Published in final edited form as:
Psychol Aging. 2015 September ; 30(3): 699–711. doi:10.1037/pag0000043.

The STEP model: Characterizing simultaneous time effects on practice for flight simulator performance among middle-aged and older pilots

Quinn Kennedy[1,2], **Joy Taylor**[1,3], **Art Noda**[1], **Jerome Yesavage**[1,3], and **Laura C. Lazzeroni**[1]

[1]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, 94305-5550

[2]Operations Research Department, Naval Postgraduate School, Monterey, CA, 93943

[3]Department of Veterans Affairs Health Care System and Sierra-Pacific MIRECC, Palo Alto, CA, 94304

Abstract

Understanding the possible effects of the number of practice sessions (practice) and time between practice sessions (interval) among middle-aged and older adults in real world tasks has important implications for skill maintenance. Prior training and cognitive ability may impact practice and interval effects on real world tasks. In this study, we took advantage of existing practice data from five simulated flights among 263 middle-aged and older pilots with varying levels of flight expertise (defined by FAA proficiency ratings). We developed a new STEP (Simultaneous Time Effects on Practice) model to: (1) model the simultaneous effects of practice and interval on performance of the five flights, and (2) examine the effects of selected covariates (age, flight expertise, and three composite measures of cognitive ability). The STEP model demonstrated consistent positive practice effects, negative interval effects, and predicted covariate effects. Age negatively moderated the beneficial effects of practice. Additionally, cognitive processing speed and intra-individual variability (IIV) in processing speed moderated the benefits of practice and/or the negative influence of interval for particular flight performance measures. Expertise did not interact with either practice or interval. Results indicate that practice and interval effects occur in simulated flight tasks. However, processing speed and IIV may influence these effects, even among high functioning adults. Results have implications for the design and assessment of training interventions targeted at middle-aged and older adults for complex real world tasks.

Keywords

Practice effects; age; expertise; real world performance

In the cognitive aging field, a considerable amount of research has been conducted on the ways in which practice effects, defined as improved performance due to factors such as test familiarization and sophistication, memory for specific test items, or learned testing

---

Correspondence concerning this article should be addressed to Quinn Kennedy, Stanford/VA Aviation Lab, VA Palo Alto Health Care System (151Y), 3801 Miranda Avenue, Palo Alto, CA 94304-1207. quinnk@stanford.edu.

strategies, influence cognitive test performance (Calamia, Markon, & Tranel, 2012; Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007). Participant age and the interval between tests are two factors that consistently diminish the strength of practice effects on cognitive performance (Calamia, et al., 2012; Ferrer, Salthouse, Stewart, & Schwartz, 2004; Rabbitt, Diggle, Smith, Holland, & Mc Innes, 2001). In the context of understanding longitudinal change in cognitive performance among adults, practice effects are considered to be a nuisance factor that should be teased out (Rabbitt, et al., 2001; T.A. Salthouse & Tucker-Drob, 2008). However, for real world tasks, such as driving a rental car in an unfamiliar area, practice effects -- familiarization with the setting, memory for specific scenarios, and learned strategies -- can be beneficial (Jamieson & Rogers, 2000; Shinar, Tractinsky, & Compton, 2005). The purpose of this study was to extend practice effect findings from the cognitive test performance literature to real world tasks in which practice effects could have positive, tangible performance consequences and important implications for training interventions targeting middle-aged and older adults.

We first briefly describe pertinent results related to practice effects and aging on repeated assessments of cognitive tests. The most comprehensive recent work in this area was an extensive meta-analyses of practice effects on performance on 31 common cognitive tests, including tests of executive function, processing speed, and working memory (Calamia, et al., 2012). Participants were tested twice on each test. Participants ranged in age from 18 to 85; most tests had participants ranging in age from early 20’s to upper 70’s. In addition to modeling practice effects as a function of the ability being assessed, Calamia et al (2012) included possible moderators of practice effects, such as age and test-retest interval. On average, practice effects occurred in a magnitude of .24 standard deviation increase in performance from time 1 to time 2 (based on test-retest interval of one year), with the greatest practice effect on memory tests involving multiple learning trials (California Verbal Learning Test, CVLT) and Rey Auditory Verbal Learning Test (AVLT) and least practice effect on some tests of visuospatial ability. Moderate practice benefits occurred on tests of processing speed and executive function.

Age and interval between tests generally diminished the strength of practice effects across a range of cognitive tests (Calamia, et al., 2012). Age had a negative impact on 27 of the 31 cognitive test scores examined. Across all tests examined, a healthy 70 year old would have about half the expected practice effect benefit of a 40 year old. Longer intervals weakened practice effects on all cognitive tests, such that an interval of about five years would eliminate any expected practice effect advantage. Test-retest interval particularly affected practice effects on memory tests such as the CVLT and processing speed tests, whereas tests of executive function were less affected. Finally, previous research indicates that the magnitude of interval effects does not appear to depend on age (Salthouse, 2011).

These results clearly demonstrate that the benefits of practice vary depending on the type of cognitive function tested, with the strongest benefits on verbal learning/ memory, modest benefits for executive function and processing speed, and weakest benefits for visuospatial ability. Practice effects have important implications for clinical diagnoses, such as mild cognitive impairment, longitudinal development studies (predicting terminal decline), and intervention trials (Calamia, et al., 2012; Cooper, Lacritz, Weiner, Rosenberg, & Cullum, 2004; Dodge, Wang, Chang, & Ganguli, 2011).

Practice effects also play a role in real world tasks, which can range from everyday activities such as driving and work-related skills to expert activities such as professional dancing, competitive athletics, or gaming. Here, it helps to distinguish between "deliberate practice" to acquire or maintain superior performance in a particular skill domain (Ericsson, 2006) and other types of practice relevant to performance stability or compentency. These other types of practice include task repetition to increase task familiarity (Calamia, et al., 2012), refine strategies (Shinar, et al., 2005), study sessions to improve factual recall (Pavlik & Anderson, 2008), training to demonstrate competency in a procedure or a variation of procedure, and practice to maintain a satisfactory level of fitness or competence. Also, the time period involved for deliberate practice typically entails a decade or more to reach a level of expert performance, rather than weeks or months as in the other types of practice described above (Ericsson, 2006). Hereon, we primarily focus on the type of practice that is used to for performance stability or competency stable or competent performance. Below, we review the small literature in this domain.

Little is known about the effect of practice in everyday activities such as driving and work-related skills. We found only a few studies that examined practice effects in real world tasks and they typically involved young adults. One study examined practice effects among aviation students; however, practice effects were assessed in paper and pencil tests rather than actual flight performance, and the range of age and flight experience among the students was very limited (Momen, 2009). The other studies investigated practice effects on admissions tests used to select candidates for entry level job positions or for the premedical track at the undergraduate level (Hausknecht, Trevor, & Farr, 2002; Puddey, Mercer, Andrich, & Styles, 2014). Hausknecht et al (2002) examined whether practice effects were evident among candidates for an entry level position in a law enforcement agency. Candidates who did not pass a cognitive abilities test or an oral communications test were allowed to retake the test, administered once a year. The scores among candidates who retook the test on average, improved with each additional try up until their fourth attempt. Furthermore, the number of attempts until successful admittance into the training program was positively associated with performance in the training program (Hausknecht, et al., 2002). Similarly, consistent practice effects were found among young adults who took the annual Undergraduate Medicine and Health Sciences Admission Test (UMAT) at least twice (Puddey, et al., 2014). The vast majority scored significantly better the second time than initially, and better on the third time than second time. Practice effects beyond the third test were not found. These studies suggest that practice effects may occur in real world, high stakes situations.

There is some evidence to suggest that the benefits of practice may be affected by interval length. One meta-analysis found that practice benefits on highly complex tasks (e.g. milk pasteurization procedure) were less affected by the interval length than less complex tasks (simple motor tasks) (Donovan & Radosevich, 1999). Again, participants in the studies of complex work-related tasks tended to be young adults whose performance had no real world consequences. (Mumford, Costanza, Baughman, Threlfall, & Fleishman, 1994).

Thus, there is evidence that practice and interval do affect performance on real world tasks. Yet, the above studies examining real world tasks do not illuminate whether age interacts with practice and interval effects. Participants were relatively young, with a mean age ranging from 18 years (Puddey, et al., 2014) to 32 years(Hausknecht, et al., 2002) . Therefore it is unclear as to whether these results would have been found with an older sample.

Additionally, the results do not speak to effects of intervals that vary within or across individuals by days, weeks or months. Typically, participants had a constant time interval between each practice or retest session (Hausknecht, et al., 2002; Momen, 2009; Puddey, et al., 2014). When the interval varied across participants, the total practice period was limited to less than a week (Mumford, et al., 1994). Yet in many real world settings, the interval between each practice or retest session can vary from days to months. For example, work schedules may take priority over occupational assessments.

Finally, it is not known whether adults with extensive specialized education, such as pilots with expert-level FAA proficiency ratings, are less affected by the negative influences of interval when these adults repeatedly take assessments relevant to work-related skills. Research on deliberate practice typically has not examined the effect of time interval between practice sessions among experts or adults with professional experience. Many training models exist, however, they tend to focus on optimizing and predicting regimented training schedules for the acquisition of a particular skill (Pavlik & Anderson, 2008; Schmidt & Bjork, 1992). A statistical model that estimates simultaneous practice and interval effects on assessed performance of complex domain relevant tasks among adults with specialized training is needed, particularly when interval can vary widely within or between individuals.

For over a decade, the Stanford/VA Aviation Lab has conducted a longitudinal study assessing yearly flight simulator performance and cognitive function among middle-aged and older general aviators with a range of FAA proficiency ratings. Flight simulators are used by commercial, military, and general pilots to both maintain current flight control skills and to train for situations that would be too dangerous or costly to do in an actual aircraft (Federal Aviation Administration, 2014; Bell & Waag, 1998). The flight scenario presented in the simulator assessment was designed to capture the essence of “Line-Oriented Flight Training” in which a number of representative flight tasks are performed during a single flight (U.S. Department of Transportation, 1990). The tasks include both normal flight maneuvers (e.g. takeoff and approach) and emergency procedures (e.g. traffic avoidance). Prior to conducting the baseline flight simulator assessment, the pilots completed six two-hour practice sessions in the flight simulator to allow pilots to acclimate to our flight simulator and scenario so that the baseline would be a truer assessment of their flight simulator performance. In this study, practice involved task repetition to increase familiarity with the scenario and provide pilots time to refine their strategies for interleaving multiple flight tasks. For each flight, pilots completed a similar scenario involving the same four flight tasks: ATC communication, traffic avoidance, approach to landing, and response to emergencies. The exact ATC items and the timing of traffic and emergencies differed in each flight. Participants typically completed their practice flights during a one to three week period, after which they had a three week break before returning for the baseline visit. Thus, the practice component of this longitudinal study provides an ideal set of data for studying practice and interval effects among middle aged and older adults in a task that is relevant to real world experiences.

Besides examining the effects of practice and interval on simulated flight performance, we also were interested in the extent to which practice and interval effects depend on individual differences in cognitive abilities. For theoretical and empirical reasons, we focused on processing speed, executive function, and intraindividual variability (IIV) in processing speed as potential moderator variables. The capacity to improve with practice may be related to fluid intelligence, which entails the ability to adapt to new situations and the use of reasoning to make predictions and solve problems. Fluid intelligence declines with age (Horn, 1982) and this decline has been linked to age-related decline in working memory (Horn, 1982), which in turn has been linked to declines in processing speed (T. A. Salthouse, 1996) and executive functions (Hasher & Zacks, 1988). IIV in processing speed also has been found to predict older adults’ fluid cognitive performance, even after controlling for processing speed (Bielak, Hultsch, Strauss, MacDonald, & Hunter, 2010b). Examination of processing speed, executive function, and IIV as moderators may help elucidate how age moderates effects of practice.

Previous studies also have determined that processing speed, IIV in processing speed, and executive function play a role in flight simulator performance (Causse, Dehais, & Pastor, 2011; Kennedy et al., 2013; Taylor, O’Hara, Mumenthaler, Rosen, & Yesavage, 2005; Yesavage et al., 2011) more so than other cognitive measures such as episodic and working memory (Taylor et al., 2011; Yesavage, et al., 2011). Possession of advanced FAA proficiency ratings also plays a role, leading to higher and more stable simulation performance assessed yearly (Taylor, Kennedy, Noda, & Yesavage, 2007). Therefore, a secondary goal of this study was to determine if the three cognitive factors, along with age and flight expertise (defined by FAA proficiency ratings) moderated practice or interval effects on flight simulator performance.

In sum, the purpose of this study was threefold: (1) develop a statistical model for examining simultaneous effects of practice and interval between practice sessions on performance; (2) characterize their effects on assessments of flight simulator performance, measured over several weeks, among general aviators ranging in age and flight expertise (i.e., specialized training/education); and (3) based on theories regarding the roles of processing speed, executive function, and IIV on age related decline of fluid cognition (Bielak, Hultsch, Strauss, Macdonald, & Hunter, 2010a; Hasher & Zacks, 1988; T. A. Salthouse, 1996), determine if age, flight expertise, and cognitive abilities moderate practice effects on flight simulator performance.

Method

Participants

We report results from 263 pilots who were part of the ongoing longitudinal Stanford/VA Aviation Study, approved by the Stanford University Institutional Review Board.

Enrollment criteria were age between 40 and 69 years, current FAA medical certificate (Class III or higher) which entails an assessment of pilots' vision, hearing, and physical and mental health, and current flying activity between 300 and 15,000 hours of total flight time. All participants gave written informed consent to participate in the study, with the right to withdraw at any time. At entry, each participant was classified into one of three levels of aviation expertise depending on which FAA pilot proficiency ratings had been attained by study entry: (1) least expertise: VFR (rated for flying under visual flight rules only); (2) moderate expertise: IFR (also rated for instrument flight); and (3) most expertise: CFII and/or ATP (certified flight instructor of IFR students or rated for flying air-transport planes). As reported in our previous work (Taylor, Kennedy, Noda, & Yesavage, 2007), all of the VFR pilots were recreational pilots, although a small minority were employed in aviation-related jobs such as aircraft sales or mechanics. Within the moderate expertise group, the majority of these IFR pilots were recreational pilots, whereas approximately one-tenth were currently certified flight instructors or had been aviators during military service. Within the most expert group, approximately one-half were currently air-transport pilots, CFIIs, or their job duties included piloting: the other half of the expert pilots had other careers. Table 1 provides demographic and flight experience characteristics of the sample.

Equipment

Pilots "flew" in a FAA approved Frasca 141 flight simulator (Urbana, IL). Motion, vibration, and sound elements were not incorporated into this simulator protocol. The simulator was linked to a computer specialized for graphics (Dell Precision Workstation and custom C++ OpenGL Linux software) that generated a "through-the-window" visual environment and continuously collected data concerning the aircraft's position and communication frequencies. The simulator is located in a quiet, darkened room kept at a comfortable temperature with the cockpit independently lit from the projector display. The display is projected on a screen 15′ in front of the pilot. The simulation occurred during normal working hours from 0900 to 1600 at the pilot's preference. Previous work in our lab indicates that the flight simulator has validity as it distinguishes performance between novice and expert aviators, and between younger and older aviators (Taylor, et al., 2007; Taylor, et al., 2005).

Measures

Flight simulator performance—Each flight lasted 75 minutes and consisted of a scenario with 19 flight segments (legs) around the airport, including leg 1: take-off, legs 2–17: enroute flying, leg 18: approach, leg 19: landing. During enroute flying, pilots were given a new ATC command every three minutes with new course (heading), altitude, radio frequency, and in 50% of the legs, a new transponder (identification) code. In order to increase the pilots' workload on legs 2–17, we confronted them with three different emergency situations that occurred randomly: carburetor icing, drop of engine oil pressure, or the sudden approach of air traffic (total of 19 occurrences in 48 minutes). In summary: each flight was designed to include representative normal flight maneuvers (e.g. takeoff and approach) and emergency situations. Due to the inclusion of emergencies and frequent ATC communications, flights were designed to be relatively cognitively demanding, compared to an uneventful cross-country flight.

The scoring system of the flight simulator-computer system produces 23 variables that measure deviations from ideal positions or assigned values (e.g., altitude in feet, heading in degrees, airspeed in knots), or reaction time in seconds (Yesavage, Taylor, Mumenthaler, Noda, & O'Hara, 1999). Because these individual variables have different units of measurement, the raw scores for each variable were converted to *z*-scores, using the baseline visit mean and SD of 141 participants enrolled during 1996–2001 (scores on the morning and afternoon flights were averaged). The *z*-scores on the individual measures were aggregated on the basis of previous principal component analyses into four component measures (Yesavage et al., 2002; Yesavage, et al., 1999): (1) accuracy of executing the air traffic control (ATC) communications regarding the heading, altitude, radio frequency, and transponder code; (2) traffic avoidance; (3) scanning cockpit instruments to detect engine emergencies; (4) executing a visual approach to landing. A flight summary score, the average of the above four component measures, was used as the primary performance measure. Thus, one global and four component measures of flight performance were assessed. All measures are in scaled sd units.

Practice—For a given flight, the number of unassisted flights previously completed was used as the measure of practice.

Interval—The interval for a given flight was measured as the number of days since the more recent previous, unassisted flight. Because flights 4 and 5 occur on the same day, we assigned a .2 day interval to flight 5 for all participants.

Processing speed measure—Processing speed was a composite measure of speeded performance during 11 visual scanning and perceptual comparison tasks found in CogScreen-AE (Kay, 1995), in sd units. Performance on all of these tasks is measured as response "throughput," which is the number of correct responses made per minute. The 11 tasks were throughput components of the Pathfinder, Shifting Attention, Symbol Digit Coding, Visual Sequence Comparison, Matching to Sample, and Manikin tasks. Full descriptions of the tasks are available online (http://www.cogscreen.com/) and in the CogScreen-AE manual (Kay, 1995).

Executive function—The Discovery subtest of the Shifting Attention Test (Cogscreen-AE, (Kay 1995)) was used to measure cognitive flexibility (ability to shift to a new rule), and the ability to maintain the set. As in the Wisconsin Card Sorting Test, participants use trial and error to discover which of multiple stimulus dimensions (such as object color) is currently relevant and then use that dimension as the sorting rule until feedback indicates it is no longer relevant (see Taylor et al. 2005) for additional information). Three types of performance were measured: (1) number of completed rule sets, (2) number of failures to maintain set, and (3) the percentage of correct responses. These three performance measures were standardized and averaged into a composite measure of executive function.

Basic IIV measure—Intra-individual variability in basic reaction time was measured with reaction times on the CogScreen-AE Pathfinder task (Kay, 1995), in sd units (see Kennedy, et al., 2013 for details on the creation of this variable). The Pathfinder task is a sequencing and visual scanning task. The participant uses a light pen to: (a) sequentially connect numbers (Pathfinder Number), (b) connect letters in alphabetic order (Pathfinder Letters), and (c) sequence an alternating set of numbers and letters (Pathfinder combined).

Procedures

Prior to the baseline visit for the Stanford/VA Aviation Safety Longitudinal Study (Yesavage, et al., 1999), the pilots completed six two-hour sessions of practice training in the flight simulator. Over the first three training sessions, pilots were gradually introduced to the flight simulator and typical scenarios they would encounter in the actual study. During these training sessions, the research assistant would provide assistance when needed. During the fourth to sixth training sessions, pilots completed three practice flights with no assistance provided. These practice flights, referred to as flight 1, flight 2, and flight 3, were analyzed in the present study. Approximately three weeks after completing flight 3, pilots returned to the lab for the baseline visit. During this 6-hour visit, pilots completed flight 4, CogScreen-AE, and finally flight 5. For each flight, the same four performance components were measured: ATC communication, traffic avoidance, response to engine emergencies, and approach to landing. However, the flight scenarios varied in that the communications were not repeated and the timing of traffic and emergencies varied from flight to flight.

Statistical methods

We developed the STEP (Simultaneous Time Effects on Practice) model for longitudinal data comprised of multiple tests on each individual. The STEP model was designed to explain the jagged, non-monotone patterns of performance exhibited by individual pilots as seen in Figure 1. The STEP model contains a Practice effect that depends on the number of previous practice sessions and a Time Interval or “Time” effect, based on the time since the most recent previous practice session. Scores improve by amount P (Practice effect) with each additional test. The increase with practice is offset by amount T (Time Interval effect) for each additional day since the previous practice session. In the base linear model used here, the expected performance at practice session *j* for subject *i* is:

$$\mathrm{E}\,\mathrm{Y}_{\mathrm{ij}}=\mathrm{I}+R_{\mathrm{i}}+P*\mathrm{prev}_{\mathrm{ij}}+T*\mathrm{diff}_{\mathrm{ij}}$$

where I is an fixed intercept for all subjects and R_i is a random effect for the baseline performance level of subject *i*. For flights $j = 2,\ldots5$, $\mathrm{prev}_{ij} = j-1$ is the number of previous flights at flight *j* and diff_{ij} is the elapsed time between previous flight $j-1$ and current flight *j*. For flight 1, prev_{i1} and diff_{i1} both equal 0 so that performance on flight 1 is a function of the individual’s baseline ability only without practice or interval effects.

In the first phase of designing the STEP model, we examined the model’s fit to the overall flight summary score without covariates. During this phase, we considered alternatives such as using the log of the interval or allowing intervals prior to the most recent one to have an effect. These alternatives either fit the data less well or added complexity to the model without appreciable improvement as reflected by the likelihood ratio test. Similarly, allowing the practice effect to vary by flight number or allowing a distinct initial practice effect of flight 1 on flight 2 also did not significantly improve the fit of the model.

Following this initial development phase, we extended the STEP model by adding covariates, such as age, both as main effects and in interaction with practice (e.g. age by practice) and with interval (e.g. age by interval). We considered using both mixed effects models and general estimating equations (GEE) to fit the step model. Results in all cases were almost identical. Results of mixed effects models, with a random subject effect and fixed practice, interval and covariate effects, are reported here.

Initial analyses fit the STEP model without covariates for each of the five flight performance measures. A second set of analyses added main effects of all covariates to the initial STEP model, using subjects with complete-case data. A final set of analyses examined various interactions added to the main-effect covariate model, including interactions of practice with interval and, for each covariate, separate interactions with practice and with interval. Covariates were age (scaled in years), expertise level, processing speed, IIV, and executive function[1].

Results

Preliminary Results

Two sets of preliminary analyses were conducted. The first set was conducted to check whether there were expertise or age differences in the interval (i.e. number of days) between each flight. Table 2 shows the mean number of days between each flight for all pilots, by expertise level and by age group (split at 60 years for tabulation). Table 3 shows correlation coefficients among covariates, intervals, and simulator scores at flight 1. A significant difference among expertise groups was found in the number of days between flights 3 and 4 ($F(2,260) = 4.35$, $p = .014$), in which the most expert group (CFII/ATP) had significantly more days between these two flights than the experienced group (CFI/IFR). We note that the correlations between expertise and each interval were negligible (r's ≤ .11). Although the correlations of age with each interval were modest (r's ≤ .18), older pilots had significantly fewer days than younger pilots between flights 2 and 3 and between flights 3 and 4.

The second set of preliminary analyses examined whether age or expertise was associated with the cognitive covariates. Consistent with our previous findings, age was correlated with processing speed ($r = -.38$, $p < .001$), IIV ($r = .15$, $p = .026$), and executive function ($r = -.17$, $p = .008$). Expertise was not associated with any cognitive covariate.

Results of STEP model without covariates

Results from the STEP model without covariates indicated a positive effect of practice, in which for each additional flight, flight performance increased by .075 SD ($\beta_2 = .075$, $SE(\beta_2) = .006$, $p < .001$) and a negative effect of interval, in which for each additional day between flights, pilots' performance decreased by .002 SD ($\beta_1 = -.002$, $SE(\beta_1) = .001$, $p < .001$) on overall flight simulator performance. Figure 2 illustrates how the main effects of

---

[1] Because gains related to practice can be thought of as a "memory" effect, we also included a memory measure (CogScreen-AE Symbol-Digit Recall), as a covariate in the main-effects models and separately for potential interactions with practice or interval. Memory was not a significant covariate; it was not associated with the summary score or any of its components and had no significant interactions with practice or interval. Reported results are for models that do not include memory as a covariate in order to avoid overfitting the data.

practice and interval together contribute to the expected flight summary score and how the expected performance depends on the timing of an individual pilot's flight schedule. The fitted practice effect is shown by the immediate vertical increase in expected performance following flight 1. Each subsequent flight occurs at a nadir following an interval during which the benefit of the most recent previous session decays. Flights 1–4 also result in an immediate gain in expected performance due to the practice effect and the added refresh effect of setting the interval back to zero. The jagged, stair-like shape of the underlying or latent STEP model is not directly observed. Instead, we observe a trajectory joining the nadirs of the stair at those times when actual flights take place, as represented by the black dotted line, yielding a variety of patterns. Thus, the pattern of performance for each pilot over time depends on when his or her flights take place. Figure 3 shows the four performance subscores using the same flight schedule (and scale) as seen in the first panel (upper, left corner) of Figure 2. Practice effects are greater for the communication and emergency subscores and smaller for traffic avoidance and approach, as shown by the relative heights of the vertical increase immediately following flight 1. In contrast, interval effects are greatest for approach and negligible for traffic avoidance, as shown by the slopes between flights. Practice effects and interval effects without covariates were very similar to those with covariates, which are described in much greater detail below.

Results of STEP model with covariates

Table 4 reports the results of the STEP model for each of the five flight performance measures with main effects for all covariates. Thus, reported practice, interval and other effects in this model are each adjusted for the effects of all other covariates. Results show a consistent pattern in which practice, expertise level, and processing speed positively impacted flight simulator performance on all measures. Processing speed and expertise level were strong positive covariates. In contrast, age, interval, and IIV reliably diminished flight simulator performance, with the following exceptions: the emergency measure, which was not significantly affected by age or interval; the traffic avoidance measure, which was not significantly associated with interval; and the approach measure, which was not significantly associated with IIV. Executive function had a positive impact only for flight summary score and communications. Results for practice and interval in the covariate model were numerically very similar to those in the no-covariate model shown visually in Figures 2 and 3. However, the covariate effects lower or raise the entire stair pattern of the STEP model according to an individual's covariate values.

Results of STEP model with interaction terms

Lastly, we examined the covariates as potential moderators of the effects in the STEP model. For each covariate separately, we added two interaction terms to the main-effects covariate model: (1) a covariate by practice interaction and (2) a covariate by interval interaction. We detected evidence for five such interactions involving age, processing speed, and IIV with practice and/or interval. As shown in Table 5, these interactions were observed in the traffic avoidance, communication, emergency and approach subscores, as well as the summary score. We also considered interactions between the practice effect and the interval terms, but no significant interactions were detected for any flight measure. As in the preceding main-effects model, reported effects in this model are adjusted for the effects of all other covariates.

Figure 4 shows fitted models for the traffic avoidance subscore, in which the beneficial effects of practice were moderated by age (*B* = −.006, *SE(B)* =.002, *p* = .004) and by processing speed (*B* = .049, *SE(B)* =.023, *p* = .029). Younger pilots and those with faster cognitive processing speed tended to have greater improvement in traffic avoidance with practice than older pilots and those with slower processing speed. At one SD above the mean in age or one SD below the mean in processing speed, the model shows almost no gain from additional practice. Due to the correlation of age and cognitive processing speed (*r* = −.38, p < .001), it is impossible to separate their contributions based on these data. Figure 5 shows fitted models for components of flight simulator performance in which interval effects were moderated by processing speed and IIV. First, for communication performance, slower processing speed was associated with a faster decline of performance during the interval (*B* = .003, *SE(B)* = .001, *p* = .024). Analogously, less stable IIV led to faster declines in overall performance (*B* = −.001, *SE(B)* = .001, *p* = .038) and approach performance (*B* = −.004, *SE(B)* = .001, *p* = .003). Even with the inclusion of these interaction terms in the STEP model, the main effects of practice and interval, and the covariates age, expertise, processing speed, executive function and IIV, remained strong when they were also present in the covariate main-effects only model.

Discussion

In a simulated task designed to assess real world performance among middle-aged and older adults with specialized training, consistent practice and interval effects were found. The communications, traffic avoidance, and emergency components especially benefited from practice, in comparison to the practice effect observed on the approach component of flight performance. The communications and approach components were most sensitive to longer time intervals between practice flights. Notably, communications was strongly influenced by both practice and interval. Like many real-world tasks, different flight task components, such as communication and traffic avoidance, involve a varying mix of motor reflexes, working memory and strategy for successful execution. Thus, results suggest that frequent training with relatively short intervals are beneficial for cognitively demanding memory tasks, such as communications. For more reflexive tasks, such as traffic avoidance, the number of practice sessions appears to be key while interval is less important. For perspective, the STEP model estimates that, for the traffic component, nearly twice as many days could elapse before the gain from the previous training flight would completely fade.

All four flight components showed the influence of age. Regarding age and moderation of practice effects, the benefits of practice diminished with age for traffic avoidance performance, extending findings based on laboratory paper and pencil tests to tasks simulating complex real world activities (Calamia et al,2012). On the other hand, we did not observe that practice effects diminished with age for the other three flight components. Importantly, the number of practice sessions were found be especially important for those pilots with relatively slow processing speed.

Additionally, the length of interval was especially important for those pilots with either relatively slow processing speed or high IIV. Pilots with relatively slow processing speed or relatively high IIV were adversely affected by longer intervals whereas the higher cognitively functioning counterparts were relatively immune to interval effects. These results suggest that people with slow processing speed or unstable IIV require shorter intervals between training sessions to maintain or improve performance on complex real world tasks.

Expertise level did not impact practice or interval effects among this sample of pilots. Expert pilots consistently maintained their performance advantage regardless of the number of previous practice sessions or length of interval. This result is consistent with the work of Ericsson (2006). According to Ericsson (2006), it is the use of deliberate practice rather than routine practice (as was completed by pilots in this study) that leads to an increased advantage in performance in experts compared to less experienced pilots.

It is possible that processing speed and IIV may be better indicators of a person’s need for additional practice, or for less time in between practice sessions, than chronological age. In this sample of well educated, healthy adults, age was correlated with both of these cognitive measures. However, processing speed and IIV were the two variables that more frequently interacted with practice and/or interval than age per se. If processing speed and IIV are modifiable cognitive abilities, our results suggest that the combination of processing speed and IIV training with task-specific practice could lead to significant improvements in performance of real world tasks among middle-aged and older adults. This targeted combination of training may be particularly helpful for reflexive tasks such as traffic avoidance, which appears to benefit from training, especially for those with faster processing speed, yet is relatively immune to fading of training with time. From a practical standpoint, the multitude of apps that aim to improve processing speed and other cognitive abilities, as well as relatively inexpensive home flight simulators, mean that pilots can complete this type of training in their own homes. Driving is one real world task in which improved or even maintained performance through this type of intervention could enable older adults to drive safely for longer periods of time than without completing such an intervention. Aspects of driving that place demands on speed of processing, divided attention, and visual-perceptual processing are most susceptible to age-related declines in performance, such as steering control (Ni, Andersen, McEvoy, & Rizzo 2005), staying a constant distance behind a lead car (Dashtrup, Lees, Dawson, Lee, & Rizzo, 2009) and collision detection, particularly at faster speeds or in fog (Anderson, Cisneros, Saidpour, & Atchley, 2000 & Bian, Ni, Guindon, & Andersen, 2009). Thus, a combination of cognitive training with driving simulations that focus on these driving skills may be most beneficial for older drivers. The STEP model also may have application in other settings, such as in the design of regimens intended to maintain older adults’ professional competency or adaptation to new technology used in everyday life.

Results also provide support for the general utility of the STEP model. The model can distinguish separate effects of practice and interval and can evaluate their relative contributions to performance on the diverse components of a complex real-world task. Additionally, the estimated practice and interval effects could be incorporated into models that seek to determine optimal training schedules for learning and retention such as the ACT-R model (Pavlik & Anderson, 2008; Schmidt & Bjork, 1992). The STEP model could be tested on repeated sessions involving deliberate practice among adults with different levels of domain-related expertise (Jastrzembski, Addis, Krusmark, Gluck, & Rodgers, 2010). The model also can be used to identify other relevant person characteristics that moderate the effects of either practice or interval on performance. Furthermore, the STEP model realistically captures the variable and inconstant performance trajectories seen in many real-world contexts.

Limitations of this study include a methodological design that does not enable us to tease apart the contributions of age and cognitive processing speed on the effects of practice. Additionally, although five practice sessions were examined, the effects of practice and interval cannot be generalized beyond five sessions. Furthermore, each practice session used an alternate form for the flight scenario, which may have reduced the magnitude of the practice effect. The moderator analyses also involved a large number of interaction tests and should be viewed as exploratory. Results also should be interpreted with caution as expertise and age differences were found to be correlated with the intervals between some pairs of flights. Furthermore, the linear model used for interval effects was chosen for its statistical stability, rather than its theoretical plausibility. Finally, the effects of practice and interval on flight simulator performance may differ from real world tasks, such as work-related training.

In summary, results from this study extend current knowledge in both the cognitive testing and applied psychology fields regarding practice and interval effects on an adult sample with specialized training. We demonstrate that practice effects are beneficial in a task assessing real world performance and that long intervals can be particularly damaging to performance for adults with relatively slow processing speed or high IIV. Because practicing a task requires sustained attention, these results point to the need to examine IIV as a moderator of practice and interval effects. Intervals -- and their effect on performance -- can vary considerably within and between persons in real world tasks. The STEP model provides a method for characterizing the simultaneous effects of practice and interval under these circumstances. Finally, consistent with previous work, basic cognitive functions, such as processing speed, executive function and IIV appear to underlie age-related declines in real world performance (Causse, et al., 2011; Kennedy, et al., 2013; Taylor, et al., 2005; Yesavage, et al., 2011) (Kennedy, et al., 2013; Shinar, et al., 2005; Yesavage, et al., 2011). Results have implications for the design, implementation, and assessment of training interventions targeted at middle-aged and older adults for complex real world tasks.

Acknowledgments

This research is supported in part by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, the Department of Veterans Affairs Sierra-Pacific Mental Illness Research, Education, and Clinical Center (MIRECC), and by Grant Number R37 AG 12713 from the National Institute on Aging. These sponsors solely provided financial support or facilities to conduct the study. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Aging or the National Institutes of Health. We thank Katy Castile, Jake Volz, and Daniel Heraldez for assistance in data preparation.

References

Bell HH, Waag WL. Evaluating the effectiveness of flight simulators for training combat skills: A review. The International Journal of Aviation Psychology. 1998; 8(3):223–242.

Bielak AA, Hultsch DF, Strauss E, Macdonald SW, Hunter MA. Intraindividual variability in reaction time predicts cognitive outcomes 5 years later. Neuropsychology. 2010a 2010-19104-001. 10.1037/a0019802

Bian, Z.; Ni, R.; Guindon, A.; Andersen, G. Aging and the detection of collision events in fog. Proceedings from the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design; 2009. p. 69-75.

Bielak AA, Hultsch DF, Strauss E, MacDonald SW, Hunter MA. Intraindividual variability is related to cognitive change in older adults: evidence for within-person coupling. Psychology and Aging. 2010b; 25(3):575–586. 2010-18944-004. 10.1037/a0019503 [PubMed: 20853965]

Calamia M, Markon K, Tranel D. Scoring higher the second time around: meta-analyses of practice effects in neuropsychological assessment. Clinical Neuropsychology. 2012; 26(4):543–570.10.1080/13854046.2012.680913

Causse M, Dehais F, Pastor J. Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots. International Journal of Aviation Psychology. 2011; 21(3): 217–234.

Cooper DB, Lacritz LH, Weiner MF, Rosenberg RN, Cullum CM. Category fluency in mild cognitive impairment: reduced effect of practice in test-retest conditions. Alzheimer Disease & Associated Disorders. 2004; 18(3):120–122. 00002093-200407000-00003. [PubMed: 15494616]

Dashtrup, E.; Lees, M.; Dawson, J.; Lee, J.; Rizzo, M. Differences in simulated car following behavior of younger and older drivers. Proceedings from the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design; 2009. p. 76-82.

Dodge HH, Wang CN, Chang CC, Ganguli M. Terminal decline and practice effects in older adults without dementia: the MoVIES project. Neurology. 2011; 77(8):722–730.10.1212/WNL.0b013e31822b0068 [PubMed: 21832224]

Donovan JJ, Radosevich DJ. A Meta-analytic review of the distribution of practice effect: Now you see it, now you don't. Journal of Applied Psychology. 1999; 84(5):795–805.

Ericsson, KA. The influence of experience and deliberate practice on the development of superior expert performance. In: Ericsson, KA.; Charness, N.; Hoffman, RR.; Feltovich, PJ., editors. The Cambridge Handbook of Expertise and Expert Performance. New York, NY: Cambridge University Press; 2006. p. 685-705.

Federal Aviation Administration. Approval of flight simulators and flight training devices. 2014 FAR Sec125.297.

Ferrer E, Salthouse TA, Stewart WF, Schwartz BS. Modeling age and retest processes in longitudinal studies of cognitive abilities. Psychology and Aging. 2004; 19(2):243–259.10.1037/0882-7974.19.2.2432004-14948-001 [PubMed: 15222818]

Hasher, L.; Zacks, RT., editors. Working memory, comprehension, and aging: A review and a new view. Vol. 2. New York, NY: Academic Press; 1988.

Hausknecht JP, Halpert JA, Di Paolo NT, Moriarty Gerrard MO. Retesting in selection: a meta-analysis of coaching and practice effects for tests of cognitive ability. Journal of Applied Psychology. 2007; 92(2):373–385. 2007-03270-007. [PubMed: 17371085]

Hausknecht JP, Trevor CO, Farr JL. Retaking ability tests in a selection setting: implications for practice effects, training performance, and turnover. Journal of Applied Psychology. 2002; 87(2): 243–254. [PubMed: 12002953]

Horn, JL. The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In: Trehub, FIMCS., editor. Aging and Cognitive Processes. New York, NY: Plenum; 1982. p. 237p. 230p. 278

Jamieson BA, Rogers WA. Age-related effects of blocked and random practice schedules on learning a new technology. Journal of Gerontology Series B: Psychological Sciences and Social Sciences. 2000; 55(6):P343–353.

Jastrzembski, TS.; Addis, K.; Krusmark, M.; Gluck, KA.; Rodgers, S. Prediction intervals for performance prediction. Paper presented at the Proceedings of the 10th International Conference on Cognitive Modeling; Philadelphia, PA. 2010.

Kay, GG. CogScreen Aeromedical edition professional manual. Odessa, FL: Psychological Assessment Resources; 1995.

Kennedy Q, Taylor J, Heraldez D, Noda A, Lazzeroni LC, Yesavage J. Intraindividual variability in basic reaction time predicts middle-aged and older pilots' flight simulator performance. Journal of Gerontology Series B: Psychological Sciences and Social Sciences. 2013; 68(4):487–494.10.1093/geronb/gbs090

Momen N. The effects of alternative input devices and repeated exposures on the test of basic aviation skills (TBAS) performance. Military Medicine. 2009; 174(12):1282–1286. [PubMed: 20055069]

Mumford MD, Costanza DP, Baughman WA, Threlfall KV, Fleishman EA. Influence of abilities on performance during practice: Effects of massed and distributed practice. Journal of Educational Psychology. 1994; 86(1):134–144.

Ni, R.; Andersen, GJ.; McEvoy, S.; Rizzo, M. Age related decrements in steering control: The effects of landmark and optical flow information; Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design; 2005. p. 143-149.

Pavlik PI, Anderson JR. Using a model to compute the optimal schedule of practice. Journal of Experimental Psychology: Applied. 2008; 14(2):101–117. [PubMed: 18590367]

Puddey IB, Mercer A, Andrich D, Styles I. Practice effects in medical school entrance testing with the undergraduate medicine and health sciences admission test (UMAT). BMC Medical Education. 2014; 14:48–63. [PubMed: 24618021]

Rabbitt P, Diggle P, Smith D, Holland F, Mc Innes L. Identifying and separating the effects of practice and of cognitive ageing during a large longitudinal study of elderly community residents. Neuropsychologia. 2001; 39(5):532–543. S0028-3932(00)00099-3. [PubMed: 11254936]

Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychological Review. 1996; 103(3):403–428. [PubMed: 8759042]

Salthouse TA. Effects of age on time-dependent cognitive change. Psychological Science. 2011; 22(5):682–688.10.1177/0956797611404900 [PubMed: 21467547]

Salthouse TA, Tucker-Drob EM. Implications of short-term retest effects for the interpretation of longitudinal change. Neuropsychology. 2008; 22(6):800–811. [PubMed: 18999354]

Schmidt RA, Bjork RA. New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. Psychological Science. 1992; 3(4):207–217.

Shinar D, Tractinsky N, Compton R. Effects of practice, age, and task demands, on interference from a phone task while driving. Accident Analysis & Prevention. 2005; 37(2):315–326. S0001-4575(04)00096-X. [PubMed: 15667818]

Taylor JL, Kennedy Q, Adamson MM, Lazzeroni LC, Noda A, Murphy GM, Yesavage JA. Influences of APOE epsilon4 and expertise on performance of older pilots. Psychology and Aging. 2011; 26(2):480–487.10.1037/a0021697 [PubMed: 21668123]

Taylor JL, Kennedy Q, Noda A, Yesavage JA. Pilot age and expertise predict flight simulator performance: a 3-year longitudinal study. Neurology. 2007; 68(9):648–654. [PubMed: 17325270]

Taylor JL, O'Hara R, Mumenthaler MS, Rosen AC, Yesavage JA. Cognitive ability, expertise, and age differences in following air-traffic control instructions. Psychology and Aging. 2005; 20(1):117–133. [PubMed: 15769218]

U.S. Department of Transportation, F. A. A. Line operational simulations: Line oriented flight training, special purpose operational training, line operational evaluation. 1990.

Yesavage JA, Jo B, Adamson MM, Kennedy Q, Noda A, Hernandez B, Taylor JL. Initial cognitive performance predicts longitudinal aviator performance. Journal of Gerontology Series B: Psychological Sciences. 2011; 66(4):444–453.10.1093/geronb/gbr031

Yesavage JA, Mumenthaler MS, Taylor JL, Friedman L, O'Hara R, Sheikh J, Whitehouse PJ. Donepezil and flight simulator performance: effects on retention of complex skills. Neurology. 2002; 59(1):123–125. [PubMed: 12105320]

Yesavage JA, Taylor JL, Mumenthaler MS, Noda A, O’Hara R. Relationship of age and simulated flight performance. Journal of the American Geriatric Society. 1999; 47(7):819–823.

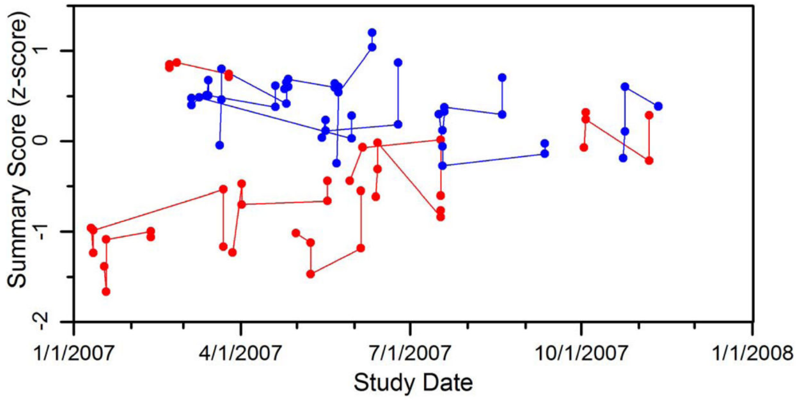

Figure 1.

Summary score trajectories across the five flights for a subset of pilots, those seen between January 2007 and January 2008. Circles represent summary score and lines connect summary scores from consecutive flights. Red indicates summary score trajectories of pilots aged 60+ years; blue indicates those of pilots younger than 60 years. Many pilots exhibit a jagged non-monotone pattern of performance.

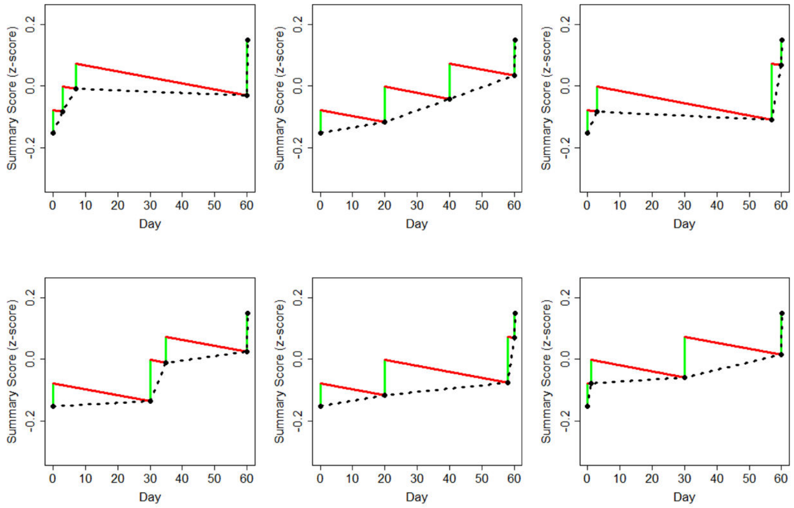

Figure 2.

Fitted STEP model results without covariates for overall flight simulator performance. Figure 2 shows expected performance of six hypothetical pilots with different intervals between flights 1 – 4. The flight schedule in the upper left panel was the most typical and therefore is used in Figures 3 – 5: For 143 pilots (54%), flight 3 occurred within 10 days of flight 1 and 20 to 90 days before flight 4. Note that flights 4 and 5 always occurred on the same day. Solid black dots occur at the time of actual flights and show predicted performance for that flight. The solid lines reflect the underlying STEP model learning process. Solid green vertical lines at flight times show the practice effect immediately following a flight and the refresh of the interval effect (if any) following a flight. Solid red lines between flight times show the interval effect. Dotted black lines connect expected performance at hypothetical flight times, and the trajectory that would be observed as a consequence of the underlying process.

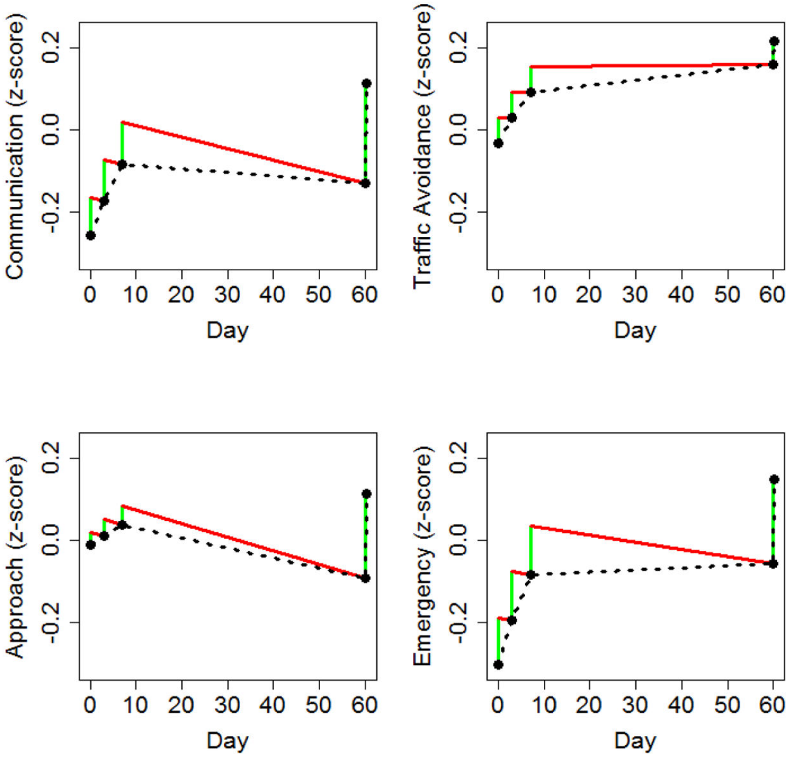

Figure 3.

STEP model results for each individual flight performance measure without covariates, using the same flight schedule as in the upper left panel of Figure 2.

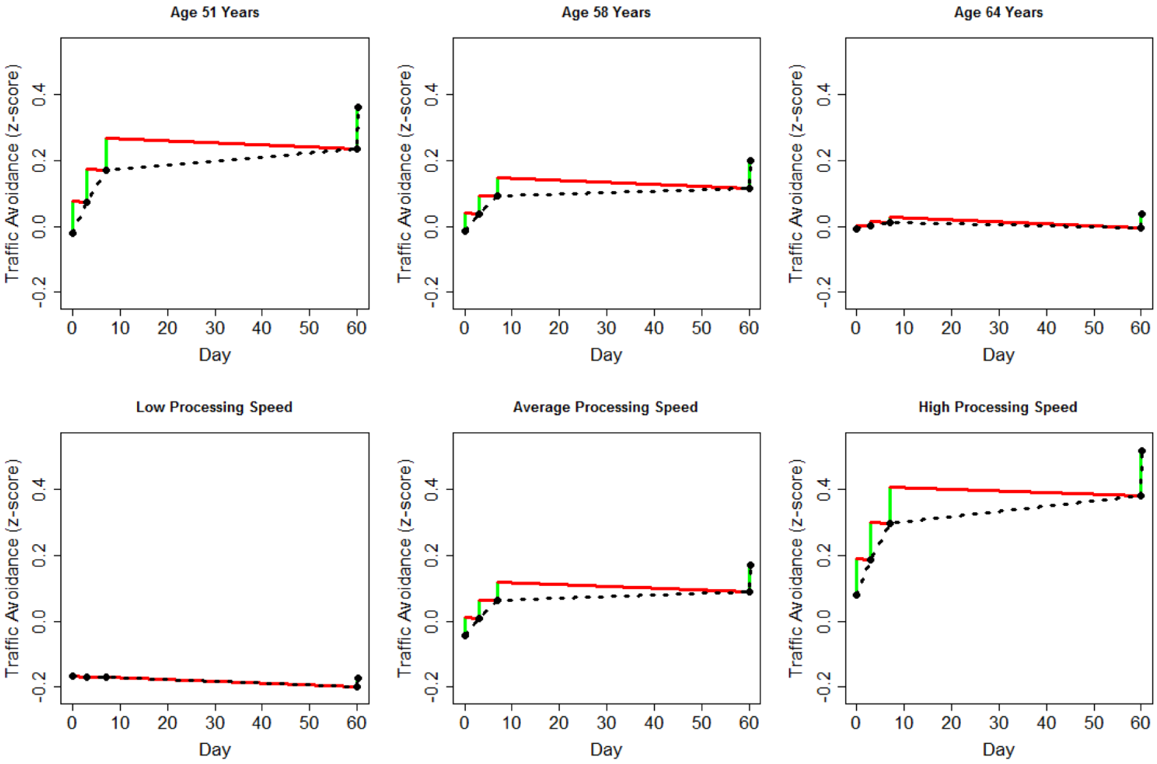

Figure 4.

Results for interaction effects with practice of age and cognitive processing speed on traffic avoidance. For age, predictions are shown for an individual at age 51, one SD below the sample mean; at age 58, at the sample mean; and at age 64, at one SD above the sample mean. For processing speed, predictions are for an individual one SD below the mean; at the mean; and one SD above the mean based on the values of a reference sample enrolled early in the study. Older adults and those with slower processing speed do not benefit from practice.

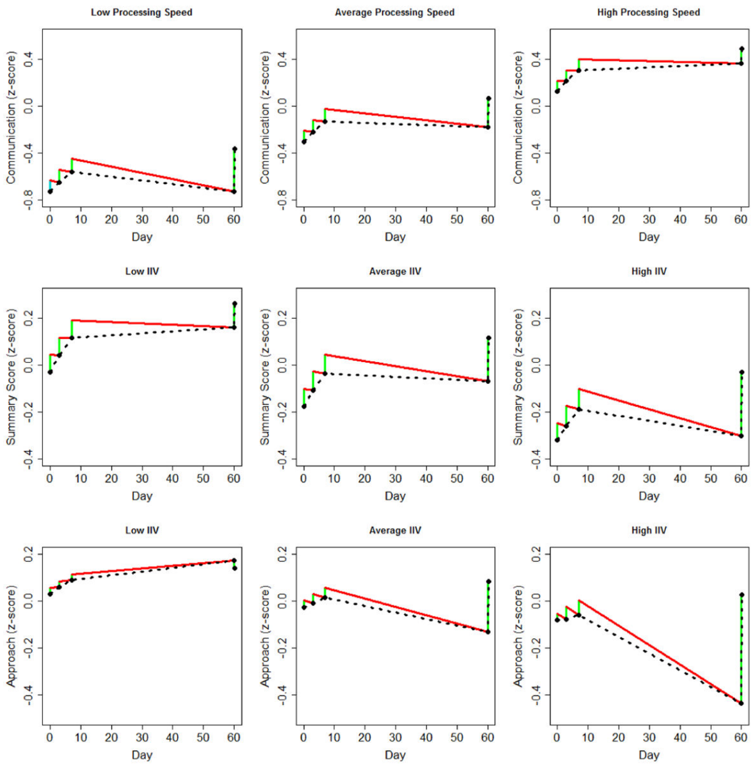

Figure 5.

Results for interaction effects with interval of processing speed and IIV for the communications measure (top row), summary score (middle row) and approach measure (bottom row). Predictions are shown for an individual with low processing speed or IIV one SD below the mean (left column); with average (mean) processing speed or IIV (middle column); and with high processing speed or IIV one SD above the mean (right column) based on the values of a reference sample enrolled early in the study. Those with fast processing speed or low IIV are relatively unaffected by interval, whereas the deleterious effect of interval is evident in those with slow processing speed or high IIV.

Table 1

Participants' Demographics and Flight Experience Characteristics by Level of Expertise.

	Pilot expertise level		
	Least (*n* = 65)	**Moderate (*n* = 145)**	**High (*n* = 53)**
Age in years mean (SD)	56.7 (7.3)	58.7 (6.3)	55.5 (6.6)
Years of education, mean (SD)	16.6 (2.3)	17.1 (2.0)	17.2 (1.9)
Women, *n* (%)	11 (16.9)	19 (13.1)	4 (7.6)
White, non-Hispanic, %	88%	99%	92%
Total log hours, mean (SD)	989 (1328)	1927 (1927)	5262 (2947)
Log hours in past month, mean (SD)	5.5 (7.2)	8.6 (10.2)	14.9 (17.3)

Table 2

Intervals (number of days) between flights, mean (sd), min - max range.

	No. days between Flights 1 and 2	**No. days between Flights 2 and 3**	**No. days between Flights 3 and 4**
Total group (*n* = 263)	7.02 (13.07) .2 – 140	5.38(8.92) .2 – 61	35.03 (16.72) 2 – 103
Expertise group			
VFR (*n* = 65)	8.44(13.98) .2 – 75	6.51 (9.75) .2 – 51	34.94 (15.23) 3 – 87
CFI/ IFR (*n* = 145)	7.37 (14.24) .2 – 140	4.64 (8.20) .2 – 61	32.97 (15.61)[a] 2 – 95
CFII/ATP (*n* = 53)	4.31 (6.93) .2 – 44	6.05 (9.67) .2 – 42	40.79 (20.07)[a] 15 – 103
Age group			
Under 60 years (*n* = 158)	7.06 (10.83) .2 – 75	6.79 (10.13)[b] .2 – 61	36.23 (17.16)[b] 3 – 103
60^+ years (*n* = 105)	6.96 (15.91) .2 to 140	3.26 (6.15)[b] .2 – 43	33.24 (15.95)[b] 2 – 74

[a] Significant expertise group difference at the .05 alpha level.

[b] Significant age group difference at the .05 alpha level.

Table 3

Correlation matrix (with p-values) of covariates, performance on flight 1measures, and interval between flights.

	Expertise	**Age**	**Sum score**	**Comm**	**Traffic Avoidance**	**Approach**	**Emergency**	**Proc Speed**	**Executive Function**	**IIV**	**Interval 1 to 2**	**Interval 2 to 3**	**Interval 3 to 4**
Expertise	1.00												
Age	−.05	1.00											
Summary score	$.33^{***}$	$-.27^{***}$	1.00										
Communication	$.37^{***}$	$-.28^{***}$	$.71^{***}$	1.00									
Traffic Avoidance	.09	−.005	$.58^{***}$	$.20^{**}$	1.00								
Approach	$.19^{**}$	$-.18^{**}$	$.65^{***}$	$.33^{***}$	$.20^{**}$	1.00							
Emergency	$.23^{***}$	$-.23^{***}$	$.70^{***}$	$.41^{***}$	$.19^{**}$	$.18^{**}$	1.00						
Proc speed	.05	$-.38^{***}$	$.40^{***}$	$.42^{***}$	$.13^{*}$	$.24^{***}$	$.28^{***}$	1.00					
Executive function	.05	$-.17^{**}$	$.23^{***}$	$.25^{***}$	.08	.08	$.21^{**}$	$.22^{***}$	1.00				
IIV	.02	$.15^{*}$	$-.16^{*}$	$-.19^{**}$	−.01	−.06	$-.17^{*}$	.02	−.12	1.00			
Interval 1 to 2	−.10	−.07	−.08	−.03	−.07	−.02	−.09	.03	.01	.09	1.00		
Interval 2 to 3	−.02	$-.18^{**}$	−.06	−.06	−.03	−.10	.03	−.003	$.16^{*}$	−.09	.04	1.00	
Interval 3 to 4	.11	$-.15^{*}$	$.13^{*}$	.08	.08	$.17^{*}$	.02	.04	−.01	.01	.05	−.06	1.00

[*] $p < .05$,

[**] $p < .01$,

[***] $p < .001$

Table 4

Results from the STEP model with covariates. Reported effects are adjusted for effects of all other covariates.

Variable (unit)	Summary Score Parameter Estimate (SE)	Communication Parameter Estimate (SE)	Traffic Avoidance Parameter Estimate (SE)	Emergency Parameter Estimate (SE)	Approach Parameter Estimate (SE)
Intercept (mean, I)	.324 (.258) (p = .210)	.476 (.373) (p = .201)	.412 (.336) (p = .220)	−.153 (.505) (p = .762)	.554 (.392) (p = .158)
Interval (day)	−.002 (.001) (p < .001)	−.003 (.001) (p < .001)	−.001 (.001) (p =.671)	−.002 (.001) (p = .167)	−.004 (.001) (p = .001)
Practice (event)	.074 (.007) (p < .001)	.092 (.009) (p < .001)	.053 (.014) (p < .001)	.119 (.014) (p < .001)	.030 (.013) (p = .024)
Age (year)	−.015 (.004) (p < .001)	−.022 (.006) (p < .001)	−.012 (.006) (p = .036)	−.011 (.008) (p = .190)	−.016 (.006) (p = .015)
Expertise (level)	.206 (.038) (p < .001)	.273 (.055) (p < .001)	.123 (.050) (p = .013)	.244 (.075) (p = .001)	.181 (.058) (p = .002)
Cognitive processing speed (sd)	.338 (.048) (p < .001)	.447 (.070) (p < .001)	.234 (.063) (p < .001)	.358 (.094) (p < .001)	.314 (.073) (p < .001)
Executive function (sd)	.064 (.029) (p = .029)	.083 (.042) (p = .049)	.028 (.038) (p = . 465)	.083 (.057) (p = .147)	.061 (.044) (p = .168)
IIV (sd)	−.162 (.036) (p < .001)	−.247 (.053) (p < .001)	−.118 (.047) (p = .013)	−.181 (.071) (p = .011)	−.102 (.055) (p = .066)

Table 5

Results from the STEP model with significant interaction terms.

Variable (unit)	Summary Score Parameter Estimate (SE)	Communication Parameter Estimate (SE)	Traffic Avoidance Parameter Estimate (SE)	Emergency Parameter Estimate (SE)	Approach Parameter Estimate (SE)
Intercept (mean, I)	.323 (.259) (p =.212)	.478 (.373) (p =.200)	.409 (.337) (p = .224)	−.352 (.405) ((p =.386)	.550 (.393) (p = .161)
Interval (day)	−.002 (.001) ($p < .001$)	−.003 (.001) ($p < .001$)	−.001 (.001) (p = .663)	.009 (.009) (p =.317)	−.004 (.001) (p = .002)
Practice (event)	.073 (.007) ($p < .001$)	.092 (.009) ($p < .001$)	.054 (.014) ($p < .001$)	.384 (.117) (p =.001)	.028 (.013) (p = .034)
Age (year)	−.015 (.004) ($p < .001$)	−.022 (.006) ($p < .001$)	−.012 (.006) (p = .037)	.002 (.007) (p =.818)	−.016 (.007) (p = .016)
Expertise (level)	.206 (.038) ($p < .001$)	.272 (.055) ($p < .001$)	.122 (.050) (p = .014)	.123 (.050) ($p < .001$)	.182 (.058) (p = .002)
Cognitive processing speed (sd)	.338 (.048) ($p < .001$)	.461 (.075) ($p < .001$)	.114 (.077) (p = .138)	.234 (.063) ($p < .001$)	.315 (.073) ($p < .001$)
Executive function (sd)	.064 (.029) (p = .030)	.084 (.042) (p =.047)	.028 (.038) (p = .457)	.027 (.038) (p =.476)	.061 (.044) (p = .165)
IIV (sd)	−.136 (.041) (p = .001)	−.247 (.053) ($p < .001$)	−.118 (.047) (p = .013)	−.118 (.047) (p =.013)	−.019 (.066) (p = .774)
IIV x Interval	−.001 (.001) (p = .038)	N/A	N/A	N/A	−.004 (.001) (p = .003)
Processing speed x Interval	N/A	.003(.001) (p = .024)	N/A	N/A	N/A
Processing speed x Practice	N/A	N/A	.049 (.023) (p = .029)	N/A	N/A
Age x Practice	N/A	N/A	N/A	−.006 (.002) (p = .004)	N/A

Note: Although all interaction terms were run for each flight performance measure, this table shows the results when only the significant interaction term was included to the model with covariates. N/A: Not applicable because the result was not significant.