September 15, 2016

# An Evaluation of Cognitive Skill Degradation in Information Automation

Katherine Volz, *Iowa State University*
Euijung Yang, *Iowa State University*
Rachel Dudley, *Iowa State University*
Elizabeth Lynch, *Iowa State University*
Maria Dropps, *Iowa State University*, et al.

# AN EVALUATION OF COGNITIVE SKILL DEGRADATION IN INFORMATION AUTOMATION

Katherine Volz, Euijung Yang, Rachel Dudley, Elizabeth Lynch, Maria Dropps, Michael C. Dorneich
Industrial and Manufacturing Systems Engineering, Iowa State University

The purpose of this research is to investigate long term effects of cognitive skill degradation through the use of automation. Even though advanced studies have looked into information automation (IA) in aviation, the amount of empirical data on the effects of these systems on the retention of cognitive skills is less deeply examined. Measurement and analysis of the effects of IA on cognitive performance is an important first step in understanding cognitive skill degradation, which should be considered during the design of these systems. The use of an automation aid is expected to result in a high level of performance degradation over time. Participants were randomly placed into three experimental groups (manual, alternating, or automation) and asked to perform flight planning calculations as an experiment task. Participants performed the task five times, once every two weeks. The manual group used the manual method throughout the experiment, the alternating group switched between the manual and automated method every trial. The automation group used the manual method for the first trial, the automated method for the three consecutive trials and then went back to using the manual method during the last trial. The automation group showed the most performance degradation and highest workload, while the alternating group presented reduced performance degradation and workload, and the manual group showed the least performance degradation and workload. This work provides the foundation for the design of guidelines and recommendations for IA systems in order to prevent cognitive skill degradation.

## INTRODUCTION

Current display and control system technology on the aviation flight decks has automated many of the tasks that pilots used to do on their own, resulting in lower workload (Kaber & Endsley, 2004), fewer errors, and increasingly safer and efficient airline operations (Wiener et al., 1991; Sherman, Helmreich, & Merritt, 1997; Helmreich, & Merritt, 2000). These systems are highly reliable and failures are extremely rare. As a result, many pilot responsibilities have shifted from direct, hands-on control of the aircraft to that of a systems monitor, intervening only when the primary system fails or cannot perform a given task as well as the human operator. Without consistent use of the piloting skills developed during training, pilot skill degradation is a looming and familiar issue (Parasuraman, Sheridan, & Wickens, 2000). While the retention of direct, hands-on piloting skill is essential in responding to emergency situations, equally as critical is the retention of the cognitive skills that allow pilots to maintain situation awareness at all times, quickly assess new situations, and make the best decision from the options available to them. This paper presents a study that investigates the potential loss of cognitive skills through the use of automation. Developing an understanding of the underlying causes of cognitive skill degradation is necessary to address this important issue during the design of information systems.

Information automation (IA) is unique from control automation (CA) and management automation (MA) that are also found on the flight deck. Whereas CA relates to the direct control (dynamics) of the aircraft and management automation deals with mission oversight, IA encompasses all aspects of data collection (e.g., from sensors, databases, or human input), processing (filtering, prediction from models, varying levels of abstraction, etc.), and presentation to the human operator(s) through any appropriate modality (e.g., visual, auditory, tactile) (Billings, 1997; Nakamura, 2013). As such, it has distinct human factors (HF) issues that must be addressed separately from CA and MA issues.

While IA in aviation is not a new concept, the amount of empirical data on the effects of these systems on the retention of cognitive skills is lacking. Anecdotal evidence that this is a potential safety issue is available through reports on the Aviation Safety Reporting System (ASRS) as well as NTSB accident investigation reports. Measurement and analysis of the effects of IA on cognitive performance is an important first step in understanding the root causes of these types of errors and in addressing them through mitigation recommendations that should be considered during the design of these systems.

## RELATED WORK

### Skill Degradation in Aviation

Automation, by the definition adopted by Parasuraman et al. (2000, pp. 287), is “a device or system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator”. If and when these systems fail, the human operator must be ready and able to perform those tasks. While pilots are trained to handle abnormal, or off-nominal, situations, once they have completed training, the opportunity to practice those error handling skills is limited due to the high reliability of automation systems and the day-to-day demands of their airline schedules. Without practice, these skills can degrade over time, so recurrent training is critical for retaining these skills (Helmreich, Merritt, & Wilhelm, 1999).

The concept of skill degradation in aviation has been discussed for several decades, but it has typically revolved

around the physical skills needed to control an aircraft (Wetzel, Konoske, & Montague, 1983). In the age of the glass cockpit and the IA functionality that comes along with these displays, the concern now includes degradation of cognitive skills required to safely complete a flight. These skills include calculating, comprehending, reasoning, prediction, and decision making (Anderson, 1982). In order to reduce pilot workload and potential for error, IA systems are designed to include many of these functions, leaving pilots with the task of monitoring the system and intervening when necessary. Monitoring highly reliable systems has been shown to be difficult for humans (Bainbridge, 1983 and Parasuraman et al., 1993).

### Consequences of Skill Degradation

The 1990's saw a high number of controlled flight into terrain (CFIT) accidents. For example, an inter-agency task force analyzed the NTSB aircraft accident database during the period January 1, 1990 through December 31, 1998 and found that of the 126 fatality accidents of commercial jet airplanes that occurred in Alaska, 89 (71%) involved CFIT (Bailey, 2000). Enhanced Ground Proximity Warning Systems have helped reduce the number of incidents and accidents attributable to CFIT, but in an analysis of fatal accidents of commercial jet airplanes worldwide from 2003 through 2012 (Boeing Commercial Airplanes, 2013), CFIT was among three categories in the number of fatal occurrences, grouped closely with loss of control in-flight and runway excursion during landing (all three of which had more than double the number of incidents than the next closest category). These types of accidents are frequently attributed to a lack of situation awareness, which is a supportive function of IA systems.

Anecdotal evidence of cognitive skill degradation can be found in the ASRS database, which allows pilots to confidentially report situations they see as potential safety issues. In one such report (ACN 1052320), a new pilot was flying with a captain who had not flown the aircraft to which they had been assigned for 30 days. This 30-day gap became an issue when they experienced moderate icing conditions, but the captain "appeared a little lost" looking up at the ice panel and eventually, after some hesitation, redirected engine heat to the wings. After some time, a warning light illuminated indicating a fault in the timer that alternates application of heat between the wings and the tail at 30 second intervals. The captain failed to notice the light on the ice panel, which was immediately above her head, so the new pilot informed her of this fault indication and they took turns manually activating heat to the flight surfaces in rotation. With the gap in flight experience with this airplane, the captain's instrument scanning skills and the ability to recognize and quickly comprehend the meaning of warning lights appeared to have degraded. Had the fault indicator gone unnoticed for longer than it did, the result could have been catastrophic.

### Knowledge Types

There are two types of knowledge that are important in the acquisition of skill: declarative and procedural (Anderson, 1982). Declarative knowledge consists of things that are facts, such as Sacramento is the capital of California and koalas are not bears, rather they are marsupials. This type of knowledge is stored in memory and is considered to be static. Procedural knowledge, on the other hand, is the term given to the knowledge of how to perform tasks or to operate devices. Akin (1986) further categorizes these two knowledge types as specific and general. Tokens and attributes are part of the declarative-specific category, while schemata and inferences fall under declarative-general knowledge. Procedural-specific knowledge includes the cause and effect relationships in a process, while heuristics belong to the procedural-general knowledge category.

In the early stages of learning, operator experiences as well as "book" knowledge form both the declarative-specific and procedural-specific knowledge about the domain in which one is operating. As operators gain experience and increase their skill level, they develop the general mental models that can be called upon to predict system behavior and respond more quickly to a given situation.

## METHOD

### Research Objective and Hypothesis

The objective of the study was to examine the effects of cognitive skill degradation over time through the use of automation. The use of an automation aid is expected to result in larger skill degradation than manual practice over time. Reliance upon an automation aid will lead to higher workload when the user is required to use manual.

### Participants

The study consisted of five visits spread out over nine weeks. A total of 59 undergraduate students from a large Midwestern university served as participants initially (32 male, 27 female), with an average age of 19.69 (range: 18-27). A total of 46 participants completed all five trials (26 male, 20 female), with an average age of 19.7 (range: 18-27). The experiment required participants to return every two weeks for nine weeks. Table 1 indicates the number of participants that came to each trial, the cumulative attrition was approximately 22%.

**Table 1. Participant attendance for each trial.**

| Trial | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Participants at Trial | 59 | 56 | 50 | 48 | 46 |

No participants stated that they had previous experience with flight planning. Participants have taken classes including Algebra (100%), Geometry (98%), Trigonometry (92%), Pre-Calculus (83%), Calculus (76%), and Statistics (61%).

### Task

Participants were asked to conduct flight planning. In order to plan a flight they had to calculate the following elements of a flight segment: heading, ground speed, flight distance, time en route, fuel consumption, or gallons burned per hour.

This experiment utilized procedural and declarative knowledge types. An example of procedural knowledge in relation to the experiment is that the participant must know how to operate the manual apparatus. The participant uses declarative knowledge by understanding the meaning of each unit number provided.

### Independent Variables

The independent variable was the *Experimental Group* (manual, alternating, and automation). Experimental groups were randomly assigned to each participant at the beginning of the study.

Participants were randomly placed into three experimental groups: manual, alternating, or automation. The manual group used an E6-B flight computer for every experiment trial. Flight planning on an E6-B was chosen as the task because it required processing of a procedure, synthesizing information, and assessing the results – all cognitive skills that may degrade over time. The alternating group switched between an E6-B flight computer and an E6-B emulator on a tablet every trial. The automation group used an E6-B flight computer for the first trial, used the automated E6-B for the three consecutive trials and then went back to using the regular E6-B during the last trial. See Table 3 below for a visual description of the schedule.

**Table 2. Experiment schedule.**

| | **Week 0** (Trial 1) | **Week 2** (Trial 2) | **Week 4** (Trial 3) | **Week 6** (Trial 4) | **Week 8** (Trial 5) |
|---|---|---|---|---|---|
| **Manual** | Manual | Manual | Manual | Manual | Manual |
| **Alternating** | Manual | Automated | Manual | Automated | Manual |
| **Automation** | Manual | Automated | Automated | Automated | Manual |

### Dependent Variables

The dependent variables (see Table 3) were workload and performance. Workload for each participant was measured subjectively by a NASA-Task Load Index (NASA-TLX). Performance was measured objectively by rate of error in calculation of flight planning task questions.

**Table 3. Dependent variables.**

| DV | Measurement | Frequency |
|---|---|---|
| **Performance** | Between groups: Percent correct | Every trial |
| | Within Group: Percent correct change between trial 1 and trial 5 | N/A |
| **Workload** | NASA – TLX | Every trial |

Performance was measured through the percent correct on three questions after each flight planning scenario. An example question relating to the calculation of heading and ground speed is, "The weather report indicates that there are winds from 240° at 38 knots (KTS). Your course is 300° and your aircraft has a true airspeed of 165 KTS. Calculate the true heading and ground speed."

### Experiment Design

This was a 1x3 (manual, alternating, and automation) between-subject design. Each condition was tested once per trial (five trials). In training and the five trials, two scenarios were given. Each scenario consisted of three questions. Throughout the experiment the only alterations made were the numbers given in the problem statements. The difficulty in the questions remained the same. When the participant was using the automated method (see Figure 1b), specific paths were given in the instructions for each trial to ensure error would not occur while using the device.

### Procedure

The duration of the experiment was nine weeks, where the participants came in every other week to perform flight planning tasks. After signing a consent form, participants completed a survey to collect basic demographic data. They were briefed about the experiment schedule and shown the instructions to a NASA Task Load Index (NASA-TLX). An initial training session was given to all the participants on how to use an E6-B via video and oral instruction. Next, the user would take a practice test to confirm their understanding which would be graded to see if they needed further assistance. If further assistance was needed, an experimenter would work with the participant individually to identify and correct any errors.

Once participants had completed the practice test with no errors, they were able to begin trial 1. There were a total of five trials, one trial every other week for a total of five visits. For every trial, the participants completed two scenarios consisting of three questions each. It was intentional to not provide immediate feedback after each trial to participants because the experimenters did not want to intervene with the participants' skill set. After the scenarios, they were instructed to fill out a NASA-TLX questionnaire.

### Testing Apparatus

An experimental booklet was assigned to each participant. Six E6-Bs and five 8.1 inch tablets were purchased. A projector and speakers were utilized during training. While participants completed trials, they sat at separated workspaces to insure privacy and individual work.

Figure 1a shows the E6-B Flight Computer developed by Aviation Supplies & Academics, Inc. (2012). Figure 1b shows the HP Stream 8 32GB Windows 8.1 4G 8 inch Tablet.

(a) (b)

**Figure 1. (a) E6-B Flight Computer. (b) HP Stream 8 Tablet.**

### Assumptions and Limitations

For the purposes of developing an empirical study to investigate skill degradation, assumptions were made. First, a given skill has been taught and learned to a sufficient level at the beginning of the study. Second, the primary cause of degradation is insufficiently consistent use of a particular skill after it has been learned.

The experiment participants were all college students performing a simple task over the course of nine weeks. In addition, solving given problems requires an ability to interpret and understand the values by employing the E6-B flight computer. However, the degree of such ability is lower than that of a pilot, because the participants were not trained pilots.

## RESULTS

Satterthwaite approximation t-tests were used for comparing results between groups and trials. Post-hoc analysis was conducted by using Tukey's test for pairwise comparisons. The results are reported as significant for alpha < .05. The error bars represent standard error between participants within a group.

### Performance

Figure 2 illustrates the average performance of the three groups over five trials. The main effect of group was significant ($F(2, 62.9) = 13.99$, $p < .001$). The main effect of trial was significant ($F(4, 171) = 39.17$, $p < .001$). The interaction was significant ($F(8, 171) = 18.81$, $p < .001$).

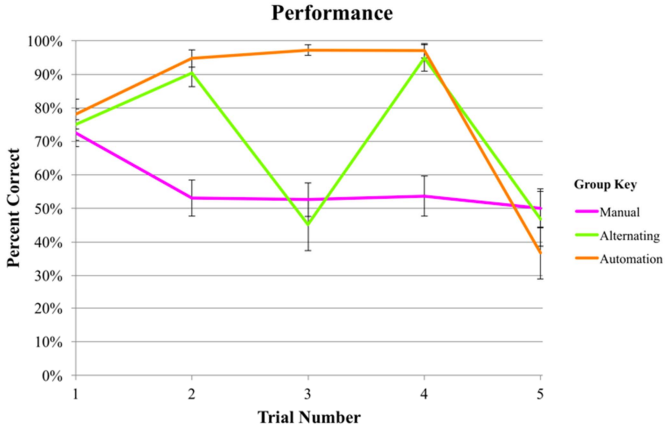


**Figure 2. Average performance of participants on flight task over five trials.**

*Comparing Trial 1 and Trial 5 within group*

For the manual group, there was a significant ($t(217.15) = 3.26$, $p = 0.0115$) decrease in performance from trial 1 ($M = 0.7250$, $SE = 0.0401$) to trial 5 ($M = 0.5000$, $SE = 0.0581$). For the alternating group, there was a significant ($t(215.1) = 3.17$, $p = 0.0153$) decrease in performance from trial 1 ($M = 0.7500$, $SE = 0.0468$) to trial 5 ($M = 0.4688$, $SE = 0.0816$). For the automation group, there was a significant ($t(218.5) = 5.92$, $p < .001$) decrease in performance from trial 1 ($M = 0.9474$, $SE = 0.0259$) to trial 5 ($M = 0.3672$, $SE = 0.0769$).

*Comparing Groups within Trial*

At trial 1, there were no significant differences between groups. For trials 2, 3, and 4, there were no significant differences between groups performing the task with the same method (manual or automated). However, there were significant differences when any group used different methods. For example, at trial 3, alternating ($M = 0.4519$, $SE = 0.0782$) and automation ($M = 0.9722$, $SE = 0.0162$) showed significant differences ($t(190.1) = -6.45$, $p < .001$). Other statistical results showed a similar pattern. At trial 5, there were no significant differences between groups.

### Average Workload

Figure 3 illustrates the average workload of the three groups over five trials. The main effect of group was not significant. The main effect of trial was significant ($F(4, 170) = 40.76$, $p < .001$). The interaction was significant ($F(8, 173) = 35.63$, $p < .001$).

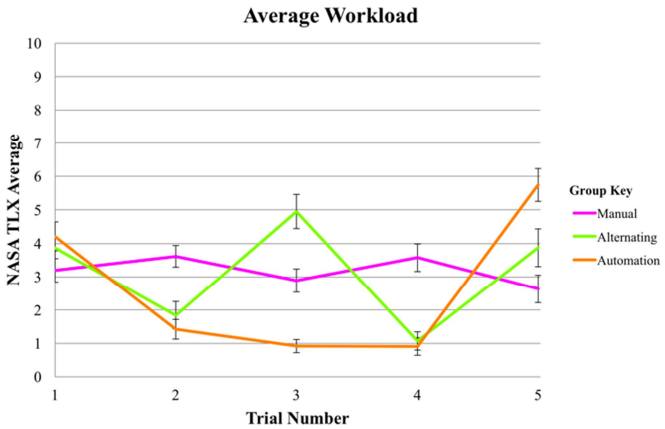


**Figure 3. Average workload of participants after flight task over five trials.**

*Comparing Trial 1 and Trial 5 within group*

For the manual group, there was a no significant difference in average workload from trial 1 to trial 5. For the alternating group, there was a no significant difference in average workload from trial 1 to trial 5. For the automation group, there was a significant ($t(221.8) = -2.97$, $p = 0.0279$) increase in average workload from trial 1 ($M = 4.2083$, $SE = 0.4318$) to trial 5 ($M = 5.7448$, $SE = 0.4841$).

*Comparing Groups within Trial*

At trial 1, there were no significant differences between groups. For trials 2 and 4, there were no significant differences in workload between groups completing the task with the same method (manual or automated). However, there were significant differences when any group used different methods. For example, at trial 4, manual ($M = 3.5787$, $SE = 0.4103$) and automation ($M = 0.9020$, $SE = 0.2640$) showed significant differences ($t(155.4) = -5.00$, $p < .001$). Other statistical results showed a similar pattern. At trial 3, every group was statistically different. Alternating group ($M = 4.9551$, $SE = 0.5046$) presented significant ($t(168.7) = 7.40$, $p < .001$) increase in workload compared to automation group

(M = 0.9190, SE = 0.1973). Alternating group showed significant ($t(167.4) = 4.01$, $p = 0.0003$) increase in workload compared to manual group (M = 2.8860, SE = 0.3520). In addition, automation group (M = 0.9190, SE = 0.1973) showed significant ($t(152.6) = -3.81$, $p = 0.0006$) decrease in workload compared to manual group (M = 2.8860, SE = 0.3520). At trial 5, there were significant differences in workload between every group. Alternating group (M = 3.8750, SE = 0.5642) presented a significant ($t(160.3) = -3.04$, $p = 0.0078$) decrease in workload compared to automation group (M = 5.7448, SE = 0.4841). In addition, automation group showed a significant ($t(154.1) = 5.67$, $p < .001$) increase in workload compared to manual group (M = 2.6389, SE = 0.4186).

## DISCUSSION

The study investigated long term effects of cognitive skill degradation through the use of automation. The results of this study show that the use of an automation aid presented the highest level of performance degradation between the first and last trial. All three groups decreased in performance when comparing trial 1 to 5. Specifically, the manual group displayed the least degradation of performance. The alternating group demonstrated moderate degradation of performance. The automation group showed the highest degradation of performance.

Although all three groups presented degraded results of performance, the manual group had the least amount of degradation whilst the automation group showed the highest degradation. This indicates that reinforcing practice of a task manually can be used to mitigate skill degradation. The automated method of both the automation and alternating groups was not helpful to lessen skill degradation when completing manual tasks.

A potential reason why the manual group showed a decrease in performance from trial 1 to trial 2 is due to lack of feedback during training. It was intentional to not provide immediate feedback after each trial to participants because the experimenters did not want to intervene with the participants' skill set. It is possible that no feedback after each trial led to lower performance because participants did not know whether their approach was correct or not.

Training resulted in a consistent, yet mediocre level of performance of the manual group in trials 2 to 5. While the results demonstrate a skill degradation between the automation and manual groups, future work should focus on highly trained participants (i.e. high skill level).

Reflecting upon workload, as more time elapsed between using the manual method average workload increased. Using the automated method provides participants with a lower workload whilst using the aid, however, relying on automation increases workload when switching back to the manual method.

Further studies need to establish how to mitigate skill degradation. The results of this study can be used to guide design. Measurement and analysis of the effects of IA on cognitive performance is an important first step in understanding the root causes of these types of errors and in addressing them through mitigation recommendations that should be considered during the design of these systems. These discussions should be addressed in future work, and will expand the understanding of long term effects of skill degradation.

## ACKNOWLEDGEMENTS

The authors would like to thank Nicole Kittleson and Jeremy Hadler for supporting the data analysis. This work was funded by the Industrial and Manufacturing Systems Department's Undergraduate Research Assistant program (Leslie Potter and Devna Popejoy-Sheriff, program managers).

## REFERENCES

Akin, Ö. (1986). Psychology of architectural design. London: Pion.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological review, 89(4), 369-406.

Bailey, L. L., Peterson, L. M., Williams, K. W., and Thompson, R. C. (2000) Controlled Flight Into Terrain: A Study of Pilot Perspectives in Alaska. FAA Final Report. DOT/FAA/AM-00/28. Office of Aviation Medicine, Washington, D.C.

Bainbridge, L. (1983). Ironies of automation. Automatica, 19(6), 775-779.

Billings, C. E. (1997). Aviation automation: The search for a human-centered approach.

Bloom, B. S., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain.

Boeing Commercial Airplanes. (2013). Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations 1959-2012. Available: http://www.boeing.com/news/techissues/pdf/statsum.pdf (accessed February, 2014).

Fitts, P. M., & Posner, M. I. (1967). Human performance. Brooks/Cole Pub. Co.

Helmreich, R. L., & Merritt, A. C. (2000). Safety and error management: The role of crew resource management. Aviation resource management, 1(2000), 107-119.

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. The international journal of aviation psychology, 9(1), 19-32.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theoretical Issues in Ergonomics Science, 5(2), 113-153.

Nakamura, D. (2013). Operational Use of Flight Path Management Systems. Washington DC: FAA.

Overbaugh, R., & Schultz, C. (2013). Old Dominion University. Available: http://ww2.odu.edu/educ/roverbau/Bloom/blooms_taxonomy.htm (accessed February, 2014).

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 30(3), 286-297.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. IEEE Transactions on Systems, Man, and Cybernetics, SMC 13, 257-266.

Sherman, P. J., Helmreich, R. L., & Merritt, A. C. (1997). National culture and flight deck automation: Results of a multination survey. The International journal of aviation psychology, 7(4), 311-329.

Wetzel, S. K., Konoske, P. J., & Montague, W. E. (1983). Estimating skill degradation for aviation antisubmarine warfare operators (AWs): Loss of skill and knowledge following training (No. NPRDC-SR-83-31). NAVY PERSONNEL RESEARCH AND DEVELOPMENTCENTER SAN DIEGO CA.

Wiener, E. L., Chidester, T. R., Kanki, B. G., Palmer, E. A., Curry, R. E., & Gregorich, S. E. (1991). The impact of cockpit automation on crew coordination and communication. Volume 1: Overview, LOFT evaluations, error severity, and questionnaire data.