

THE EFFECT OF FEEDBACK ON THE ACCURACY OF CHECKLIST COMPLETION DURING INSTRUMENT FLIGHT TRAINING

WILLIAM G. RANTZ, ALYCE M. DICKINSON, GILBERT A. SINCLAIR, AND
RON VAN HOUTEN

WESTERN MICHIGAN UNIVERSITY

This study examined whether pilots completed airplane checklists more accurately when they receive postflight graphic and verbal feedback. Participants were 8 college students who are pilots with an instrument rating. The task consisted of flying a designated flight pattern using a personal computer aviation training device (PCATD). The dependent variables were the number of checklist items completed correctly. A multiple baseline design across pairs of participants with withdrawal of treatment was employed in this study. During baseline, participants were given postflight technical feedback. During intervention, participants were given postflight graphic feedback on checklist use and praise for improvements along with technical feedback. The intervention produced near perfect checklist performance, which was maintained following a return to the baseline conditions.

DESCRIPTORS: feedback, flight checklists, pilot training

In aviation, the checklist is used during different segments of flight to sequence specific, critical tasks and aircraft configuration adjustments that correspond to specific environmental demands (Degani & Wiener, 1990). It is divided into subsections with task checklists that correspond to all flight segments and, in particular, critical segments such as takeoff, approach, and landing.

The complexity of these task checklists cannot be overstated. Standard procedures common to some cockpits are not compatible

with other cockpits or with newer generation cockpits. In addition, the checklists can be very long. For example, on some checklists, the “before engine start” subsection has 76 items for the first flight of the day and 37 items for subsequent flight segments (Degani & Wiener, 1990). Thus, it is not surprising that many aviation experts have addressed their importance and design, as well as the practices and policies that surround their use (Adamski & Stahl, 1997; Degani, 1992, 2002; Degani & Wiener; Gross, 1995; Turner, 2001; U.S. Federal Aviation Administration [FAA], 1995, 2000). Even so, the incorrect use of flight checklists is still often cited as the probable cause or a contributing factor to a large number of crashes (Degani; Degani & Wiener, 1993; Diez, Boehm-Davis, & Holt, 2003; Turner). Similarly, many investigations by the U.S. National Transportation Safety Board (NTSB) have revealed that the aircraft were not properly configured for flight, which usually results from improper use of checklists (NTSB, 1969, 1975, 1982, 1988a, 1988b, 1989, 1990, 1997).

Studies by Lautmann and Gallimore (1987) and Helmreich, Wilhelm, Klinect, and Merritt (2001) provide more direct evidence of improper use of checklists by flight crews. In a

Reprints and materials including the checklists, technical flight pattern parameters and narration, checklist performance protocol, and checklist observation form can be obtained from William Rantz, Western Michigan University College of Aviation, 237 Helmer Road, Battle Creek, Michigan 49015 (e-mail: William.rantz@wmich.edu).

We thank colleagues and students who provided assistance and support that made this study possible. These include Brad Huitema, Vladimir Risukhin, John Austin, Sean Laraway, Susan Syznerski, Axel Anderson, Brandon Jones, and Alex Merk. We also thank Dave Gaurav and James Burgess for their technical support in the experimental laboratory.

Address correspondence to Ron Van Houten, Department of Psychology, Western Michigan University, Kalamazoo, Michigan 49008 (e-mail: ron.vanhouten@wmich.edu).

doi: 10.1901/jaba.2009.42-497

study funded by the Boeing aircraft manufacturer, Lautmann and Gallimore surveyed 12 airlines and concluded that procedural errors involving use of the checklist contributed to a substantial number of aircraft crashes and incidents.

In an effort to identify particular errors flight crews commit, the U.S. National Aeronautics and Space Administration sponsored a series of studies in which crews were observed while flying. Observers using the line-oriented safety audit (LOSA) that recorded checklist behaviors throughout the flight (Helmreich, Klinect, Wilhelm, & Jones, 1999; Helmreich et al., 2001) recorded crew errors. Between 1997 and 1998, LOSAs were conducted at three airlines with 184 flight crews on 314 flight segments (Helmreich et al.). Seventy-three percent of the flight crews committed errors. The number of errors ranged from 0 to 14 per flight, with a mean of 2. Rule-compliance errors were the most frequently occurring errors, accounting for 54% of all errors (Helmreich et al.). Checklist errors constituted the highest number of errors in this category.

Despite widespread recognition that checklist errors occurred relatively frequently and were major contributing factors to many crashes, the design of checklists "escaped the scrutiny of the human factors profession" until the 1990s (Degani & Wiener, 1993, p. 28). Degani and Wiener (1990, 1993) observed flight crews while flying, interviewed flights crews from seven major U.S. airlines, and analyzed how the design of checklists contributed to aircraft crashes and incidents that were reported in three aviation databases. Their analytic guidelines became the industry standard (Patterson, Render, & Ebright, 2002).

Although Degani and Wiener (1990) did not pursue the behavioral factors that influence checklist use, they recognized their importance, indicating that safety culture issues related to support of misuse or nonuse of checklists were a core problem that led some pilots to misuse the checklist or not use it at all. They also noted

that the promotion of a positive attitude toward the use of the checklist procedure was an important element that was often overlooked. Regardless, an extensive search of the aviation checklist literature did not reveal any studies that have examined whether behavioral interventions could increase the appropriate use of flight checklists.

Although a number of behavioral studies have employed checklists as part of or the sole independent variable in a treatment plan (Anderson, Crowell, Hantula, & Siroky, 1988; Austin, Weatherly, & Gravina, 2005; Bacon, Fulton, & Malott, 1982; Crowell, Anderson, Abel, & Sergio, 1988; Shier, Rae, & Austin, 2003) only a few studies have focused on checklist use as a dependent variable (Burgio, Whitman, & Reid, 1983).

In aviation, incorrect checklist use can lead to fatal consequences. In addition, the completion of checklists during flight is more behaviorally challenging than in the settings in which checklist use has been evaluated, due to constantly changing environmental demands, distractions, and schedule pressures. For example, in one fatal crash, the taxi checklist was not completed because of several interruptions (new weather information, checking aircraft and runway data; Degani & Wiener, 1990). Yet, to date, no study has examined whether behavioral interventions can improve checklist use. Performance in aircraft simulators is one method that could be used to evaluate pilots' use of checklists. In recent decades, personal computer aviation training devices (PCATDs) have emerged as an effective, low-cost platform for training instrument flight skills. For the last 10 years, researchers have demonstrated positive transfer from PCATDs to the actual aircraft (Taylor et al., 1999). Thus, up to 10 hr of simulated flight experience gained while using an FAA-approved PCATD can be applied toward qualifying for certain pilot ratings under Part 61 or Part 141 of the federal aviation regulations (FAA, 1997).

One advantage of simulation training is that it allows more complete monitoring and feedback of pilot behavior. For that reason, training was performed on a PCATD in this experiment. Also, in the future, it is likely that early flight training will begin in a simulator, allowing good checklist performance to be established before further training in the aircraft. Finally, it is likely that the operation of aircraft will increasingly be done by pilots who sit at a console on the ground. Many pilots in the armed forces already fly attack and surveillance unmanned aerial vehicles (UAV) in this manner, and some civilian use may follow (Bone & Bolkcom, 2003). Checklist procedures need to be followed in these situations to prevent crashes. Operators of UAVs are trained on simulators and actual flight is indistinguishable from simulation. Therefore, in this case, transfer would not be an issue. Given the dearth of studies on behavioral interventions to improve pilots' use of checklists, the current experiment examined whether postflight graphic feedback and verbal praise could increase the accuracy and quality of checklist use by pilots during training on a flight simulator.

METHOD

Participants

Participants were 8 undergraduate students (7 male and 1 female) who were enrolled in commercial flight courses in the aviation flight science program at Western Michigan University. The age of the participants ranged from 20 to 26 years. Recruitment flyers and in-class announcements were used to notify potential participants of the opportunity to volunteer for the study. Criteria for inclusion included a private pilot certificate, instrument rating, and a minimum of 2 hr PCATD experience flying some type of instrument landing approach.

A private pilot certificate and instrument rating were prerequisites for performing the simulated instrument flight patterns used as the experimental task. Instrument flight refers to the use of

flight instruments to maintain straight and level flight, turn, climb, and descend while vision is obscured by clouds, precipitation, or other environmental conditions. The FAA requires that pilots have a minimum of 125 flight hours before they can obtain instrument rating; thus, all participants had these minimum flight hours. The minimum 2 hr of PCATD experience ensured that participants had some understanding of how the flight software program functioned and what responses were required to perform technical flight skills on the PCATD, enabling them to perform fluently sooner than those who would not have had such exposure.

Setting

The experimental setting was a room (3.6 m by 4.9 m) that was used as the PCATD flight laboratory. The laboratory was located in a hanger adjacent to Western Michigan's Flight Education Center in Battle Creek. Within the room, dividers restricted the vision of the participant to the PCATD flight simulation testing equipment.

Apparatus

PCATD equipment. The PCATD equipment consisted of a Pentium II 300-megahertz processor, four megabytes of SRAM video memory, and 64 megabytes of SDRAM memory. Other PC equipment included a monitor (25.4 cm by 35.6 cm) and two JUSTer SP-660 3D speakers. Operating software was Microsoft Windows 95, and the simulation software was On-Top Version 8. Flight support equipment for the PCATD included a Cirrus yoke, a throttle quadrant, an avionics panel, rudder pedals, and a number of additional controls used to configure the aircraft. The On-Top software permitted the simulation of several different aircraft. The aircraft that was simulated in the current study was the Cessna C-172. The Cessna was chosen due to its vast popularity in the flight-training field as well as the fact that it was the primary aircraft used in the Western Michigan training fleet.

Technical flight parameters, which depicted how well participants flew the designated flight patterns vertically and horizontally, were recorded for each flight. The simulation software automatically recorded these technical parameters and enabled them to be printed.

Flight patterns. There were six different flight patterns. Each flight pattern was divided into six segments: (a) pretakeoff, (b) climb, (c) cruise, (d) prelanding, (e) landing, and (f) after landing. Each pattern took approximately 15 min to complete. To simulate actual flight patterns realistically and to ensure that the patterns were flown in a consistent way across trials and participants, the experimenter provided typical air traffic control instructions throughout each flight pattern. These instructions were transmitted using an intercom.

The flight checklist. The flight checklist contained 40 checklist items divided into sections that corresponded to each of the six flight segments. This checklist, based on the checklist for the Cessna 172 R, was similar to the one used in Western Michigan's flight-training curriculum. The checklist was mounted in plain sight 25 cm to the left of the flight computer monitor. A paper checklist (rather than an electronic checklist) was used because paper checklists are the most common type of checklist used in aviation and in other industries for complex processes (Boorman, 2001).

Dependent Variables

The main dependent variable was the number of checklist items completed correctly per flight. Two secondary dependent variables were the percentage of total errors for each of the six flight segments during each experimental phase (baseline, feedback, and withdrawal of treatment) per participant and the percentage of baseline trials in which participants performed each of the checklist items incorrectly.

For an item to be scored correct, participants had (a) to respond to the correct flight equipment, (b) to respond appropriately with respect to that equipment, and (c) to respond at

the appropriate time of the flight segment. For example, if the checklist item required turning the heading indicator to the direction corresponding to the compass reading, and the participant turned the heading indicator (the correct equipment) to the corresponding compass heading position (the correct response), the item was scored correct. However, if the participant turned the heading indicator (the correct equipment) to the wrong heading (an error), the item was scored as an error. If the participant turned the heading indicator (the correct equipment) to the corresponding compass heading position (the correct response) at the incorrect time in the flight or checklist sequence, the item was also scored as an error. An error was also scored if a checklist item was omitted.

The first and third authors, both experienced pilots, served as observers. They scored checklist behaviors using an observation form that included each item and columns to check whether the item was completed correctly, completed incorrectly, or omitted. The observers occupied a room adjacent to the participant's room and observed participants remotely via Web cameras. The two cameras had built-in microphones that allowed the observers to see and hear both the nonverbal and verbal responses that were required to complete the checklist. One camera was mounted on the computer monitor approximately 51 cm in front of the participant to capture hand and arm movements. The other was positioned 89 cm behind the participant to observe the participant's interaction with the flight panel. To observe the frequencies entered into the communication and navigation radios, which could not be seen clearly via the cameras, observers viewed a dual computer monitor that mimicked the participant's computer screen. The settings on the computer screen mimicked the setting on the radio stack, which was manipulated by the pilot. All flights were recorded and stored digitally for the purposes of conducting interobserver agreement checks.

Independent Variable

The independent variable was the presence or absence of postflight (a) graphic feedback on the total number of checklist items completed correctly per flight, (b) graphic feedback on the number of items completed correctly and errors for each of the six flight segments per flight, and (c) praise for improvement in the number of checklist items completed correctly. Procedural details are described below.

Experimental Design

A multiple baseline design across pairs of participants with a withdrawal of treatment was used. Sessions lasted approximately 1 hr, and participants flew three different flight patterns per session. Each flight was considered a trial. The order of exposure to the six flight patterns was randomized in blocks of six for each participant. A withdrawal phase was included to assess whether checklist performance would be maintained after postflight feedback was withdrawn.

Procedure

Baseline. Participants were told that the PCATD aircraft was not programmed for any system failures and that the flight pattern was a radar-vectored instrument flight, with an instrument landing system approach to a full-stop landing. They were also told that their behavior during the flight would be observed and recorded using Web cameras. They were then shown the flight checklist and asked to use it as they would during regular flights. In addition, they were told that the experimenter would provide them with some postflight information after each flight and that it would take him about 3 to 5 min to prepare that material. They would thus have a short break after each flight. Although this break was not necessary to provide the postflight technical information during this phase, it was necessary to permit the experimenter to summarize the participant's checklist performance during the intervention phase. Thus, the same postflight break was

scheduled during this phase as well. After instructing the participant, the experimenter left the room.

After the participant completed a flight, the experimenter printed out a technical diagram of the flight pattern flown by the participant. This diagram was automatically created by the simulator software and displayed the lateral and vertical flight paths. The experimenter then entered the room, gave the diagram to the participant, and discussed the technical merits of the flight, praising good performance. No feedback was given to the participant about the use of the flight checklist. This protocol was repeated for each flight during the baseline phase.

Postflight checklist feedback. In addition to giving participants postflight technical feedback, the experimenter provided feedback on the use of the flight checklist. After each flight, the experimenter printed a line graph that displayed the number of correctly completed items for that flight and each preceding flight, including baseline. The experimenter also printed a bar graph that displayed the number of items completed correctly and the number of errors for each of the six flight segments for that particular flight. As in the baseline phase, the experimenter printed out a technical flight diagram as well. The experimenter then entered the room. He showed the technical flight diagram to the participant first and discussed the technical merits of the flight. He then showed the two checklist feedback graphs to the participant and praised any improvements. Praise was provided only following improvements. This protocol was repeated for each flight.

Withdrawal of treatment. Feedback was no longer provided for use of the flight checklist after each flight. This phase was identical to the baseline phase.

Interobserver Agreement

For each trial, either the first or third author was the experimenter or primary observer. The

other than independently watched randomly selected recordings of the flights and scored performance using the checklist observation form. After a participant completed the study, numbers corresponding to each trial were placed in a container and 10% were randomly drawn. Care was insured that at least one session during each condition was selected for each participant. This ensured that (a) 10% of the sessions were rescored for each participant and (b) the trials that were rescored were randomly selected, with the restriction that at least one session during each condition was rescored. Interobserver agreement was calculated separately for the occurrence of correct checklist behavior and for the occurrence of errors. Agreement on the occurrence of correct checklist behaviors was computed by dividing the number of agreements on the occurrence of each checklist behavior completed correctly by the number of agreements on the occurrence of each checklist behavior completed correctly plus the number of disagreements. Interobserver agreement on the occurrence of the percentage of checklist errors was computed by dividing the number of agreements on checklist errors by the number of agreements on checklist errors plus the number of disagreements. Mean agreement on the occurrence of correct checklist use was 90% (range, 61% to 100%). Mean agreement on the occurrence of errors was 82% (range, 0% to 100%). The occasion on which mean agreement on errors was 0% was for a session in which one observer scored one error and the other observer did not score any errors.

Integrity of the Independent Variable

To be sure that the technical flight and checklist feedback was administered correctly, the experimenter read from prepared scripts. In addition, participants were asked to initial the technical flight diagrams and the checklist feedback graphs that were used during postflight briefing sessions and give them back to the experimenter. Participants initialed 100% of all flight diagrams and feedback graphs.

RESULTS

Figure 1 displays the total number of checklist items completed correctly for each participant. All participants increased performance accuracy over baseline when postflight checklist feedback was provided, and improvements were maintained after intervention withdrawal.

Baseline checklist performance varied considerably among participants, with P1 showing the highest level of performance and P2 showing the lowest level. Baseline trends were stable over time with the exception of P3, who showed a sudden increase in accuracy following the first two trials, and P6, who showed a slight downward trend.

Following the introduction of feedback, 5 participants (P1, P5, P6, P7, and P8) showed an abrupt change in the number of checklist items completed correctly, 2 (P2 and P4) showed a level change followed by an increasing trend, and 1 (P3) showed a gradually increasing trend. Overall the mean number of checklist items completed correctly for all participants increased from 22.2 during baseline to 39.2 (of 40) during the last three sessions of the intervention phase.

Each participant maintained high levels of correct item completion after the graphic feedback intervention was withdrawn. The mean number of checklist items completed correctly was 39.6 (of 40) after the withdrawal of treatment for P1, P2, P3, P4, P5, and P6. Because P7 and P8 were not able to attend as many sessions as the others, they were not able to move to the withdrawal condition before the end of the semester.

The percentage of total checklist errors for each flight segment for each participant during each condition is shown in Figure 2. During all flights, 1,973 total errors were observed. The percentage of total checklist errors was high and variable during baseline. The mean percentage of checklist segment errors was highest for the after-takeoff checklist (88%, range, 71% to

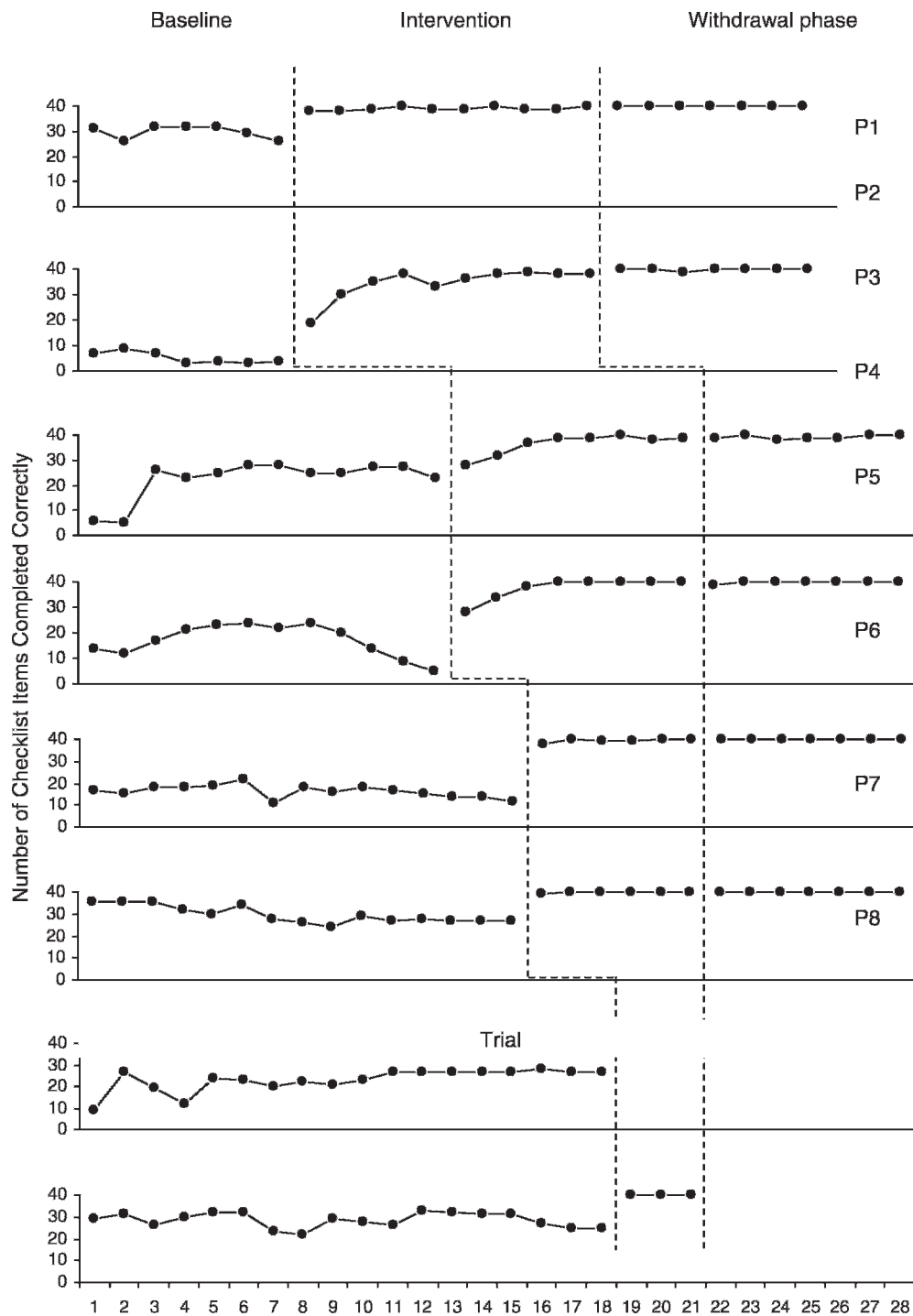


Figure 1. Total number of checklist items completed correctly by each participant during each session of the experiment.

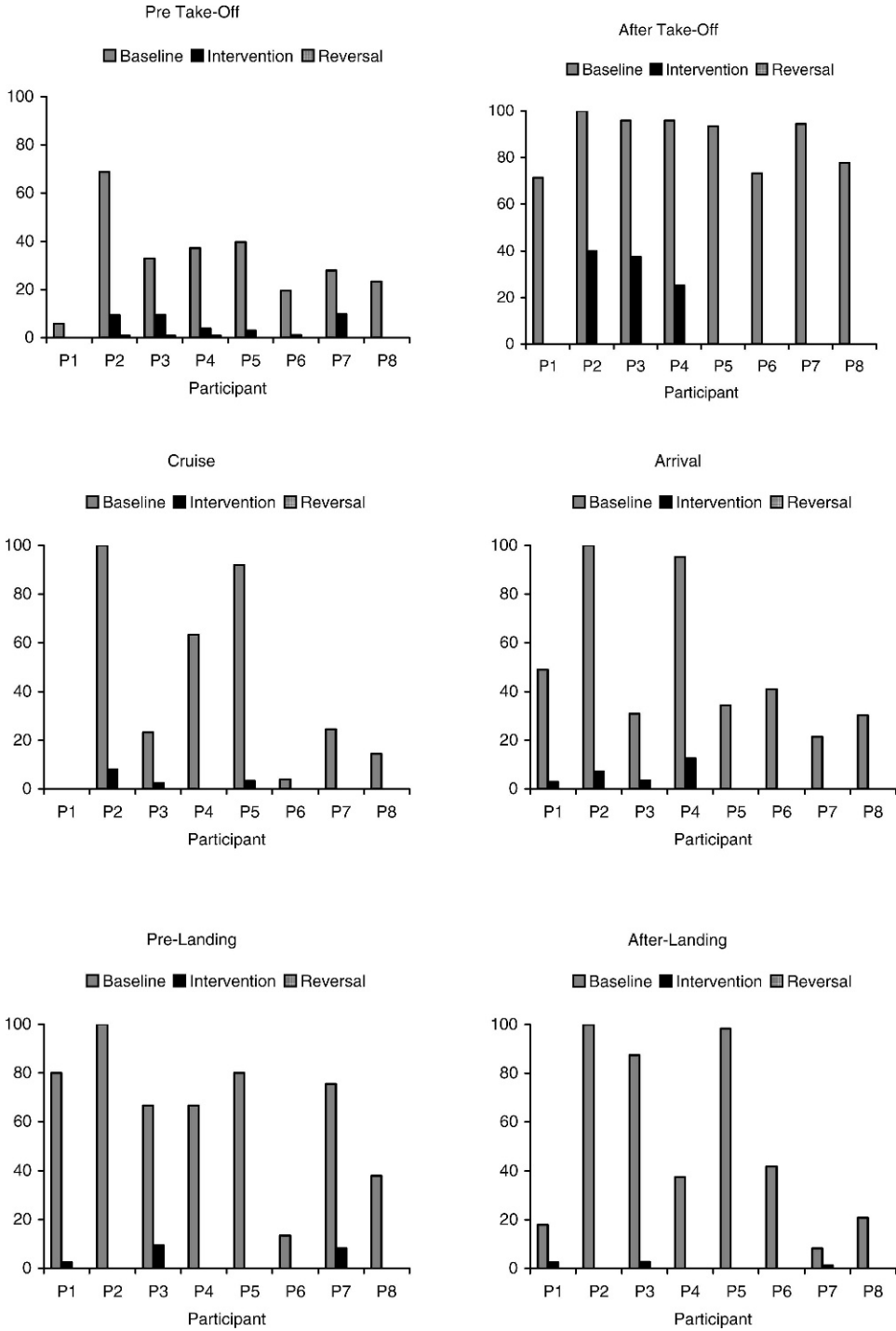


Figure 2. Percentage of errors for each condition per flight segment.

100%) and lowest for the pretakeoff portion of the flights (32%, range, 6% to 69%).

For all participants, errors decreased or were eliminated during intervention. During withdrawal of treatment, 3 participants (P1, P5, and P6) performed perfectly. P2, P3, and P4 had near-perfect performance.

Figure 3 shows the mean percentage of total errors per checklist item across individuals during baseline. Percentages that are 50% or greater are shaded for each participant. Also, the checklist item name is shaded if the percentage of error was 50% or greater for 4 or more participants. Three checklist segments emerge as having the highest percentage of errors. The highest frequency of errors occurred for the two items in the after-takeoff segment: checking flaps and engine instruments, with 99 errors each. These errors occurred on approximately 50% of the total flights across all participants. The prelanding items were the next problematic. Six of the 8 participants (P1, P2, P3, P4, P5, and P7) had high percentages of errors on all five items in this segment. The after-landing segment had the third highest errors, with 4 participants (P2, P3, P5, and P7) having high error rates on all four items.

DISCUSSION

The use of paper checklists in the flight environment remains a vital component to safe operations. As with the airline audits conducted by Helmreich et al. (1999, 2001) and Klinect, Murray, Merritt, and Helmreich (2003), the results of the current study revealed that checklist compliance varied considerably across individuals during baseline, ranging from a mean of 53% to 91%. When the intervention was introduced, accuracy increased rapidly to near-perfect levels for each participant. Furthermore, those changes were maintained after the intervention was removed, continuing for seven simulated flights.

The increases after intervention are similar to increases that have been reported in other studies when graphic feedback and praise have

been provided for desired work behaviors (Austin et al., 2005; Crowell et al., 1988; Wilk & Redmon, 1998). The present research is novel, however, in demonstrating such effects for the use of flight checklists by pilots in an extremely challenging and dynamic situation.

Although this study was conducted in a simulated environment, the results suggest that graphic feedback and praise could be used to increase the extent to which pilots use checklists accurately, potentially preventing crashes and saving lives. A large number of studies show that flight skills acquired in the simulator transfer to the aircraft; however, none of these studies has specifically examined whether checklist use taught in a simulator transfers to the flight deck of the aircraft. Further research should focus on installing data loggers and cameras on the aircraft to measure checklist use before and after training. If checklist use did generalize, it would have major implications for simulated flight training, which is less expensive and less risky than in-flight training. Even if checklist use did not improve without adding enhancements to promote generalization, these results would be directly applicable to the operators of UAVs, for whom actual flight is indistinguishable from simulation.

Baseline performance varied considerably across participants. Moreover, some participants performed very poorly. The variables that contributed to poor baseline performance are not known but may be due to (a) poor initial flight training, (b) no or infrequent feedback on checklist use during training or nontraining flights, (c) no aversive consequences for failing to use the checklist accurately in the simulated setting (e.g., no emergencies that could lead to actual danger, no crashes possible), or (d) a combination of these variables.

Five participants (P1, P5, P6, P7, and P8) showed an abrupt change in performance after the first intervention trial. In addition, all 5 maintained high levels of performance after the withdrawal of treatment. The abrupt increase in accuracy after one intervention session and the

PRE TAKE-OFF	P1	P2	P3	P4	P5	P6	P7	P8
Carburetor heat	0.00	0.43	0.08	0.25	0.00	0.00	0.06	0.00
Fuel quantity	0.00	0.71	0.00	0.33	0.00	0.00	0.11	0.00
Flight instruments	0.00	0.71	0.08	0.17	0.07	0.27	0.94	0.39
Controls	0.00	0.86	0.58	0.33	0.73	0.07	0.00	0.33
Throttle friction lock	0.00	1.00	0.17	0.75	0.80	0.07	0.06	0.06
Flaps	0.00	1.00	0.08	0.25	0.73	0.73	0.00	0.06
Elevator trim	0.00	1.00	1.00	0.75	0.93	0.13	0.94	0.06
Fuel selector	0.00	1.00	0.25	0.25	0.00	0.07	0.11	0.00
Transponder	0.00	0.57	0.17	0.17	0.13	0.00	0.11	0.00
Takeoff briefing	0.00	1.00	0.42	1.00	0.93	0.80	0.50	0.00
Avionics	0.86	1.00	0.33	0.67	0.40	0.47	0.50	0.06
Pitot heat	0.00	1.00	0.83	0.67	0.13	0.60	0.50	1.00
Strobes	0.00	0.86	0.08	0.00	0.00	0.00	0.39	0.00
Mixture	0.00	0.86	0.17	0.00	0.13	0.00	0.11	6.00
Landing Light	0.29	0.00	0.25	0.17	0.20	0.00	0.06	0.17
Compass	0.00	0.57	0.08	0.25	0.53	0.13	0.28	0.72
Time set	0.00	1.00	1.00	0.08	1.00	0.00	0.33	1.00
AFTER TAKE-OFF								
Flaps	0.57	1.00	0.92	0.92	0.93	1.00	0.94	0.78
Engine instruments	0.57	1.00	1.00	1.00	0.93	0.67	0.94	0.78
CRUISE								
Landing light	0.00	1.00	0.42	0.67	1.00	0.00	0.28	0.17
Mixture	0.00	1.00	0.25	0.75	1.00	0.00	0.22	0.11
Engine instruments	0.00	1.00	0.08	0.58	0.93	0.00	0.22	0.11
Fuel quantity	0.00	1.00	0.17	0.58	1.00	0.00	0.22	0.11
Altimeter	0.00	1.00	0.25	0.58	1.00	0.20	0.28	0.22
ARRIVAL								
Landing light	0.43	1.00	0.17	0.92	0.33	0.40	0.17	0.28
Nav radios	0.57	1.00	0.17	1.00	0.47	0.47	0.33	0.28
Engine instruments	0.43	1.00	0.25	0.92	0.33	0.40	0.17	0.28
Compass	0.43	1.00	0.25	0.92	0.33	0.40	0.22	0.33
Altimeter	0.43	1.00	0.33	0.92	0.27	0.60	0.22	0.28
Mixture	0.43	1.00	0.33	1.00	0.60	0.53	0.17	0.33
Approach brief	0.71	1.00	0.67	1.00	0.40	0.07	0.22	0.33
PRE LANDING								
Carburetor heat	0.86	1.00	0.50	0.67	0.80	0.00	0.56	0.28
Fuel selector	0.71	1.00	0.58	0.58	0.73	0.00	0.61	0.28
Mixture	0.71	1.00	0.58	0.58	0.80	0.07	0.67	0.28
Flaps	0.86	1.00	0.92	0.50	0.73	0.40	1.00	0.50
Parking brake	0.86	1.00	0.75	1.00	0.93	0.47	0.94	0.56
AFTER LANDING								
Carburetor heat	0.14	1.00	0.83	0.50	0.93	0.20	0.94	0.11
Flaps	0.14	1.00	0.83	0.58	1.00	0.20	1.00	0.22
Lights and pitot heat	0.14	1.00	0.92	0.17	0.87	0.47	0.94	0.17
Parking brake	0.29	1.00	0.92	0.25	1.00	0.80	0.83	0.22

Figure 3. Percentage of errors per checklist item across individuals during baseline. Items are shaded if a participant missed that item 50% or more of the time, and checklist items are shaded if 4 or more participants missed that item 50% or more of the time.

maintenance of high levels of performance following intervention removal suggest that checklist use was being controlled by rules rather than by direct-acting contingencies, and those participants formed new rules after receiving feedback and praise. The nature of the changes in rules is not known. However, these new rules may have brought checklist behavior under the dual control of the checklist item and the relevant antecedent stimulus in the flight segment and continued to affect behavior once the experimenter-provided feedback and praise were withdrawn (Galizio, 1979; Shimoff, Catania, & Matthews, 1981).

In contrast to these 5 participants, 1 participant (P3) showed a gradual increase in performance over time after successive intervention sessions. This transition may indicate selection by consequences. Two participants (P2 and P4) showed an abrupt change in performance after the first intervention trial followed by a gradually increasing trend, which may indicate that checklist use was being controlled by both rules and direct-acting consequences. On the other hand, rule-governed behavior can appear to be contingency-shaped behavior (Shimoff, Matthews, & Catania, 1986). The fact that all of these participants maintained their performance after the removal of feedback suggests that all of them formulated new rules about the importance of checklist use as a function of the treatment contingency.

Although it is likely that all participants developed new rules, the types of rules they developed may have been different. The rules may have been related to safety. For example, participants may have developed a rule like "My checklist use is not good. If I perform this poorly when I actually fly, I might crash." If this type of rule was developed, accurate checklist use might well generalize to actual flight. Alternatively, the rules may have been related to evaluation by the experimenters, for example, "If I perform poorly, I will look bad to the experimenters who are experienced pilots." If so, accurate checklist performance would be

unlikely to generalize to actual flight. It is also possible that both types of rules multiply controlled checklist use. Future research should investigate the nature of the rules established.

Another topic for further study is to what extent variations in high or extreme workload conditions influence checklist errors. During baseline, feedback, and withdrawal of treatment, the flight patterns and the condition of the aircraft systems did not vary and were considered normal operations. Particular checklist segments (e.g., after takeoff, arrival, and prelanding) had higher errors during baseline. These checklist segments occur during elevated workload conditions in which competing stimuli tend to increase and become more variable. It is possible that the stimuli and behaviors associated with an increased workload interfere with the stimulus control of stimuli associated with the appropriate checklists. Lapses in standardized recognition of prompts or lack of feedback for recognition of prompts during pilot training may contribute to timing errors and missed items. Thus, recognition of prompts should be emphasized in training, particularly for less salient antecedent stimuli.

There are several avenues of future research. The ones most directly related to the current study include (a) replicating the current study and ascertaining whether checklist compliance transfers to actual flight, (b) replicating the current study during actual training flights when flight conditions (e.g., weather and airport traffic) differ, and (c) determining how long gains in checklist accuracy would continue in the absence of postflight feedback and praise.

Although the design and composition of checklists, their position and placement on the flight deck, and standard operating procedures that require their use may encourage accurate checklist completion, they do not ensure it (Degani & Wiener, 1990). In this study, postflight graphic feedback and praise increased checklist compliance to near-perfect levels. This is the first time that this type of behavioral

intervention has been used to alter checklist use. The intervention was a package, and thus it is not possible to determine the effects of the individual components. Nonetheless, the results of the current study are clear: Graphic feedback and praise can increase the accurate use of flight checklists in a simulated flight setting. Further research is needed to determine whether the results generalize to actual flight and whether the results would be similar when workload demands are elevated due to abnormal flight conditions. There are many other questions in aviation safety that may be addressed with behavior-analytic research methods, including paper versus electronic checklist, behavior-based safety applied to crew performance, improving performance in aircraft maintenance, and specific procedures for initial flight instruction.

REFERENCES

- Adamski, A., & Stahl, A. (1997). Principles of design and display for aviation technical messages. *Flight Safety Digest*, 16(1), 1–29.
- Anderson, C. D., Crowell, C. R., Hantula, D. A., & Siroky, L. M. (1988). Task clarification and individual performance posting for improving cleaning in a student-managed university bar. *Journal of Organizational Behavior Management*, 9(2), 73–90.
- Austin, J., Weatherly, N. L., & Gravina, N. E. (2005). Using task clarification, graphic feedback, and verbal feedback to increase closing-task completion in a privately owned restaurant. *Journal of Applied Behavior Analysis*, 38, 117–120.
- Bacon, D. L., Fulton, B. J., & Malott, R. W. (1982). Improving staff performance through the use of task checklists. *Journal of Organizational Behavior Management*, 4(3/4), 17–25.
- Bone, E., & Bolkom, C. (2003). *Unmanned aerial vehicles: Background and issues for Congress*. Washington, DC: Library of Congress, Congressional Research Service.
- Boorman, D. (2001). Today's electronic checklists reduce likelihood of crew errors and help prevent mishaps. *ICAO Journal*, 1, 17–21.
- Burgio, L. D., Whitman, T. L., & Reid, D. H. (1983). A participative management approach for improving direct-care performance in an institutional setting. *Journal of Applied Behavior Analysis*, 16, 37–53.
- Crowell, C. R., Anderson, D. C., Abel, D. M., & Sergio, J. P. (1988). Task clarification, performance feedback, and social praise: Procedures for improving the customer service of bank tellers. *Journal of Applied Behavior Analysis*, 21, 65–71.
- Degani, A. (1992). *On the typography of flight-deck documentation* (NASA Contractor Rep. 177605). Moffett Field, CA: NASA Ames Research Center.
- Degani, A. (2002, April). *Pilot error in the 90s: Still alive and kicking*. Paper presented at the meeting of the Flight Safety Foundation of the National Business Aviation Association, Cincinnati, OH.
- Degani, A., & Wiener, E. L. (1990). *Human factors of flight-deck checklists: The normal checklist* (NASA Contractor Rep. 177549). Moffett Field, CA: NASA Ames Research Center.
- Degani, A., & Wiener, E. L. (1993). Cockpit checklists: Concepts, design, and use. *Human Factors*, 35(2), 28–43.
- Diez, M., Boehm-Davis, D., & Holt, R. (2003). Checklist performance on the commercial flight-deck. *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 323–328). Columbus: The Ohio State University.
- Galizio, M. (1979). Contingency-shaped and rule-governed behavior: Instructional control of human loss avoidance. *Journal of the Experimental Analysis of Behavior*, 31, 53–70.
- Gross, R. (1995). Studies suggest methods for optimizing checklist design and crew performance. *Flight Safety Digest*, 14(5), 1–10.
- Helmreich, R. L., Klinec, J. R., Wilhelm, J. A., & Jones, S. G. (1999). *The line/LOS error checklist, Version 6.0: A checklist for human factors skills assessment, a log for off-normal events, and a worksheet for cockpit crew error management* (Tech. Rep. No. 99-01). Austin: University of Texas, Human Factors Research Project.
- Helmreich, R. L., Wilhelm, J. A., Klinec, J. R., & Merritt, A. C. (2001). Culture, error, and crew resource management. In E. Salas, C. A. Bowers, & E. Edens (Eds.), *Improving teamwork in organizations* (pp. 305–331). Hillsdale, NJ: Erlbaum.
- Klinec, J. R., Murray, P., Merritt, A. C., & Helmreich, R. L. (2003). Line operation safety audits (LOSA): Definition and operating characteristics. *Proceedings of the 12th International Symposium on Aviation Psychology* (pp. 663–668). Columbus: The Ohio State University.
- Lautmann, L., & Gallimore, P. (1987). Control of the crew-caused accident: Results of a 12-operator survey. *Boeing Airliner*, 1–6.
- Patterson, E. S., Render, M. L., & Ebright, P. R. (2002). Repeating human performance themes in five health care adverse events. *Proceedings of the 46th meeting of the Human Factors and Ergonomic Society* (pp. 1418–1422). Santa Monica, CA: Human Factors and Ergonomic Society.
- Shier, L., Rae, C., & Austin, J. (2003). Using task clarification, checklists and performance feedback to increase tasks contributing to the appearance of a grocery store. *Performance Improvement Quarterly*, 16(2), 26–40.
- Shimoff, E., Catania, A. C., & Matthews, B. A. (1981). Uninstructed human responding: Sensitivity of low-rate performance to schedule contingencies. *Journal of the Experimental Analysis of Behavior*, 36, 207–220.

- Shimoff, E., Matthews, B. A., & Catania, A. C. (1986). Human operant performance: Sensitivity and pseudosensitivity to contingencies. *Journal of the Experimental Analysis of Behavior*, 46, 149–157.
- Taylor, H., Lintern, G., Hulin, C., Talleur, D., Emanuel, T., & Phillips, S. (1999). *Transfer of training effectiveness of personal computer-based aviation training devices*. Washington, DC: U.S. Federal Aviation Administration.
- Turner, T. (2001). *Controlling pilot error: Checklists and compliance*. New York: McGraw-Hill.
- U.S. Federal Aviation Administration. (1995). *Human performance considerations in the use and design of aircraft checklists*. Washington, DC: Author.
- U.S. Federal Aviation Administration. (1997). *Advisory circular 61-126: Qualification and approval of personal computer-based aviation training devices*. Washington, DC, Author.
- U.S. Federal Aviation Administration. (2000). *Advisory circular 120-71: Standard operating procedures for flight deck crew members*. Washington, DC: Author.
- U.S. National Transportation Safety Board. (1969). *Aircraft accident report: Pan American World Airways, Inc., Boeing 707-321C, N799PA, Elmendorf Air Force Base, Anchorage, Alaska, December 26, 1968* (Report No. NTSB/AAR-69/08). Washington, DC: Author.
- U.S. National Transportation Safety Board. (1975). *Aircraft accident report: Northwest Airlines, Inc., Boeing 727-25, 264US, Near Thiells, New York, December 1, 1974* (Report No. NTSB/AAR-75/13). Washington, DC: Author.
- U.S. National Transportation Safety Board. (1982). *Aircraft accident report: Air Florida, Inc., Boeing 737-222, N62AF, Collision with 14th Street Bridge, Near Washington National Airport, Washington, DC, January 13, 1982* (Report No. NTSB/AAR-82/08). Washington, DC: Author.
- U.S. National Transportation Safety Board. (1988a). *Aircraft accident report: BAE-3101 N331CY, New Orleans International Airport, Kenner, Louisiana, May 5, 1987* (Report No. NTSB/AAR-88/06). Washington, DC: Author.
- U.S. National Transportation Safety Board. (1988b). *Aircraft accident report: Northwest Airlines, Inc., McDonnell Douglas DC-9-82, N312RC, Detroit Metropolitan Wayne County Airport, Romulus, Michigan, August 16, 1987* (Report No. NTSB/AAR-88/05). Washington, DC: Author.
- U.S. National Transportation Safety Board. (1989). *Aircraft accident report: Boeing 727-232, N473DA, Dallas-Fort Worth International Airport, Texas, August 31, 1988* (Report No. NTSB/AAR-89/04). Washington, DC: Author.
- U.S. National Transportation Safety Board. (1990). *Aircraft accident report: USAir Inc., Boeing 737-400, LaGuardia Airport, Flushing, New York, September 20, 1989* (Report No. NTSB/AAR-90/03). Washington, DC: Author.
- U.S. National Transportation Safety Board. (1997). *Aircraft accident report: Wheels up landing, Continental Airlines Flight 1943, Douglas DC-9 N10556, Houston, Texas, February 19, 1996* (Report No. NTSB/AAR-97/01). Washington, DC: Author.
- Wilk, L. A., & Redmon, W. K. (1998). The effects of feedback and goal setting on the productivity and satisfaction of university admissions staff. *Journal of Organizational Behavior Management*, 18(1), 45–68.

Received January 29, 2007

Final acceptance March 31, 2008

Action Editor, Chris Ninness