maneuver poorly such that pitch and/or power were not managed well. To ensure that the participant did not anticipate that each GA would be performed poorly, four scenarios in each set were flown well. Set II scenarios used the same climb distance as those in Set I but changed the airport.

Procedure

After a short briefing on the study activities and collection of demographic data, the participant was taken to a Boeing 737-800 Level D full-flight simulator to fly the Set I GAs. A researcher sat behind the flightcrew and presented ATC clearances during the scenario. Clearances were tied to a maneuver event, such as reaching an altitude.

For the "bad" scenarios, the PF would fly poorly by increasing pitch and/or decreasing thrust, relative to maneuver targets, to keep airspeed low in the early part of the maneuver. Also, in one of the "bad" scenarios, the PF disengaged the autothrottle at the end of the maneuver to let airspeed increase uncontrolled. If the PM spoke up during the maneuver to indicate that they thought there was a problem, the PF would acknowledge and manage the problem. After each GA maneuver, the participant was asked to assign a number to the PF's performance of the maneuver, where 1 was the worst and 5 was the best. The PF was unaware of the rating given.

After a short break, the participant was asked to complete a short, 737-specific test to determine their current understanding of airplane behavior/performance during a GA. Participants then took the assigned self-paced training module. After a second short break, the participant returned to the simulator to perform the second set of eight GAs. Finally, the participant returned to the briefing room and completed the same short test on 737 GA knowledge. Participants were then asked a set of questions to gather their views of the training.

Results

The primary focus of analysis was whether the experimental training improved PM monitoring performance on poorly flown GAs. That is, "Did the experimental training condition change performance significantly more than did the control training condition?" This section (due to space limitations) describes a subset of the performance measures.

The analysis approach assessed likelihood ratio, which compares the fit of a mixed-effects regression model that includes the effect of interest with the fit of a model lacking this effect but otherwise identical. Significant improvement in model fit was measured using χ^2 (chi-square) with its associated probability and the ANOVA function.

Call-outs. In each of the "bad" scenarios, the PF performed the GA maneuver while keeping airspeed close to its value when TOGA was engaged; that is, nearly stagnant. The lower airspeed also led to delays in flap cleanup as the airplane climbed. To be clear, the airplane was not in an upset; this was not an unsafe situation. However, the PF was not following the flight director guidance and was not allowing an efficient flap retraction schedule.

We reviewed the video of each "bad" scenario to determine whether the PM called out the situation in an appropriate manner. Sometimes, the PM comment was diagnostic, e.g., "your airspeed is decreasing; pitch down a little bit." Other times, the PM made a simple statement to call the PF's attention to the situation, e.g., "check airspeed." All of these responses were scored as a call-out. Table 1 shows the number of times that the PM called out a problem to the PF.

Table 1

Comparison of PMs with Control training and 737-specific training who called out the PF's poor performance on the GA maneuver (number of opportunities in parentheses).

Control	Set I	Set II
Total	5 (28)	7 (32)
737-specific	Set I	Set II
Total	7 (29)	23 (32)

PM performance on call-outs improved from Set I to Set II for both training conditions. The improvement was far greater for the *737-specific* condition, more than tripling, than for the *Control* condition. These differences were evaluated using the likelihood ratio test described above, which showed that the effect of the interaction of 'training condition x set' was significant, $X^2(1)=4.68$, $p=.03$.

Time after ALT ACQ; time after ALT HOLD. During the GA maneuver in the Boeing 737, as soon as the autopilot vertical mode changes from TOGA to a different mode, an airspeed target gets set on the mode control panel (MCP). That target, if low, limits acceleration after the GA maneuver when the pilot would like to set up for a second approach. Pilots typically want to be around 210-220 kts for this maneuvering. The way the GA maneuver was flown, however, often resulted in an MCP airspeed target of 172. An effective PM should be quick to notice the low airspeed target and manage it so that the airplane is not flying slowly for an extended period.

One trigger for changing the airspeed target is when it is initially set (the MCP airspeed window changes from blank to a value), which occurs at ALT ACQ. A second trigger occurs when the airplane levels off (ALT HOLD mode). We looked at the time from each trigger until the PM reached up and dialed a higher airspeed. Note that there were negative "ALT HOLD times" for some participants, which means that the PM was changing airspeed before the airplane leveled off. Table 2 shows the time from initial MCP value (at ALT ACQ) to dialing airspeed. Table 3 shows the time from reaching the level-off altitude (at ALT HOLD) to dialing airspeed.

Table 2

Average time (seconds) after the onset of ALT ACQ before the PM started revising the MCP airspeed target, by training condition and set.

Control

Set I			Set II		
Avg	SD	n	Avg	SD	n
26.15	14.8	27	27.27	13.84	30

737-specific

Set I			Set II		
Avg	SD	n	Avg	SD	n
36.3	15.35	27	28.77	14.84	31

Table 2 shows that group performance improved from Set I to Set II for the *737-specific* training condition only; response times got slower for the *Control* condition. These differences were evaluated using the likelihood ratio test and showed that the interaction of 'training condition x set' was marginally significant, $X^2(1)=3.51$, $p=.06$.

Table 3

Average time (seconds) after the onset of ALT HOLD before the PM started revising the MCP airspeed target, by training condition and set.

Control

Set I			Set II		
Avg	SD	n	Avg	SD	n
6.33	15.24	27	10.37	12.93	30

737-specific

Set I			Set II		
Avg	SD	n	Avg	SD	n
17.81	17.43	27	11.74	15.51	31

Table 3, like Table 2, shows that group performance improved from Set I to Set II for the *737-specific* training condition only; response times got slower for the *Control* condition. Note that times were initially lower for the *Control* condition, but training did not improve performance. These differences were evaluated using the likelihood ratio test and showed the interaction of 'training condition x set' was significant, $X^2(1)=4.21$, $p=.04$.

We also found significantly improved performance for the *737-specific* training group on improved maneuver performance and the GA knowledge test.

Discussion

The data analysis identified significant performance improvements tied to the experimental training condition—the 'set x training condition' interaction. A note about the "rapid assessment" method that was trained as a way to tailor expectations to the parameters of the GA being flown: Although not one of the eight participants who had been trained on the rapid assessment method attempted to apply it aloud in the Set II maneuvers, the idea showed high face validity in the subjective reviews of the training. Specifically, all eight *737-specific* trainees were positive about the idea, and felt that it could be done quickly and that it asked a useful set of questions. The reluctance to apply it during the Set II scenarios seems to reflect the concern that this type of communication should be established as SOP before it is performed in operations; participants were unwilling to introduce it unrehearsed into cockpit communications in a study environment.

The subjective reviews of the training also revealed that participants were positive about the challenge of an untrustworthy PF. In a strong sense, to evaluate the PM, the PF must perform poorly. Several of the participants expressed appreciation for the challenge of identifying and correcting the maneuver. An element of their experience, which may not occur in an airline setting, is that they were told that the PF may not fly well, and thus, were given permission to call out PF errors. This is not a common approach to airline training.

Conclusion

This study showed that training that focuses on developing clear expectations about system performance can lead to improved monitoring performance. The PMs trained in the *737-specific* condition outperformed the *Control* group on several important operational measures while flying realistic GA scenarios.

References

ASIAS (May, 2021). Approach, landing, and go-around (ALG) directed study. Enclosure 1 to Letter P231-L21-20.

Bureau d'Enquetes et d'Analyses (2013). Study on aeroplane state awareness during go-around. BEA Safety Study. https://skybrary.aero/sites/default/files/bookshelf/2404.pdf

Lombaerts, T., Mumaw, R.J., & Zaal, P. (2024). Statistical analysis of recent go-around flight data to study and evaluate skilled pilot monitoring. Proceedings of the annual meeting of AIAA AVIATION Forum, Las Vegas, NV.

Mumaw, R.J., Billman, D., & Feary, M. (2020). Analysis of pilot monitoring skills and a review of training effectiveness. NASA/TM-20210000047.