Learning to Use a Home Medical Device: Mediating Age-Related Differences with Training

Amy L. Mykityshyn, Arthur D. Fisk, and **Wendy A. Rogers,** Georgia Institute of Technology, Atlanta, Georgia

We examined the differential benefits of instructional materials for younger and older adults learning to use a home medical device. Participants received training on use of a blood glucose meter via either a user manual (a text guide with pictures) or an instructional video. Performance was measured immediately and then after a 2-week retention interval. Type of instruction was critical for determining older adults' performance. Older adults trained using the manual had poorer performance than did all other groups. After only 1 calibration, older adults who received video training performed as accurately as did the younger adults. Older adults' performance was more influenced by the retention interval; however, the benefit of the video training was maintained for the older adults across the retention interval. Confidence ratings paralleled subjective workload ratings. The data provide practical information to guide the development of training programs for systems that will be used by both younger and older adults; they also demonstrate the need for age-related usability testing even for training program design.

INTRODUCTION

Electronic devices are becoming more prevalent in society. Interaction with such devices is nearly unavoidable; finding effective methods for training a wide range of potential users is thus very important. Although many devices are portrayed as being simple to operate, they may be quite difficult to use. For example, Rogers, Mykityshyn, Campbell, and Fisk (2001) reported that a blood glucose meter advertised as simple and easy to use required 52 steps to operate successfully.

The blood glucose meter is typical of home health care devices that require a relatively long series of sequential tasks to be performed, are sensitive to user procedural errors, and may lead to health risks if used incorrectly (e.g., blood pressure monitors, infusion pumps). These devices can be more difficult for older adults to use, not only because these users are less familiar and comfortable with such devices but also because they may have to remember sequences of steps to operate them properly.

Usability of home health care devices is critical given the increased emphasis being placed on home health care. More people are relying on home medical equipment to monitor their health or to provide at-home treatment. A study commissioned by the National Research Council concluded, "Good design is especially critical for home medical equipment because users may have infirmities, high stress, limited training, and insufficient sources of advice and support" (Klatzky, Kober, & Mavor, 1996, p. 3). These problems can be exacerbated for older adults, who may be taking medication or have chronic health problems in addition to cognitive and information-processing deficits associated with normal aging (Bogner, 1999). The field of human factors has the potential to play a major role in the improved, user-centered design of this class of equipment (Gardner-Bonneau & Gosbee, 1997). In many cases,

Address correspondence to Wendy A. Rogers, School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332-0170; wr43@prism.gatech.edu. **HUMAN FACTORS,** Vol. 44, No. 3, Fall 2002, pp. 354–364.

however, changes to medical devices take years to implement, test, and receive approval before they are available to the general public. The most immediate impact on the safe and effective use of these devices may therefore be in the design of training methods to support their proper use (e.g., attainment of required performance accuracy).

What is the optimum method for training individuals to use sequential devices? The format of the training materials can affect how well the user is able to learn a task. Manufacturers often provide a training video with their devices, although there have been no empirical studies to assess whether such training is useful for this class of tasks. Video training has been shown to be useful in sports contexts (e.g., Scott, Scott, & Howe, 1998), provided that the video is designed so that the user's perspective is presented (Walker & Fisk, 1995). Video may have some advantages over traditional paper-based instruction methods. It can be used either to provide feedback about a performance just completed or to provide a model of the correct or desired performance (Franks & Maile, 1991). Video may provide a model of the motor response expected of the user and may be especially effective for trainees who have difficulty reading.

One drawback to the use of video is that it has been described as a "passive" training tool, rather than an "active" one. Active participation is important for learning to occur (Schneider, 1985). The act of reading and interpreting text and pictures, because it is more effortful, may result in more active engagement with the material, thus resulting in improved learning and retention. However, Sweller and colleagues have indicated that if the need for integration is too high, learning will be negatively influenced (e.g., Chandler & Sweller, 1996).

Training effectiveness may also be dependent on the cognitive abilities of the trainees. For example, a large body of research performed over the past quarter century has shown that older and younger adults differ in various cognitive abilities (Craik & Salthouse, 2000). Age-related declines in attentional processes, working memory capabilities, discourse comprehension, inference formation, memory processes, and information-processing speed may impact the ability of older adults to acquire new skills (Czaja, 1996). As a result, training may be differentially effective for older and younger adults.

Several recent studies have found age-related differences in training effectiveness in learning to perform basic laboratory tasks (e.g., Fisk & Rogers, 1991) as well as for learning to use various everyday products (e.g., see Jamieson & Rogers, 2000; Mead, Batsakes, Fisk, & Mykityshyn, 1999; Mead & Fisk, 1998; Rogers, Fisk, Mead, Walker, & Cabrera, 1996). With respect to the specific role of the format of training materials, Morrell, Park, and Poon (1990) found that compared with younger adults, older adults had more difficulty with information presented in a format that combined pictures and words, relative to a verbal description alone.

Together, training studies conducted with older adults suggest that the older trainees are more influenced by the specific characteristics of the training system and training materials. Consequently, it is critical to include older adult samples in the development and testing of instructional materials for systems that they will use.

OVERVIEW OF STUDY

The transition from training principles to training programs is often not direct (e.g., Salas, Cannon-Bowers, & Blickensderfer, 1997). Recommendations from the literature cannot simply be implemented to design a training program for a particular task or situation. However, one can build from that knowledge base to develop alternative training programs and then empirically test the differential benefits of such programs. That was the approach taken in this study. Based on analysis of the task characteristics (e.g., the sequential nature of the task) and an understanding of the unique needs of the trainee population (i.e., older adults), we developed two training programs. We compared performance on the two programs for younger adults and older adults, measured time and accuracy using the device, and assessed performance immediately after training and after a 2-week retention interval. Retention testing is used as a means of determining the robustness of differences between training techniques (Schmidt & Bjork, 1992).

The two programs we tested consisted of training with a user's manual and training with an instructional video. The literature on training and aging yielded two competing hypotheses: First, the user manual might yield better performance because it requires more active integration on the part of the user (Schneider, 1985). However, the integration requirements might overload the working memory capacity of the older trainees, thereby reducing the efficacy of the manual (e.g., Morrell et al., 1990). Second, the instructional video might be more beneficial for learning because the information was integrated for the trainee (e.g., Chandler & Sweller, 1996); moreover, such integration might be relatively more beneficial for older adults because of age-related working memory declines.

METHOD

Participants

Participants in the study were 30 younger adults between the ages of 17 and 24 years (M = 19.6, SD = 1.9) and 30 older adults between the ages of 65 and 74 years (M = 68.5, SD = 2.8). Participation was limited to individuals who had never used a blood glucose monitor and did not live with anyone who uses such a device. The younger adults received monetary or course credit compensation. The older adults were paid volunteers from the community. Medication screening eliminated participants who were taking medications that have more than minimal effects on attention (see Batsakes, Hancock, Rogers, & Fisk, 2002). All participants passed a visual acuity test at the level equivalent to 20/40 for far and near vision (corrected or uncorrected).

Given that the training manipulation was between groups, we wanted to be certain that there were not any training group differences in demographic characteristics, such as age, education, and health, or in basic abilities, such as perceptual speed, working memory, and verbal skills. Table 1 presents these data along with the t test of differences between the two experimental groups for each age. There were no training group differences for either the younger or older adults. However, typical overall age-related differences in abilities were observed, as indicated in the note to Table 1.

Design

Training group (text-based user manual or instructional video) was a between-participant independent variable, and age group was a quasi-independent variable. Calibration number was a within-participant variable (Calibrations 1 and 2 were immediately after training; Calibrations 3 and 4 were after a 2-week retention interval). Dependent variables were training time, performance on the calibration task (calibration time and accuracy), performance on the written knowledge test, subjective ratings of workload (the National Aeronautics and Space Administration Task Load Index, or NASA TLX; Hart & Staveland, 1988), and confidence assessments.

Calibration Procedure

The participants' goal was to calibrate the blood glucose meter. Calibration required three separate categories of activities, referred to here as *tasks* (see Rogers et al., 2001, for a detailed discussion of the tasks required to use a blood glucose meter). The tasks for each calibration test were (a) set up the meter (requiring 8 separate steps), (b) perform a check strip test (10 separate steps), and (c) perform a glucose control solution test (12 separate steps).

Instructional Materials

Participants were required to calibrate a commonly available commercial home blood glucose meter. Two types of instructional materials were created for this experiment: a text-based user manual (with instruction-specific pictures) and an instructional video. Information provided by the manufacturer and from experienced users was combined to construct the training materials. The information contained in both training conditions covered the same critical steps necessary to successfully perform a blood glucose test, using the same vocabulary and presentation order. Both forms of instruction were created with consideration for instructional principles (for a review, see Wickens, Gordon, & Liu, 1998). That is, they were based on a detailed task analysis of the

TABLE 1: Demographic and Ability Data

	User Manual		Instructional Video		
	M	*SD*	*M*	*SD*	*t* value
Younger Adults					
Age	19.67	1.40	19.53	2.36	.19
Education (years)	13.20	1.15	13.13	1.46	.14
Health[a]	3.80	0.77	3.87	0.74	–.24
Alphabet span, absolute span[b]	45.00	15.52	47.13	12.71	–.41
Digit symbol substitution[c]	77.67	10.61	82.47	11.11	–1.21
Reading comprehension[d]	34.13	2.64	33.73	3.53	.35
Reverse digit span[e]	6.33	1.05	6.47	1.26	–.32
Vocabulary[f]	30.00	3.16	29.53	4.24	.34
Older Adults					
Age	67.60	2.03	69.33	3.18	–1.83
Education (years)	14.77	2.11	14.70	1.96	.09
Health[a]	3.67	0.90	3.87	0.92	–.60
Alphabet span, absolute span[b]	30.93	8.83	26.67	12.47	1.10
Digit symbol substitution[c]	54.40	13.44	53.60	11.04	.18
Reading comprehension[d]	30.53	6.52	28.00	6.88	1.04
Reverse digit span[e]	5.30	1.35	4.93	1.00	.85
Vocabulary[f]	35.07	3.24	35.20	2.43	–.13

Note. Overall age differences (significant at $p < .05$) were as follows: Older adults performed better on vocabulary and had more years of education; younger adults performed better on digit symbol substitution, reading comprehension, reverse digit span, and alphabet span. There were no differences in self-reported health.

[a]Self-rating: 1 = *poor*, 2 = *fair*, 3 = *good*, 4 = *very good*, 5 = *excellent*. [b]Working memory (Craik, 1986; LaPointe & Engle, 1990). [c]Perceptual speed (Wechsler, 1981). [d]Reading comprehension (Brown, Fishco, & Hanna, 1993). [e]Memory span (Wechsler, 1981). [f]Vocabulary (Shipley, 1986).

system (see Rogers et al., 2001); readability levels were appropriate for the intended user population; the vocabulary was simple and explicit; numbered steps were used to describe the procedures; advance organizers (Mayer, 1979) were provided to help the individual conceptualize the information; the steps were described in the same order in which they were to be performed; and redundancy of information was provided through multiple modes (text with figures or spoken words with visual images) and through the repetition of critical information. The primary distinction between the conditions was the format of the presentation: text with pictures or video.

User manual. The manufacturer's and experienced user information was combined into standard text-based instructional user information. Pictures from the manufacturer's manual augmented the printed instructions. The resultant user manual included a schematic drawing of the blood glucose meter and its major components, step-by-step instructions for performing three tasks required to calibrate the meter (isolating each step within a given task), and pictures that accompanied the text instructions depicting visually the steps required to perform each calibration task (the associated picture appeared next to the text that referred to it). The user manual was eight legal-size pages in length.

Instructional video. The transcript from the manufacturer's video was modified to form the basis for the instructional video. The video was filmed from the user's perspective and showed a close-up view of the glucose meter as someone performed each calibration task, step by step. Instructions for performing each calibration task began with a screen depicting the title of the task (i.e., "Performing a Glucose Control Solution Test"), and the descriptions of when to perform each type of calibration task (and associated steps within the task) was provided in bullet form on the screen. Thus the video provided a screen with the bulleted text of the steps before they were demonstrated; during the demonstrations the video was focused on the relevant component of the device as it was being used and the steps were described. The video was approximately 8 min in length.

Procedure

Training and immediate test. All participants were trained and tested individually on the three tasks required to calibrate the blood glucose meter. Participants were randomly assigned to either the user manual or instructional video training group. After receiving written instructions concerning general procedures, they were then given the training materials for their assigned condition and told to study the materials until they felt they could perform each of the calibration tasks. Participants in the user manual group could flip freely between the parts of the manual describing different aspects of the task. Participants in the video group could rewind the video and review it in full or in part (they were shown how to operate the videocassette recorder). When participants indicated they were through studying, the training materials were removed. Participants then rated their confidence that they could calibrate the meter (on a scale from 0% to 100% confident). They were then given the NASA TLX workload assessment scales and asked to rate the workload they experienced during training.

The immediate testing consisted of two complete calibrations of the meter (Calibrations 1 and 2). Participants rated their subjective workload after each task within the calibration. After Calibration 1 the participants rated how confident they were that they could calibrate the meter. They then attempted another complete calibration and gave workload assessments and a confidence rating. Following Calibration 2, participants completed a knowledge test of calibration tasks (including knowledge of steps within each task) and facts about the blood glucose meter. Participants did not have access to the instructional materials when answering the questions.

Retention session. Participants were retested approximately 14 days after training (range was 12–16 days, $M = 13.8$, with no significant differences in retention interval between training or age groups, $ps > .50$). During the retention session, and between testing sessions, participants did not have access to the training materials. The calibration testing procedure for this session was exactly the same as described earlier; participants completed two calibrations (Calibrations 3 and 4).

RESULTS

Training Time

Training time was measured from when each person began reviewing the training materials until they indicated that they were ready to begin calibrating the meter. Older adults spent almost 18 min with the video ($M = 1061$ s, $SD = 200.5$), and younger adults spent approximately 12 min ($M = 721.7$ s, $SD = 231.2$). Older and younger adults spent approximately 14 min with the manual ($M = 855.5$ s, $SD = 214.6$, and $M = 839.3$ s, $SD = 341.2$, respectively). An Age × Training Group analysis of variance (ANOVA) revealed a main effect of age; older adults spent more time studying the materials than did younger adults, $F(1, 56) = 7.40$, $p < .01$. The main effect for training group was nonsignificant ($p = .50$), but there was an Age × Training Group interaction, $F(1, 56) = 6.12$, $p < .02$. Follow-up analysis revealed that for the older adults, the video group spent more time studying than did the manual group, $F(1, 28) = 7.35$, $p < .01$; there was no study time difference between the younger adult training groups ($p = .28$). Training time did not correlate significantly with calibration accuracy or calibration time for either age group.

Performance after Training

Performance success in calibrating the meter was measured in terms of calibration accuracy and calibration time. Calibration accuracy was indexed by the percentage of total steps completed (i.e., how much of each calibration task was performed). Participants received credit for a task even if it was performed after a skipped step. Calibration time was measured from when participants turned on the meter until they indicated that they had completed the task or that they could do no more; this measure reflects the total time for all participants, not only those who calibrated the meter completely correctly. Composite variables for accuracy and time were created by summing across the three calibration tasks.

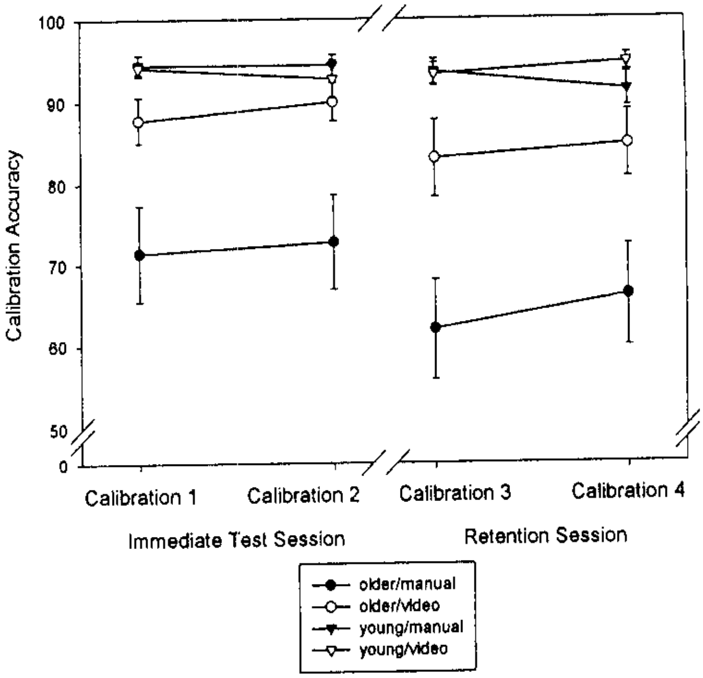

Figure 1. Calibration accuracy for the training and retention sessions for each age and training group.

Calibration accuracy. The calibration accuracy rates are presented in Figure 1. To assess the initial benefits of training, we conducted an Age × Training Group ANOVA for Calibration 1 (i.e., calibration immediately following training). Younger adults were more accurate than older adults, $F(1, 56) = 19.08$, $p < .01$, and the video groups were more accurate than the manual groups, $F(1, 56) = 5.71$, $p < .02$, but there was a significant Age × Training Group interaction, $F(1, 56) = 6.11$, $p < .02$. Follow-up analysis revealed that the training group difference was significant for the older adults, $F(1, 28) = 6.22$, $p < .02$, but not for the younger adults ($p = .85$). Thus video training was differentially beneficial for older adults.

The pattern of performance was virtually identical the second time participants calibrated the meter following training (Calibration 2). There were significant effects of age, $F(1, 56) = 12.73$, $p < .01$, training group, $F(1, 56) = 5.16$, $p < .03$, and their interaction, $F(1, 56) = 7.60$, $p < .01$. The benefit of video training for the older adults was thus maintained for the second attempt at calibrating the meter. In fact, the performance of the older adults in the video training group was not significantly different from that of the younger adults in either the video training group ($p = .40$) or in the manual training group ($p = .10$).

Calibration time. Figure 2 depicts the calibration times; for the initial calibration, older adults were substantially slower than younger adults, as indexed by the significant effect of age, $F(1, 56) = 28.03$, $p < .01$. There was no main effect of training group ($p = .20$), nor was there an interaction with age ($p = .77$). Calibrations times improved from Calibration 1 to Calibration 2. An Age × Training Group × Calibration (1–2) repeated-measures ANOVA revealed a significant effect of calibration, $F(1, 56) = 41.18$, $p < .01$, and a significant Age × Calibration interaction whereby older adults improved more from Calibration 1 to Calibration 2 than did younger adults, $F(1, 56) = 18.03$, $p < .01$. For Calibration 2, younger adults again completed the calibrations more quickly than did older adults, $F(1, 56) = 14.72$, $p < .01$; however, there was not a training group difference ($p = .16$), nor was there an Age × Training Group interaction for Calibration 2 ($p = .17$).

Performance after 2-Week Retention Interval

Participants returned after a 2-week retention interval. They were not provided with any additional training but were asked to calibrate the meter. The degree to which their learning was retained was assessed by comparing calibration

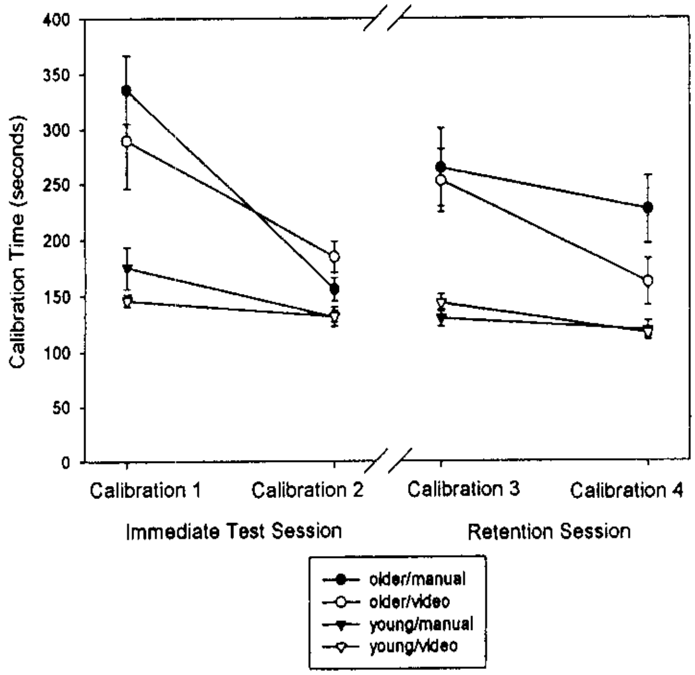

Figure 2. Calibration time (in seconds) for the training and retention sessions for each age and training group.

accuracy for Calibration 2 (before retention) and Calibration 3 (after retention; see Figure 1), with an Age × Training Condition × Calibration ANOVA. Calibration accuracy was lower after the retention interval, $F(1, 56) = 9.32$, $p < .01$; however, this effect interacted with age such that older adults showed more decline in performance, $F(1, 56) = 8.76$, $p < .01$. The rate of change across retention did not interact with training group ($p = .34$), nor was there a significant Age × Training Group × Calibration interaction ($p = .67$). In sum, accuracy of performance showed more decline for older adults than for younger adults but not differentially across the training groups.

The pattern of change in calibration time over the retention interval was similar to the pattern for accuracy (see Figure 2). That is, times for Calibration 3 were significantly slower than for Calibration 2, $F(1, 56) = 16.21$, $p < .01$, and this effect interacted with age such that older adults' times were slowed by a larger magnitude than were those of younger adults, $F(1, 56) = 12.73$, $p < .01$. However, the training groups were not differentially slowed; the Training Group × Calibration ($p = .56$) and Age × Training Group × Calibration ($p = .25$) interactions were not significant.

Older adults were more influenced by the retention interval than were younger adults in terms of both performance accuracy and time needed to complete the calibrations. There was little evidence that the training groups showed differential losses across the retention interval. Of importance is that the benefit for the video training group of older adults was maintained, as is evident by an analysis of performance during the final calibration (Calibration 4). After what can be considered one "refresher" calibration (Calibration 3), analysis of performance at Calibration 4 showed a marginal Age × Training Group interaction for calibration accuracy (see Figure 1), $F(1, 56) = 3.79$, $p < .057$. The follow-up analysis showed that the training group difference was significant for the older adults ($p < .02$) but not for the younger adults ($p = .15$). As for calibration time, the older adults in the video training group were faster than those in the manual training group, although the Age × Training Group interaction was not statistically significant ($p = .10$), perhaps because of the high variability in performance (see Figure 2).

Task Knowledge

Participants completed a knowledge test at the end of the training session and again at the end of the retention session (the maximum score was 14). At the end of the training session, test scores were 75%, 81%, 91%, and

91% for the older adult manual group (M = 10.5, SD = 2.0), older adult video group (M = 11.3, SD = 1.4), younger adult manual group (M = 12.7, SD = 1.0), and younger adult video group (M = 12.7, SD = 1.0), respectively. Younger adults were better able to answer the knowledge questions than were older adults, $F(1, 56) = 25.56$, $p < .01$, but there was no training group difference ($p = .28$) or Age × Training Group interaction ($p = .28$).

At the end of the retention session, the test scores were 69%, 80%, 88%, and 86% for the older adult manual group (M = 9.7, SD = 1.4), older adult video group (M = 11.2, SD = 1.5), younger adult manual group (M = 12.3, SD = 1.2), and younger adult video group (M = 12.0, SD = 1.4), respectively. Thus age-related performance differences on the knowledge test remained, $F(1, 56) = 21.94$, $p < .01$. There was no significant main effect of training group ($p = .12$), but there was a significant Age × Training Group interaction, $F(1, 56) = 6.15$, $p < .02$. Follow-up analysis of the interaction showed a significant advantage for the video over the manual group only for the older adults, $F(1, 28) = 7.36$, $p < .01$; the younger adult training groups did not differ ($p = .50$).

Confidence Ratings

Prior to each calibration, participants rated their confidence to do the calibrations (see Table 2). An Age × Training Group × Calibration (1–4) ANOVA yielded main effects of calibration ($p < .01$), age ($p < .01$), and training group ($p < .02$). Of more interest, however, were the interactions. The Age × Calibration interaction was significant, $F(3, 168) = 3.24$, $p < .024$. The age groups did not differ significantly in their confidence immediately following the retention interval (Calibration 3), but at all other occasions the younger adults had higher confidence. The Age × Training Group interaction was significant, $F(1, 56) = 6.49$, $p < .01$, because although older adults were generally less confident than younger adults that they could perform the task, the age difference was less pronounced for older adults who received training via the video. Older adults who received training with the user manual consistently had the lowest confidence ratings of any group. Post hoc comparisons showed that for each confidence assessment, older adults in the video group were more confident in their abilities than were the older adults in the manual group (all $ps < .06$).

Workload Assessments

Participants provided workload assessments on the six subscales of the NASA TLX (mental demand, physical demand, temporal demand, performance, effort, and frustration level) after every calibration task. Table 2 presents the composite scores averaged across the subscales and across the tasks for each complete calibration. An Age × Training Group × Calibration (1–4) ANOVA revealed that older adults consistently rated their workload higher than did younger adults, $F(1, 56) = 28.24$, $p < .01$. There was also a main effect of training group, $F(1, 56) = 12.85$, $p < .01$, but this effect interacted with age, $F(1, 56) = 10.53$, $p < .01$. Participants in the manual training groups rated their workload higher than did those in the video training group, and these differences were larger for older adults. The calibration effect was significant, $F(3, 168) = 17.72$, $p < .01$, because workload ratings were higher immediately after training and immediately after retention. However, this effect did not interact with age or with training group ($ps > .74$).

Analyses of the subscales revealed significant training group differences for all of them: mental demand, physical demand, temporal demand, performance, effort, and frustration; in all cases workload was higher for the manual training group (all $ps < .01$). The Age × Training Group interactions were also significant for mental demand, performance, effort, and frustration (all $ps < .01$); older adults in the manual training group reported the highest levels of workload.

DISCUSSION

The primary purpose of this study was to directly compare two media (written vs. video instructions) for training younger and older adults to use a relatively complex sequential system. To summarize the main findings, the type of training did not influence the performance of younger adults, either immediately after training or after a 2-week retention interval. Younger adults were able to learn

TABLE 2: Confidence Ratings and Workload Assessments

	Immediate Test Session				Retention Session			
	Calibration 1		Calibration 2		Calibration 3		Calibration 4	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
	Confidence Ratings[a]							
Younger adults								
User manual	88.9	6.7	95.3	6.8	76.4	14.8	89.3	18.3
Instructional video	86.8	14.0	95.5	5.8	75.5	21.2	89.7	17.3
Older adults								
User manual	62.2	19.0	59.3	27.4	57.8	25.3	53.7	35.9
Instructional video	79.1	16.4	82.3	25.3	74.0	19.6	77.0	23.4
	Workload Assessments[b]							
Younger adults								
User manual	14.1	8.7	7.6	6.8	14.4	8.2	8.2	8.4
Instructional video	12.6	9.2	8.8	8.0	12.3	9.9	7.2	7.0
Older adults								
User manual	37.7	13.5	25.8	14.9	33.8	15.5	30.2	16.8
Instructional video	18.3	11.1	10.6	7.3	20.3	15.3	11.9	9.9

[a]Confidence was rated on a 0–100 scale. Ratings were made immediately prior to each calibration. [b]Workload is presented as the mean of the NASA TLX subscale ratings. Higher numbers represent higher perceptions of workload. Ratings were made immediately prior to each calibration.

how to perform the calibration tasks on the blood glucose meter with both types of training. For the older adults, the type of training provided was an important variable. The older adults who received the video instructions had more system knowledge and were faster, more accurate, more confident, and under less perceived workload than were the older adults who received the written instructions. The training group differences were maintained after the 2-week retention interval, although both groups of older adults did show declines in performance attributable to nonuse during the interval. This pattern of differential effectiveness of training programs for older adults is consistent with previous findings (e.g., Jamieson & Rogers, 2000; Mead & Fisk, 1998).

These results have implications for the design of training programs as well as for understanding the maintenance of new skills by older individuals. With respect to training program design, we followed established instructional principles for designing both the written and the video instructions (e.g., see Wickens et al., 1998). As a result, participants were able to learn how to use the system reasonably well, as compared with following instructions provided by system manufacturers (see Rogers et al., 2001). However, performance was not perfect for any of the groups, younger or older (all accuracy rates were significantly lower than 100%, $p < .01$). Following instructional design recommendations is only the starting point. Designers must recognize that the development of training materials should be an iterative process whereby materials are tested with typical users of the system, benefits of training are assessed, and the materials are refined as needed (Salas et al., 1997).

The present evidence suggests that all else being equal, video training results in better learning by older adults, relative to a written manual with pictures. The benefits were evident both initially as well as following the retention interval. Why should video training be superior for older adults? The information content did not differ across the two training conditions; the primary difference was that with the video training, the trainee could see the actions being performed, whereas in the manual training group the trainee had only still pictures of the different states of the monitor as the calibration was performed. One could argue that the manual training required more active elaboration on the part of the trainee and, hence, should have yielded superior performance or performance at least equal to that observed after video training. This might explain the absence of differences between the two younger adult training groups. That is, the younger adults in the manual training condition may have been able to visualize or imagine the sequence of actions as they would occur (i.e., make inferences from the pictures about the necessary actions).

For the older adults, the fact that training with the user manual required visualization, imaging, inferring, and integration may account for the poorer performance of this group. Age-related differences in working memory and reading comprehension may also contribute to poorer performance following written instructions. Video training may provide environmental support for the learner by explicitly providing the task sequence, thereby minimizing reliance on both working memory (for visualizing) and reading comprehension (for drawing necessary inferences). The benefit of the video training is also consistent with the general idea that if complex information must be integrated, learning will benefit if the training materials facilitate integration for the learner (e.g., Chandler & Sweller, 1996). Older adults who received video training did spend more time with the training materials, as compared with the older adults who received user manual training. However, training time was not correlated with either accuracy or calibration time; thus training time differences cannot account for the performance advantage observed for the video condition. The degree to which video benefits are evident for other classes of learning tasks remains an empirical question, but it is certainly one worth pursuing.

It is important to note that not just any video will result in superior performance. Rogers at al. (2001) reported that the video provided by the manufacturer of the blood glucose meter was not sufficient for training users to operate the system. Younger and older adults watched the manufacturer-provided video for the specific calibration task, performed the calibration, and then went on to the next calibration task; overall calibration accuracy was 75% correct for younger adults and only 25% correct for older adults. To put it in perspective, contrast performance after viewing the manufacturer's video with even the "poor" performance of the older adults in the present text-based training condition, which averaged above 70%. The manufacturer's video did not show the system from the user's perspective, did not provide advance organizers, did not reiterate key points, did not allow dual coding of information (text and auditory display of information), and contained irrelevant information, such as advertisements. For a video to be effective, it must be designed with consideration of these instructional principles in mind, as was the video in the present study.

Testing participants after a 2-week interval during which the system was not used assessed skill maintenance. Performance declined for the older adults. Future studies will be necessary to determine the specific point at which declines are evident. From a practical perspective, individuals may be given instructions to use a home health care device and not need to use it for several days – it is conceivable that in the interim they will have forgotten critical aspects of the task. Evaluations of instructional materials should include retention intervals.

From a theoretical perspective, the apparent performance equivalence of the younger and older adult video training groups after training and two calibration experiences did not translate to equivalent retention of that information. This suggests that immediate performance may not provide a good index of learning (Schmidt & Bjork, 1992) and that skill learning may not be retained as well by older adults as by younger adults (Fisk, Hertzog, Lee, Rogers, & Anderson-Garlach, 1994). Understanding age-related differences in skill acquisition and retention contributes to the knowledge base about the effects of aging and provides practical information to guide the development of training programs for systems that will be used by both younger and older adults.

ACKNOWLEDGMENTS

This research was supported in part by the following grants from the National Institutes of

Health (National Institute on Aging): Grant P50 AG11715 (Center for Aging and Cognition: Health, Education, and Technology), Grant P01 AG17211 (Center for Research and Education on Aging and Technology Enhancement), and Grant R01 AG07654. The article is based on a thesis submitted by the first author as partial fulfillment of the master of science degree. The authors thank Chris Mayhorn for his constructive comments on the paper.

REFERENCES

Batsakes, P. B., Hancock, H. E., Rogers, W. A., & Fisk, A. D. (2002). A medications screening tool for cognitive aging researchers. *Psychology and Aging, 17*, 169–173.

Bogner, M. S. (1999). How do I work this thing? Cognitive issues in home medical equipment use and maintenance. In D. C. Park, R. W. Morrell, & K. Shifren (Eds.), *Processing of medical information in aging patients: Cognitive and human factors perspectives* (pp. 223–232). Mahwah, NJ: Erlbaum.

Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denney Reading Test.* Chicago: Riverside.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*, 151–70.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive abilities* (pp. 409–422). Amsterdam: Elsevier.

Craik, F. I. M., & Salthouse, T. A. (2000). *The handbook of aging and cognition* (2nd ed.). Mahwah, NJ: Erlbaum.

Czaja, S. J. (1996). Aging and the acquisition of computer skills. In W. A. Rogers, A. D. Fisk, & N. Walker (Eds.), *Aging and skilled performance: Advances in theory and applications* (pp. 201–220). Mahwah, NJ: Erlbaum.

Fisk, A. D., Hertzog, C., Lee, M. D., Rogers, W. A., & Anderson-Garlach, M. (1994). Long-term retention of skilled visual search: Do younger adults retain more than old adults? *Psychology and Aging, 9*, 206–215.

Fisk, A. D., & Rogers, W. A. (1991). Toward an understanding of age-related memory and visual search effects. *Journal of Experimental Psychology: General, 120*, 131–149.

Franks, I. M., & Maile, L. J. (1991). The use of video in sport skill acquisition. In P. W. Dowrick (Ed.), *Practical guide to using video in the behavioral sciences* (pp. 231–243). New York: Wiley.

Gardner-Bonneau, D., & Gosbee, J. (1997). Health care and rehabilitation. In A. D. Fisk & W. A. Rogers (Eds.), *Handbook of human factors and the older adult* (pp. 231–255). San Diego, CA: Academic.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: Elsevier.

Jamieson, B. A., & Rogers, W. A. (2000). Age-related effects of blocked and random practice schedules on learning a new technology. *Journals of Gerontology: Psychological Sciences, 55B*, P343–P353.

Klatzky, R. L., Kober, N., & Mavor, A. (1996). *Safe, comfortable, attractive, and easy to use: Improving the usability of home medical devices.* Washington, DC: National Academy.

LaPointe, L. B., & Engle, R. (1990). Simple and complex word spans as a measure of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1118–1133.

Mayer, R. E. (1979). Twenty years of research on advance organizers: Assimilation theory is still the best predictor of results. *Instructional Science, 8*, 133–167.

Mead, S. E., Batsakes, P. B., Fisk, A. D., & Mykityshyn, A. L. (1999). Application of cognitive theory to training and design solutions for age-related computer use. *International Journal of Behavioral Development, 23*, 553–573.

Mead, S. E., & Fisk, A. D. (1998). Measuring skill acquisition and retention with an ATM simulator: The need for age-specific training. *Human Factors, 40*, 516–523.

Morrell, R. W., Park, D. C., & Poon, L. W. (1990). Effects of labeling techniques on memory and comprehension of prescription information in young and old adults. *Journals of Gerontology, 45*, 166–172.

Rogers, W. A., Fisk, A. D., Mead, S. E., Walker, N., & Cabrera, E. F. (1996). Training older adults to use automatic teller machines. *Human Factors, 38*, 425–433.

Rogers, W. A., Mykityshyn, A. L., Campbell, R. H., & Fisk, A. D. (2001). Only three easy steps? User-centered analysis of a "simple" medical device. *Ergonomics in Design, 9*(1), 6–14.

Salas, E., Cannon-Bowers, J., & Blickensderfer, E. L. (1997). Enhancing reciprocity between training theory and practice: Principles, guidelines, and specifications. In J. K. Ford, S. W. J. Kozlowski, K. Kraiger, E. Salas, & M. S. Teachout (Eds.), *Improving training effectiveness in work organizations* (pp. 291–322). Mahwah, NJ: Erlbaum.

Scott, D., Scott, L. M., & Howe, B.L. (1998). Training anticipation for intermediate tennis players. *Behavior Modification, 22*, 243–261.

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

Schneider, W. (1985). Training high-performance skills: Fallacies and guidelines. *Human Factors, 27*, 285–300.

Shipley, W. C. (1986). *Shipley Institute of Living Scale.* Los Angeles: Western Psychological Services.

Walker, N., & Fisk, A. D. (1995). Human factors goes to the gridiron. *Ergonomics in Design, 3*(3), 8–13.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale – Revised.* New York: Psychological Corp.

Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). *An introduction to human factors engineering.* New York: Longman.

Amy L. Mykityshyn is a research engineer at the Georgia Tech Research Institute. She received her M.S. in engineering psychology from the Georgia Institute of Technology in 2000.

Arthur D. Fisk is a professor in the School of Psychology at the Georgia Institute of Technology. He received his Ph.D. in psychology from the University of Illinois in 1982.

Wendy A. Rogers is a professor in the School of Psychology at the Georgia Institute of Technology, where she received her Ph.D. in experimental psychology in 1991.

Date received: May 3, 2001
Date accepted: February 7, 2002