EXPERTISE AND AGE EFFECTS ON PILOT MENTAL WORKLOAD IN A SIMULATED AVIATION TASK

Donald L. Lassiter
Methodist College
Fayetteville, NC

Daniel G. Morrow
University of New Hampshire
Durham, NH

Gary E. Hinson
Methodist College
Fayetteville, NC

Michael Miller
Catholic University of America
Washington, DC

David Z. Hambrick
Georgia Institute of Technology
Atlanta, GA

ABSTRACT

This study investigated the effects of expertise and age on cognitive resources relevant to mental workload of pilots engaged in simulated aviation tasks. A secondary task workload assessment methodology was used, with a PC-based flying task as the primary task, and a Sternberg choice reaction time task as the secondary task. A mixed design using repeated measures was employed, with age and expertise as between-subjects factors and workload as the within-subjects factor. Pilots ranging in age from 21 to 79 years and 28 to 11,817 hours of flight time served as subjects. Of interest was whether expertise would mitigate the adverse effect of aging on pilots' mental workload handling ability as defined by two measures of secondary task performance: choice reaction time and accuracy. Results indicated that expertise did mitigate the effects of age regarding secondary task accuracy. Implications of results are discussed, and directions for future research are presented.

INTRODUCTION

The task of flying involves the time-sharing of several tasks, i.e., it is a multiple-task situation, placing a great deal of mental workload demand on the pilot. The concept of mental workload implies that limitations exist in the human information processing framework (Gopher & Donchin, 1986). It has often been reported in the aging literature that physical abilities, perceptual processes, and memory processes decrease with age (e.g., see Wickens, Braune, & Stokes, 1987; Tsang, 1992). To some degree, laboratory research with non-pilots has indicated that aging detrimentally effects the ability of the human information processing system to handle appreciable amounts of mental workload.

A frequent finding in the expertise literature is that expertise improves performance of domain-relevant tasks by reducing workload demands on short-term memory capacity (e.g., see Yekovich, Walker, Ogle, & Thompson, 1990). Domain-relevant tasks are tasks that tap the specific encapsulated knowledge within an expert's particular area of expertise, or domain (Rybash, Hoyer, & Roodin, 1986), whereas domain-general tasks do not tap a specific area of knowledge, but rather draw on generalized background knowledge and ability. To access the knowledge to perform a domain-relevant task might require less effort for an expert than a novice, who may not possess the required knowledge in as useful a form to perform the same task. This is probably because the knowledge possessed by the novice is far less structured and automatized than that of the expert (Rybash et al., 1986).

Flying can be considered a well-defined task domain, so it may be that expertise in flying could counteract the detrimental effects of aging resulting from increases in mental workload of domain-relevant aviation tasks. The number of studies looking at the relationships among expertise, aging, and workload involving pilots as subjects has been small, but is now increasing (see Morrow, Leirer, & Altieri, 1992; Morrow, Leirer, Altieri, & Fitzsimmons, 1994; Tsang & Shaner, 1994; Tsang, 1995). However, there is still a need for systematic research programs involving pilots to investigate the relationships among these variables (Tsang, 1992).

Two specific aims of the current research effort were to determine: 1) if aging adversely affects the ability to handle increases in mental workload in a simulated aviation task; and 2) if expertise in piloting (as defined by the number of hours of flight time) can mitigate the adverse effect of aging on the ability to handle increases in mental workload. Such mitigation would be demonstrated by a significant interaction between expertise and age regarding the ability to handle mental workload.

METHOD

This study utilized the secondary task workload methodology (i.e., the subsidiary task technique) using a Sternberg (1969) choice reaction time task as the secondary task, and flying courses of two different levels of difficulty on Microsoft's Flight Simulator 4.0 (Microsoft, 1990) as the primary task. The Sternberg task has been used extensively as a secondary task in mental workload research and has shown sufficient sensitivity and diagnosticity (e.g., Wickens et al., 1987). The rationale for picking this task was that it may tap cognitive resources related to working memory and monitoring and compete with the primary task for these resources. The major independent variables in this study were age (in years), expertise (in hours of total flight time), and workload. Workload consisted of six levels based on combinations of levels of difficulty of the primary task (3 levels) and secondary task (2 levels). The three levels of primary task were zero (i.e., it was not performed with the secondary task; the secondary task was performed alone); easy course, and difficult course. The two levels of secondary task were 2 letters (low load) and 4 letters (high load) in the memory set. Therefore, the six levels of workload were: 1) secondary task alone (low load); 2) secondary task alone (high load); 3) primary task (easy course) with secondary task (low load); 4) primary task (easy course) with secondary task (high load); 5) primary task (difficult course) with secondary task (low load); and 6) primary task (difficult course) with secondary task (high load). The two major dependent variables in the study reflected secondary task performance: choice reaction time (in msecs) and accuracy (percent correct responses).

Subjects

A total of 42 paid volunteers served as subjects in this study. All had general aviation experience. The age range was 21 to 79 years, while the range of expertise was 28 to 11,817 hours of flight time. Subjects were recruited from the local general aviation community, primarily from a large local flying club. Other subjects came from the sizable active/retired military aviation community in the area. All subjects had general aviation experience, and all but three subjects had such experience within the twelve months prior to participation in this study. None of the subjects reported having prior laboratory testing, although those with military experience reported having simulator time. None of the subjects reported having any experience with the PC-based flying simulation used in this study.

Apparatus

Equipment for this study consisted of two computers for presenting the primary and secondary tasks, as well as collecting and analyzing data. The lab was partitioned into subject and experimenter stations. The subject station (see Figure 1) had two monitors, one each for the primary and secondary tasks, as well as a flight yoke and pedals for performing the primary task. Two large telegraph keys were mounted next to the yoke for responding to the secondary task. The experimenter station had two monitors which allowed the experimenter to see what the subject viewed, as well as mice/keyboards for controlling the simulation.

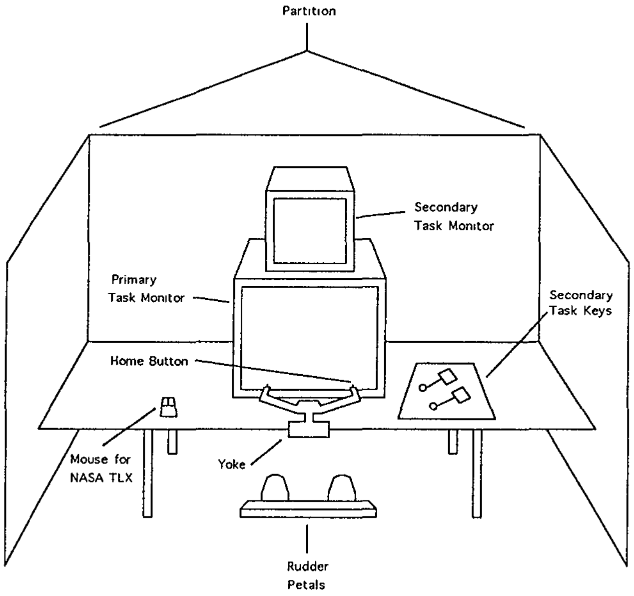


Figure 1. The subject station. The mouse was used to collect data not reported here in another phase of the study.

Procedure

Subjects participated in a total of seven two-hour experimental sessions: five practice sessions and two data collection sessions. Subjects were run one at a time, one session per subject per day. The first practice session consisted of a standardized briefing (as the subject read along), updating the subject's biographical information form, and the subject signing a consent form. Following instructions read aloud by the experimenter, the subject became familiar with the simulator apparatus and primary task software. Then the subject practiced the primary task, flying eight ten-minute courses once (four easy and four difficult courses). Before practicing each course, the subject was given a map of the course and the experimenter read aloud detailed instructions describing that course. During this first session, the software program monitoring the subject's performance did not interrupt the session if performance was not up to criterion. Feedback concerning performance was given verbally to the subject. After this first session, the subject was given maps of the eight courses and instructed to memorize them before the second session. An example of a course map is shown in Figure 2. The purpose of memorizing these courses was to have subjects use their working memory to call up and maintain a representation of the course as they flew it, therefore competing with working memory demand of the secondary task (described below). The subject's memory was tested before flying each course in subsequent sessions by having the

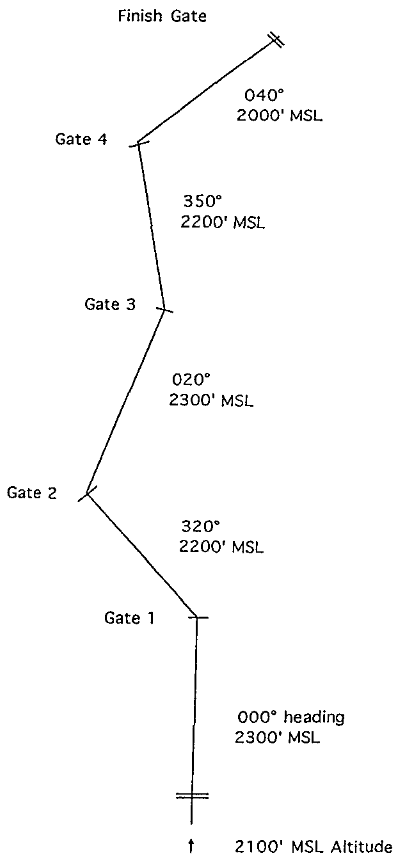


Figure 2. A map of one of the eight courses used in the study.

subject sketch the course map from memory without error. As sessions proceeded, subjects had no difficulty reproducing the maps. An easy course consisted of five two-minute splits, where at each of the four middle splits, the subject had to execute a change in heading of 60 degrees. Gates located at the positions of the course changes provided guidance, but were not visible to the subject until they were only a short distance away to insure that the subject relied on instruments and memory to fly the course. For these easy courses, all gates were at the same altitude. A difficult course was the same length as an easy course, but the subject had to make different changes in heading at each gate, as well as change altitude. Four variants of the courses within each level of difficulty were constructed by using mirror and inverted rotations. The second and third practice sessions each gave the subject ten courses to fly. From the ninth course flown in the second session onward, the software program monitoring the subject's performance interrupted and terminated the flight if the subject did not fly to criterion, and the subject was required to start that course over. The criterion was defined as flying each 20-mile course with no more than a 2000 feet deviation from ideal course, a 20 degree deviation from course (to allow for turns), a 200 feet deviation from assigned altitude, and a 10kts/hr deviation from assigned airspeed. The fourth practice session contained four additional ten-minute runs of primary task practice. Thus, over the first three and a half sessions of the experiment, 32 total practice runs of the primary task were performed. The purpose of the extensive practice was to insure that subjects flew the primary task at criterion and could maintain that performance during the dual-task sessions of primary and secondary task performance that came later. (The secondary task must not intrude on primary task performance; otherwise, any differences in secondary task performance between single- and dual-task conditions would not be interpretable). Then, after a standardized set of instructions read aloud, the subject was given four 10-minute blocks of practice on the secondary task by itself (two blocks each of memory sets of two and four letters). The subject was given the letters to memorize right before each run. On a trial, the subject was presented a brief flash of a single letter on the secondary task monitor. While keeping the right thumb on the red "home" button located on the right handle of the flight yoke, the subject decided if the letter presented was a member of the memory set. The subject then released the home button and pressed either the "yes" or "no" telegraph key located next to the flight yoke. Subjects were instructed to make their decisions and key presses as quickly as possible. Choice reaction time (CRT) was operationally defined as the time between onset of the letter flash and release of the home button. Movement time was defined as the time between release of the home button and pressing the "yes" or "no" telegraph key. Of interest in this study was CRT. Also, accuracy of the subject's responses was recorded. An accurate response was defined as either responding "yes" when a memory set letter was presented (a "hit") or "no" when it was not (a "correct rejection"). All other response types (including infrequent non-responses) were considered incorrect responses. The percentage of correct responses for each run of secondary task performance in the study was calculated. The fifth practice session consisted of practicing the secondary task alone for two runs, then practicing the primary and secondary tasks together for four runs in a dual-task situation. These dual-task runs consisted of combining both levels of primary task (easy and difficult course) and both levels of secondary task (memory set of two and four letters). The sixth and seventh sessions were each comprised of six data collection runs like those in session five. All seven sessions utilized counterbalancing of primary task courses and secondary task stimuli across the runs to minimize possible order effects. After the last session, the subject was debriefed and compensated.

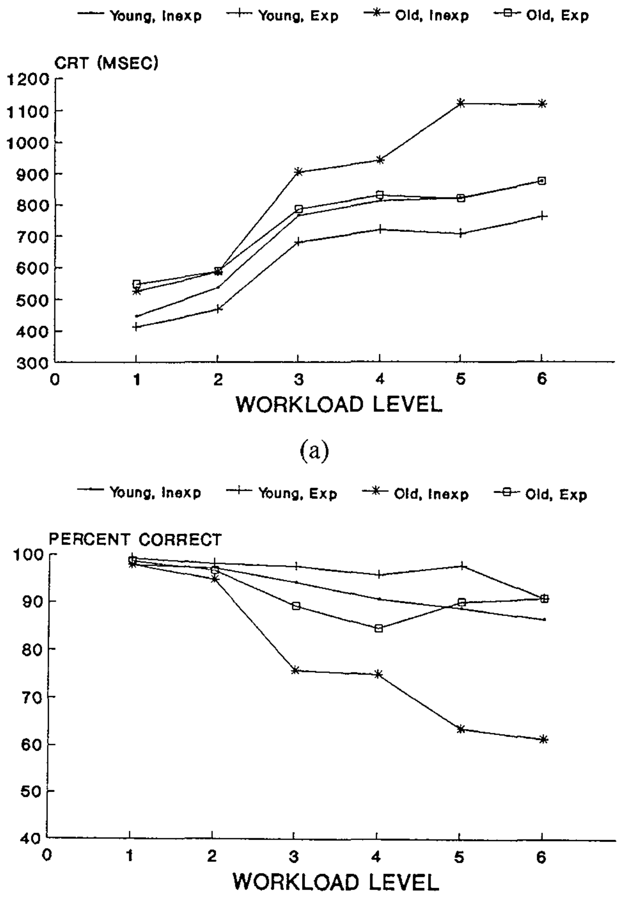
RESULTS

The data were analyzed by a mixed design ANOVA with two between subjects variables (Age and Expertise) and one repeated measures (within subjects) variable (Workload) with six levels. Age and Expertise were allowed to freely vary in the subject sample because a regression analysis had been originally planned. To perform the ANOVA, however, Age and Expertise were partitioned into two levels each using the median split technique to form four groups of subjects (young, less expertise; young, more expertise; old, less expertise; and old, more expertise). Initial analysis of the data from the entire sample revealed a virtual absence of significant findings involving expertise, perhaps because a median split was used to partition expertise into two levels - a weak manipulation. Therefore, another analysis was performed on a subset of the data consisting of those subjects with extreme amounts of expertise (i.e., least and most hours of flight time) within each age group. Within each age group, ten subjects were selected: the five with the most and the five with the least flight time. For this subset of subjects, the age range was 21 to 75 years, while the expertise range was 28 to 11,817 hours of flight time. For the groups, the mean ages and hours of flight time were: 1) young, less expertise: 27.2 years and 60.02 hours; 2) young, more expertise: 36.6 years and 5277.88 hours; 3) old, less expertise: 58 years and 217.6 hours; and 4) old, more expertise: 52.6 years and 7323.4 hours.

First, analyses were conducted to test if the secondary task intruded on primary task performance in the dual-task conditions (i.e., workload levels 3 - 6). These analyses were done because a requirement of the secondary task methodology is that the secondary task does not intrude on primary task performance. Separate ANOVAs were conducted comparing primary task performance measures in the dual-task conditions with those same measures collected during the last practice runs of performing the primary task alone. Primary task performance measures analyzed were deviations from criterion values (i.e., root-mean-square-errors, or RMSEs) for heading, altitude, airspeed, and distance off course. None of these analyses found a significant main effect for task condition (i.e., primary task alone vs. dual-task) on any of these measures, thus indicating that the addition of the secondary task did not intrude on these measures of primary task performance. Analyses were then performed to see if primary task performance varied across the different levels of workload in the dual-task conditions. Results indicated that as workload increased, distance off course (a general summary measure of primary task performance) and heading were unaffected, but altitude (F[3, 48] = 3.31, p < .03) and airspeed (F[3, 48) = 25.62, p < .001) deviations increased with workload.

Next, mixed design ANOVAs were conducted separately on CRT and percent correct responses on the secondary task for those conditions where subjects performed the secondary task alone during the last two sessions. Here, since the primary task was not performed in these conditions, Workload was equivalent to memory load of the secondary task (2 or 4 letters), and thus had two levels. For CRT, the results indicated significant effects for Age (F[1, 16] = 8.30, p < .011) and Workload (F[1, 16] = 21.12, p < .001). No other effects were observed. The presence of an age effect was expected, as the literature has repeatedly shown that age slows CRT. For percent correct responses, the results indicated only a significant effect for Workload (F[1, 16] = 10.35, p < .005). The absence of an expertise effect and interactions involving age and expertise in these analyses demonstrated that the Sternberg task was domain-general, an important finding indicating that the expert pilots in this study were not simply more capable people in terms of secondary task performance.

Next, mixed design ANOVAs with Workload as a six-level within-subjects factor were conducted separately on CRT and percent correct responses on the secondary task. For CRT, the following results were observed (see Figure 3a). Age (F[1, 16] = 11.53, p < .004), Workload (F[5, 80] = 99.14, p < .001), and Expertise (F[1, 16] = 6.84, p < .019) all had significant main effects. Also, there was a significant Expertise x Workload interaction (F[5, 80] = 4.90, p < .001), indicating that the CRT of pilots with more expertise was less affected by increases in workload. Regarding CRT, however, expertise did not significantly mitigate the effects of age on the ability to handle increases in workload, as indicated by the absence of a significant three-way interaction of Age, Expertise, and Workload (although it approached significance). For percent correct (see Figure 3b), Age (F[1, 16] = 13.32, p < .002), Expertise (F[1, 16] = 11.06, p < .004), and Workload (F[5, 80] = 16.39, p < .001) were all significant. Also, these interactions were significant: Age x Workload (F[5, 80] = 4.58, p < .001); Expertise by Workload (F[5, 80] = 5.22, p < .001); and a three-way interaction of Age x Expertise x Workload (F[5, 80]



(a)

(b)

Figure 3. Secondary task (a) choice reaction time performance and (b) percent correct performance of subjects as a function of workload level. See text for subject group and workload level definitions.

= 2.52, p < .036). This three-way interaction is discussed below. Also, repeated measures, mixed design ANOVAs with Workload as a four-level within-subjects factor (i.e., restricted to dual-task conditions) were conducted. For CRT, Age (F[1, 16] = 10.30, p < .005), Expertise (F[1, 16] = 9.09, p < .008), and Workload (F[3, 48] = 8.65, p < .001) were all significant. For percent correct, Age (F[1, 16] = 14.24, p < .002) and Expertise (F[1, 16] = 11.62, p < .004) were significant.

DISCUSSION

The goal of this study was to find evidence that expertise mitigates the affects of aging on pilot performance; specifically, a three-way interaction among age, expertise, and workload. Evidence of this interaction was provided by the analysis of subjects with extreme levels of flight expertise, with six levels of workload, for percent correct (accuracy) but not choice reaction time (although it approached significance; see Figures 3b and 3a, respectively). Figure 3b shows that: 1) the accuracy of younger subjects with more expertise was least affected by increases in workload; 2) the accuracy of older subjects with less expertise was most affected as workload increased; and 3) younger subjects with less expertise and older subjects with more expertise were moderately affected by the increase in workload. So it would seem that this subset of the data indicates that in going from a single-task situation performing a choice reaction time task to a dual-task situation performing a simulated aviation task and choice reaction time task, expertise (i.e., hours of flight time) mitigated the detrimental effects of age to some degree regarding cognitive resources involved in working memory and monitoring. This important finding coincides with those reported by Tsang and Shaner (1994) and Morrow et al. (1994). Although this three-way interaction was not found in the strictly dual-task workload manipulations, strengthening these manipulations in future studies may uncover it.

To improve the research, the following refinements will be incorporated: 1) an extreme groups design to better manipulate age and expertise; 2) stronger primary task workload manipulations; 3) more primary task practice for subjects to stabilize specific measures of primary task performance across levels of workload; 4) embedding the secondary task display into the primary task display; and 5) higher fidelity simulation to increase domain relevancy and support for domain-specific strategies used by expert pilots.

The results of this preliminary study provide a partial indication that expertise mitigates the adverse effects of aging regarding the ability to handle mental workload (regarding working memory and monitoring) in a simulation task. These findings will help "fine tune" the methodology and direct further research investigating additional cognitive resources involved in flying that may be affected by age, expertise, and workload. Ultimately this research may provide support for development of: 1) performance criteria for aging pilots to augment the age criteria currently in place in several branches of aviation; and 2) training regimens for older pilots to help them maintain their skills and certification.

ACKNOWLEDGEMENTS

This research was supported by National Institute of Aging Grant AG12388. We would like to thank John Demos for his help in data collection.

REFERENCES

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In K. R. Boff, L. Kaufmen, & J. P. Thomas (Eds.), Handbook of Perception and Human Performance (pp. 41-1, 41-44). New York: John Wiley & Sons.

Microsoft (1990). Flight Simulator 4.0. Redmond, WA: Microsoft Corporation.

Morrow, D. G., Leirer, V. O., & Altieri, P. A. (1992). Aging, expertise, and narrative processing. Psychology and Aging, 7, 376-388.

Morrow, D. G., Leirer, V. O., Altieri, P. A., & Fitzsimmons, C. (1994). When expertise reduces age differences in performance. Psychology and Aging, 9, 134-148.

Rybash, J. M., Hoyer, W. J., & Roodin, P. A. (1986). Adult cognition and aging. New York: Pergammon Press.

Sternberg, S. (1969). On the discovery of processing stages: Some extensions of Donders' method. Acta Psychologica, 30, 276-315.

Tsang, P. S. (1992). A reappraisal of aging and pilot performance. The International Journal of Aviation Psychology, 2(3), 193-212.

Tsang, P. S. (1995, April). Boundaries of cognitive performance as a function of age and piloting expertise. Paper given at the Eighth International Symposium on Aviation Psychology. Columbus, OH.

Tsang, P. S., & Shaner, T. L. (1994, March). Age and expertise in time-sharing performance. Paper given at the Biennial Cognitive Aging Conference. Atlanta, GA.

Wickens, C. D., Braune, R., & Stokes, A. (1987). Age differences in the speed and capacity of information processing: 1. A dual-task approach. Psychology and Aging, 2, 70-78.

Yekovich, F., Walker, C., Ogle, L. & Thompson, M. (1990). The influence of domain knowledge on inferencing in low aptitude individuals. In A. Graesser & G. Bower (Eds.), The psychology of learning and motivation: Advances in research and theory (Vol. 24, pp. 259-278). San Diego, CA: Academic Press.