

# A FLIGHT SIMULATOR STUDY TO EVALUATE MANUAL FLYING SKILLS OF AIRLINE PILOTS

Andreas Haslbeck<sup>1</sup>, Paul Kirchner<sup>1</sup>, Ekkehart Schubert<sup>2</sup>, and Klaus Bengler<sup>1</sup>

<sup>1</sup>Institute of Ergonomics, Technische Universität München, <sup>2</sup>Institute of Aeronautics and Astronautics, Section Flight Guidance and Air Transportation, Technische Universität Berlin

This paper reports an experimental study with the objective to assess pilots' raw-data-based flight performance which is affected by long-term practice and structured training. Fifty-seven airline pilots with different levels of aviation experience scheduled on an Airbus fleet, representing contrary levels of practice and training, had to fly a simulated 45 minutes approach and landing scenario while flight performance data were objectively recorded. The level of practice and training was found to have a significant influence on manual flying skills. Pilots with low levels of practice and training showed a large variance in manual flight performance; pilots with high levels of practice and training demonstrated high and homogenous performance.

## INTRODUCTION

For today's aviation experts, manual flying is a critical issue. The aviation industry, primarily aircraft manufacturers and air carriers, is trying to manage the trade-off between safe and economic flight operation. On the one hand, they put emphasis on the training of their pilots' manual skills, but on the other hand, obligatory manual training hours are to be reduced with every new type of aircraft introduced, in order to reduce the effort for necessary type rating training sessions. In preventing accidents, the pilot's manual flying skills are considered as the last line of defense: if all the automation breaks down and manual operation becomes necessary, a pilot is still in charge of conducting a safe landing on his own. Two very prominent disasters, indicating lacking manual flying skills are Air France flight 447 and Asiana Airlines flight 214 (BEA, 2012; NTSB, 2014)

Although much research has been done in the field of aviation, empirical studies on manual flight performance in specific comparable scenarios are still rare. Some of the more prominent works in this area are discussed below. In the present paper, the question is addressed whether airline pilots can maintain sufficient manual flying skills by recurrent training and daily flight practice over the course of a pilot's career. The experimental results are derived from a flight simulator study, which was performed in cooperation with a major European airline.

### Acquisition of manual flight

Manual control of an aircraft is an active task relative to when pilots monitor the aircraft under automation (Flach, 1990; Sarter & Woods, 1994); also known as a closed-loop control problem (Field & Harris, 1998; Wickens, 2003). Manual flying is a psychomotor process requiring more than operating the control stick of an aircraft. Three main stages of information processing have to be considered in manual flying: perception, cognitive processing, and response execution (Childs & Spears, 1986). One model frequently referred to for these sequent stages was founded by Wickens (Wickens & Hollands, 1999). In flight school, pilots learn and intensively

train these active processes before they are introduced to automation, which then switches their task as a pilot from handling an aircraft to managing it (Childs & Spears, 1986; JAA, 2006). From this point on, pilots are faced with automation induced skill degradation (Balfe, Wilson, Sharples, & Clarke, 2012), caused by the automation taking over the responsibility for tasks previously performed by the human operators (Parasuraman & Riley, 1997).

### Automation-induced changes on the flight deck

As flying becomes more automated, pilot's manual flying skills degrade. This inverse relationship is primarily caused by the automation altering the active flying task to a passive monitoring task (Sarter & Woods, 1995). The introduction of early glass cockpits (late 1970s), flight management systems, and fly-by-wire control (late 1980s) in commercial aviation were significant automation milestones. Billings (1991) also described these changes in terms of information, management, and control automation. Automation helps the human operator in difficult situations when incapacity, workload, fatigue or inaccuracy occur – just to name a few. Well known ironies of automation describe negative automation effects like skill degradation (Billings, 1991; Endsley & Kiris, 1996) or reduced operator vigilance (Endsley, 1999).

### Evidence-based experimental studies to assess degrading manual flying skills

In the mid-1980s, empirical studies shed some light on this issue. While a large number of experiments explored exposure to automation and related situation awareness, only a few attempts were made at investigating and measuring the development and degradation of manual flying skills under automation.

An early effort to warn of diminishing flying skills was made by Childs and Spears (1986). They postulated a concern that ineffective perceptual processes lead to deteriorated motor responses. Sarter and Woods (1994) reported a study focusing on pilots' mental models of the flight management system. Their findings revealed that these mental models do not ac-

count for monitoring skills and, to some degree, manual skills, as they require active interventions by the pilot, like the loss of glideslope information. The transition from conventional flight decks towards automated ones was evaluated in a simulator study by Veillette (1995). In an experiment with non-voluntary airline pilots, he measured manual flight performance across varying degrees of automation. The results of this study showed significant differences in manual flight performance between the two groups: the pilots accustomed to automation had significantly larger deviations from the ideal flight paths – evidence of degrading manual skills due to automation. A more recent study analyzing manual flight performance was introduced by Gillen (2008). His comparison showed that pilots' self-assessment delivered higher ratings than pilots were able to perform in the simulator experiment – pilots' confidence in their own skills was subject to a bias. In addition, the pilots tested performed below certification standards, which means these subjects would have failed in a certification situation.

Ebbatson (2009) showed in a large-scale analysis of manual flight performance data a correlation between manual flying skills and practice rather than overall flight experience. Recent flight practice including manual flying occurring a few weeks prior to the experiment had more influence on the measured performance than flight hours accumulated over a pilot's entire career. Finally, Ebbatson suggests replicating his experiment with a group of long-haul pilots, where he assumes even stronger effects between manual flight performance and practice would be found.

These studies all had specific foci, but cannot deliver a comprehensive view on the performance of pilots with a low level of practice and only few opportunities for training, like pilots in long-haul operation. Attempting to fill this gap, the following study was conducted to focus on long-haul pilots. Considering different approaches already mentioned, the current study addresses the following aspects specifically (see also Haslbeck et al. (2012)):

- A randomized sample for manual flying experiments is necessary to avoid self-selection and volunteer biases, thus, participants should not be chosen on a voluntary basis. Otherwise pilots with fairly high skill levels tend to participate in such experiments.
- A highly realistic and valid standardized setting for experimental simulator studies is needed, using a certified full flight simulator with motion effects e.g. from light turbulence weather effects and having real air traffic control instructions, which would force pilots to handle a higher workload by distinguishing between remote messages and their own messages in radio communications.
- The difficulty of a scenario should deliver tasks that can be fulfilled but should also give participants a chance to fail due to their manual flying skills.
- The highly standardized progress sequence of the experimental simulation scenario should ensure that, in general, all participants face the same technical, environmental, and organizational conditions.

## METHOD

### Research Question and Aim of the Study

The main research question is: How do practice and training influence manual flying skills? The concept 'level of practice and training' (according to the German expression *Trainiertheit*) stands for the manual flying skill level of pilots and is affected by the following three aspects:

- passed time since initial flight school to account for long-time skill degradation;
- daily flight practice to consider aspects of on-the-job training of skill;
- the effect of flight simulator training lessons, when selected flying tasks and maneuvers are repeatedly practiced and tested under supervision.

Flight experience, for example in terms of flight hours, and the level of practice and training are inversely proportional: while their flight hours are continuously rising, for most pilots, sections including active handling are rare especially on the long-haul. In this context, experience is rather meant as declarative knowledge how to solve problems and tasks than implicit knowledge or skill how to fly an aircraft. Therefore, it is expected that long-haul captains (CPTs) would have a lower skill level than short-haul first officers (FOs), because it has been longer since they attended flight school including systematic initial flight training, and they have a significantly lower frequency of recent flights than their short-haul colleagues.

### Participants

To investigate the influence of practice and training, younger FOs on short-haul schedules and elder CPTs on long-haul service were chosen at random (stratified random sample), to establish an extreme groups design, representing two typical populations on both evaluated fleets. All participants occupied the same seat as they do in line operation. Two Airbus-type qualified full-flight simulators (JAR-FSTD A) were used for this study because of the very comparable cockpit designs and the resulting ease of transferring between different types (communality). 27 male CPTs participated, representing a low level of daily practice and training but a high level of operational experience. Their simulator was operated in an Airbus A340-600 configuration. For the other group, representing a high level of daily practice and training but a low level of operational experience, 30 FOs (27 male, 3 female) took part in this experiment in an Airbus A320-200 simulator. They should have been in line operation for about five years. Two randomly selected CPTs reported sick and were replaced by two equally qualified but voluntary CPTs. All participating pilots experience four simulator events per annum. The CPTs had more operational tasks (executive decisions) in their last two simulator sessions prior to the experiment, while the FOs were said to have experienced manual flying tasks. This means that the kind of training for both groups of pilots ideally met the experiment's demands. Table 1 shows the demographical data, showing flight experience as overall flight hours and years since flight school, as well as the number of individually performed landings within the past 30 days. Participating pilots

were scheduled for the experiment by their company’s flight operations department, so participation was part of their service schedule and not on a voluntary basis. All pilots were airline pilots (ATPL) and in service of the cooperating airline. All CPTs held the type rating for A330/340 family aircraft, and the FOs held the type rating for A319/320/321 family types.

Table 1. Demographical data of participants (mean values).

	age	overall flight hours	indiv. landings in past 30 days	years since flight school
CPTs (n=27)	50.4	15,019.7	3.4	24.6
FOs (n=30)	30.4	3,373.9	16.6	4.5

**Procedure**

The participating pilots were prepared the same way as for a regular flight, wearing pilot uniforms and bringing their daily used computers for the electronic flight bag system. Subjects were always instructed to be the pilot flying (PF). A confederate pilot monitoring (PM) was instructed to have a passive but cooperative role, and not to cause errors. These confederate pilots (two alternating for each group of participants) were also scheduled by the partner airline on the correspondent fleet. The first subject began the experiment (three subjects per night) approximately two and a half hours after the starting time. The whole procedure resembles longer flights with landings during the early morning hours, representing long-haul flights from the east or mid-range flights operated with short-haul aircraft in the partner airline (Haslbeck et al., 2012).

**Scenario**

After an uneventful flight from the east toward Munich Airport, the PF returned from his last break to perform the approach and landing 25 minutes prior to scheduled touchdown. All flight crews had to perform a missed approach before intercepting the ILS (guide beam provided by the instrument landing system) for a second time. At this time the approach mode could not be armed and the autopilot was disabled by a scripted event. After this point, the pilots had to perform all flying activities manually without the flight director and autopilot assistance. When the localizer was manually intercepted – providing runway centerline guidance – the measurement of manual flight performance started. A hand-flown landing (raw data ILS) with touchdown ended the 45-min. scenario.

**Dependent Measures**

Pilots were instructed to act and fly according to standard operating procedures of their airline (including licensing standards) – the same as in a real flight. Flight performance data were objectively measured by the flight simulator’s data recorder. Here, deviations from the ideal glide slope (vertical guidance), and localizer (lateral guidance) were measured. These metrics represent the resulting system performance according to a control loop including the pilot and the aircraft (Morris & Miller, 1996). Flight path deviations can be consid-

ered in two different ways: measurement of absolute values with averaging afterwards or comparing maximum deviations to licensing standards. Both approaches were pursued and are subsequently shown.

All data for the localizer and glide slope are standardized to aberrations in dots, a unit which can be monitored on the primary flight display in the cockpit and which gives pilots information about their actual attitude with respect to the ideal approach path. The individual glide slope variations for all participants are observed from 3,000 ft. AGL (above ground level) down to 200 ft. AGL, whereas localizer aberration is considered significant from 3,000 ft. AGL to the model height of the aircraft above the threshold of 50 ft. AGL.

According to partner airline manuals, guidelines and laws (JAA, 2006), a maximum variance of one dot deflection on each side of the primary flight display localizer and glide slope scale must be maintained on precision approaches at all times. All measures can be directly compared to pilots’ licensing standards (JAA, 2006; Ebbatson, 2009). To complement the comparison of maximum deviation values to legal standards, the root mean square error (RMSE) was calculated for all pilots to give a combined measure of their accuracy, equally weighting mean error and standard deviation (Hubbard, 1987; Flach, 1990). Here this measure is taken to express the differences between both groups, rather than to distinguish between the directions of both the localizer and glide slope.

**RESULTS**

Pilots’ manual flying performance in terms of maximum localizer and glide slope deviations from the manually flown ILS approach is shown in figure 1 and 2. These two diagrams evaluate pilots’ skill against licensing standards ( $\pm 1$ dot max.).

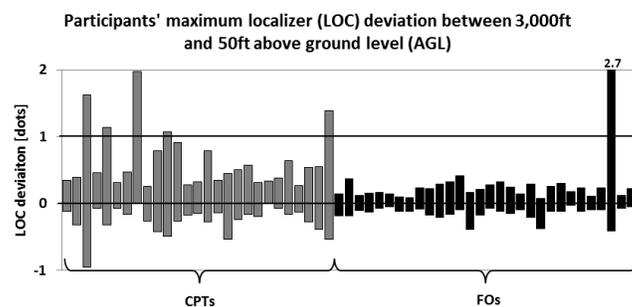


Figure 1: participants’ maximum localizer deviations

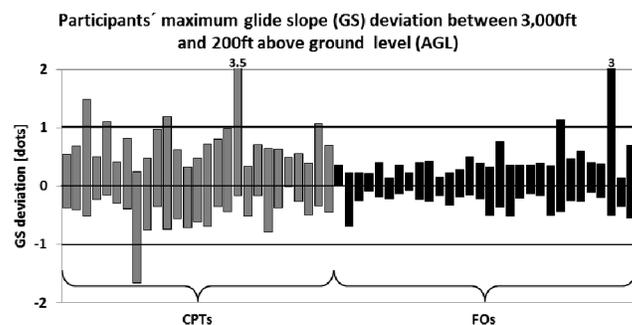


Figure 2: participant’s maximum glide slope deviations

Each bar represents the maximum deviation for each pilot from the target value in figure 1 and 2. Positive localizer deviations imply a horizontal drift to the right of the runway centerline, with negative deviations correspondingly to the left. Vertical drift information is provided via the glide slope indicator. Positive deviances equal an aircraft position higher than the ideal glide path, while negative deviances in contrast represent a lower than ideal position of the airplane. Given the results depicted in Figure 1, six out of 57 subjects violated restrictions on the allowed localizer variance. For glide slope deviation, eight out of 57 participants could not perform within the acceptable limits. A total of nine different pilots (15.8 %) did not meet the mandatory skill test requirements in this scenario. Relative to test-person groups, seven (25.9 %) out of 27 CPTs did not fulfil at least one of the binding ILS deviation parameters, while two (6.7 %) out of 30 FOs did not. For localizer and glide slope deviations, the RMSE is shown in Figure 3.

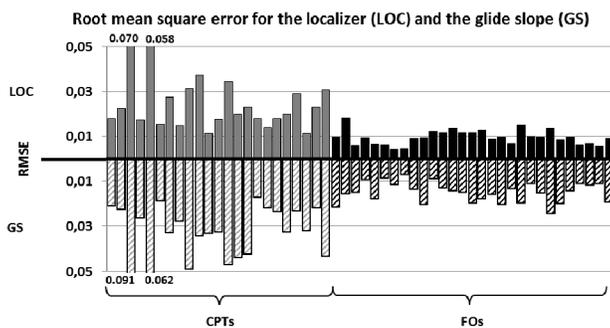


Figure 3: RSME for localizer and glide slope deviations

Table 2: statistical analysis of ILS deviations data

		mean	Independent t-Test t (df); p; r	Mann-Whitney test U; z; p; r	
max. localizer deviations	left	CPTs	.254	t(36.12)=2.55 2; p< .008; r= .39	U=249.5; z=2.489; p= .006; r= .33
		FOs	.143		
	right	CPTs	.646	t(55)=2.873; p= .003; r= .36	U=70; z=5.356; p< .001; r= .71
		FOs	.298		
RMSE localizer		CPTs	.025	t(23.86)=5.19 1; p< .001; r= .73	U=28; z=5.629; p< .001; r= .78
		FOs	.010		
max. glide slope deviation	high	CPTs	.796	t(55)=2.077; p= .021; r= .27	U=154; z=4.013; p< .001; r= .53
		FOs	.481		
	low	CPTs	.471	t(55)=3.08; p< .002; r= .38	U=211; z=3.102; p= .001; r= .41
		FOs	.271		
RMSE glide slope		CPTs	.035	t(24.48)=5.49 3; p< .001; r= .74	U=25; z=5.684; p< .001; r= .79
		FOs	.015		

The two groups were compared using parametric as well as non-parametric tests, as assumptions of normality could not be met for all deviation data. Both tests using an alpha-level of .05 indicate highly significant differences between the two

groups' manual flying performances. In addition, the effect sizes expressed by a point-biserial r show moderate to very large effects.

## DISCUSSION AND CONCLUSIONS

Evidence was found that participating CPTs with a lower level of practice and training (table 1) cause larger deviations from the ideal approach parameters than their more practiced FO colleagues; even if one limitation to this study is the fact that A320 and A340 differ in size, weight, and handling characteristics. From a technical or aeronautical perspective, these are completely different types of aircraft, while seen from a human factors view, both are successive milestones on a pilot's career and the human-machine-interfaces thinly differ because of the communality design principle of Airbus planes. When thinking about the differences in both types, the A340 is over four times heavier than its smaller counterpart and so it has larger flight path inertia. However both types are controlled by comparable roll and pitch rates, realized by larger control surfaces. In addition, the A340's higher flight path inertia might be an advantage in case of external perturbations like gusty wind. Based upon licensing standards on the one hand and flight operation realities on the other hand, all pilots have to perform within the same limits. Licensing standards neither differ between CPTs and FOs nor between long-haul or short-haul. Moreover both groups of pilots as well as both types of aircraft can use the same airports and runways. Thus requirements for manual aircraft handling are the same for all pilots. Another argument for choosing this comparison between A320 FOs and A340 CPTs is the assumed maximum range between the pilots' different skill levels (table 1) to utilize an extreme groups design.

That degrading manual flying skills have been observed in this rather small sample of pilots suggests that this is likely more prevalent than one could have suspected. As participants were active professional pilots, the results should be valid for other airline's personnel. In several cases, the deviations from ideal performance are large enough that pilots would have even failed a check situation – a dramatic finding that could reflect inadequate maintained skills. As training lessons normally cover equal contents for short-haul and long-haul pilots in longer sequences, one can assume that differences in manual flight performance are instead a consequence of everyday flight practice. This hypothesis is also supported by Ebbatson's (2009) study: accordingly, recent flight practice resulting from frequent flight operations, seem to be the most important factor in maintaining manual flying skills. Duncan, Williams and Brown (1991) have found some comparable insights in a real driving car experiment, "that adequate driving skills cannot be assumed, even for the 'average' experienced motorist, simply because they once were mastered [...]". In comparison, simulator training can instead teach the right techniques for handling the aircraft (Buckley & Caple, 2009).

A further limitation to this study is that the level of practice and training is confounded with pilot's age and experience. Tsang (2003) describes and cites findings in her comprehensive review that "older, experienced individuals do not neces-

sarily perform more poorly than their younger counterparts in tasks specific to their domain of expertise." Taylor, Kennedy, Noda, and Yesavage (2007) have reported a study investigating performance changes of pilots with different age and also under regard of different levels of expertise. Their results indicate no strong decline of landing skills by experienced (ATPL) pilots over time. In spite of these findings, in a field experiment with airline pilots, their level of practice and training will always be partially confounded with age and experience. For the measurement of manual flying skills, operational flight experience plays only a minor role and in an airline's daily operation, experience and the level of practice and training normally develop contrarily: CPTs have accumulated a vast amount of flight experience but experience only very few opportunities to practice flight skills – neither in simulator sessions nor in real operation. In spite of these limitations, the results of this study deliver a highly valuable picture of professional pilots' ability to manually control an aircraft.

Long-haul operation with its high degree of automation and pilots' long exposition to automated systems, was shown to have an eroding effect on manual flying skills; pilots with reduced flight duties and part-time schedules, like management pilots or ones who are on parental leave, should be kept in mind. Some examples to be supposed to airlines to implement strategies against deteriorating skills: additional simulator training sessions as well as type rating trainings concentrating on manual aircraft handling; combining short-haul and long-haul operation for long-haul pilots (mixed-fleet flying), especially for CPTs suffering from a lack of practice opportunities. For human-machine-interface designers, the approach of adaptive automation (Parasuraman, 2000) could also lead to a more flexible and dynamic task sharing between human and automation in the near future.

Future studies should further operationalize and analyze the influence of simulator training sessions. In addition, further groups of pilots with medium levels of practice and training, like short-haul CPTs and long-hauls FOs could complement insights in pilot's manual flying skills.

Balfe, N., Wilson, J. R., Sharples, S., & Clarke, T. (2012). Development of design principles for automated systems in transport control. *Ergonomics*, 55(1), 37–54.

BEA. (2012). Final Report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris. Le Bourget.

Billings, C. E. (1991). Human-Centered Aircraft Automation: A Concept and Guidelines: NASA-TM-103885. Moffet Field, California.

Buckley, R., & Caple, J. (2009). *The theory & practice of training*, 6th edition (6th ed.). London: Kogan Page.

Childs, J. M., & Spears, W. D. (1986). Flight-skill decay and recurrent training. *Perceptual and Motor Skills*, 62(1), 235–242.

Duncan, J., Williams, P., & Brown, I. (1991). Components of driving skill: experience does not mean expertise. *Ergonomics*, 34(7), 919–937.

Ebbatson, M. (2009). *The Loss of Manual Flying Skills in Pilots of Highly Automated Airliners* (PhD Thesis). Cranfield University, Cranfield.

Endsley, M. R. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462–492.

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors*, 37(2), 381–394.

Field, E., & Harris, D. (1998). A comparative survey of the utility of cross-cockpit linkages and autoflight systems' backfeed to the control inceptors of commercial aircraft. *Ergonomics*, 41(10), 1462–1477.

Flach, J. M. (1990). Control With an Eye for Perception: Precursors to an Active Psychophysics. *Ecological Psychology*, 2(2), 83–111.

Gillen, M. (2008). *Degradation of Piloting Skills* (Master's Thesis). University of North Dakota, Grand Forks.

Haslbeck, A., Schubert, E., Onnasch, L., Hüttig, G., Bubbe, H., & Bengler, K. (2012). Manual flying skills under the influence of performance shaping factors. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 41(Supplement 1/2012), 178–183.

Hubbard, D. C. (1987). Inadequacy of root mean square error as a performance measure. In *International Symposium on Aviation Psychology*, 4th (pp. 698–704).

Joint Aviation Requirements - Flight Crew Licensing (Aeroplane), Joint Aviation Authorities 2006.

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*, 42(3), 343–360.

NTSB. (2014). Crash of Asiana Flight 214 Accident Report Summary: Descent Below Visual Glidepath and Impact With Seawall. Public Meeting of June 24, 2014. Washington, DC.

Parasuraman, R. (2000). Designing automation for human use: empirical studies and quantitative models. *Ergonomics*, 43(7), 931–951.

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors*, 39(2), 230–253.

Sarter, N. B., & Woods, D. D. (1994). Pilot Interaction With Cockpit Automation II: An Experimental Study of Pilots' Model and Awareness of the Flight Management System. *The International Journal of Aviation Psychology*, 4(1), 1–28.

Sarter, N. B., & Woods, D. D. (1995). How in the World Did We Get into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors*, 37(1), 5–19.

Taylor, J. L., Kennedy, Q., Noda, A., & Yesavage, J. A. (2007). Pilot age and expertise predict flight simulator performance: A 3-year longitudinal study. *Neurology*, 68(9), 648–654.

Tsang, P. S. (2003). Assessing Cognitive Aging in Piloting. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 507–546). Mahwah, NJ: Lawrence Erlbaum.

Veillette, P. R. (1995). Differences in aircrew manual skills in automated and conventional flight decks. *Transportation Research Record*, (1480), 43–50.

Wickens, C. D. (2003). Pilot Actions and Tasks: Selections, Execution, and Control. In P. S. Tsang & M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 239–263). Mahwah, NJ: Lawrence Erlbaum.

Wickens, C. D., & Hollands, J. G. (1999). *Engineering psychology and human performance* (3rd). Upper Saddle River, NJ: Prentice Hall.