ABSTRACT

CARPENTER, ERIC DWIGHT. Effect of Transfer with Younger and Older Adults on Control Solution Testing using Two Blood Glucometers. (Under the direction of Christopher B. Mayhorn.)

As of 2005, 5.5% of the American population had some form of diabetes, with an increasing rate of diagnosis. Adults 65 years of age and over are at an increased risk for developing diabetes. Sixty-four participants completed 10 trials of control solution testing. Trials were videotaped and graded using a behavioral checklist constructed based on hierarchical and cognitive task analyses of glucometer stimuli. MANOVA results indicated main effects of participant age and order of glucometer use on dependent variables. Follow-up ANOVAs revealed a main effect of age on task time, errors during training, and near transfer errors, and a main effect of glucometer use on rate of near and far transfer errors committed. Results and avenues of further investigation are discussed.

Effect of Transfer with Younger and Older Adults on Control Solution Testing using Two Blood Glucometers

by
Eric Dwight Carpenter

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Psychology

Raleigh, North Carolina

2011

APPROVED BY:

Christopher Mayhorn, Ph.D., Chair

Anne McLaughlin, Ph.D., Member

Denis Gray, Ph.D., Member

BIOGRAPHY

Mr. Carpenter graduated from North Carolina State University as an undergraduate student in Spring 2006 and began his graduate career in Human Factors and Ergonomics Psychology in Fall 2006. While his motivation for entering human factors (HF) was improving training and equipment in the medical field he has applied his HF training to several domains, and remains interested in multidisciplinary projects to this day.

ACKNOWLEDGEMENTS

The author would like to thank Dr. Mayhorn for his guidance and patience during the development and implementation of this research. Thanks are due to the author's committee for raising the scientific rigor and quality of this work. Sigma Xi and the Psychology Department are thanked for their generous financial contributions that made this research possible. Finally, the author thanks his wife for her endless patience and support during this process, as well as her willingness to listen to numerous monologues concerning statistical assumptions and cognitive testing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

The Centers for Disease Control and Prevention (CDC) stated that 5.5% of the American population suffered from diabetes as of 2005 (Centers for Disease Control and Prevention, 2007). The elderly at an increased risk for being diagnosed with diabetes, with those 65 and older comprising 38% of those with diabetes, and there are no indications that the rate of diagnosis will slow anytime soon (Centers for Disease Control and Prevention, 2007). With this being the case, psychologists can either attempt to persuade the public to change their lifestyle to decrease their chances of contracting this condition, or they can aim to improve the quality of life for those diagnosed. These grim facts should grab the attention of all readers, as we are all aging (and so are our loved ones). At the same time, to provide more effective and cheaper care, there is an increasing push toward self-care with devices such as blood glucometers (Bogner, 2004; Kaye & Chenault, 2002; Scialfa, Ho, & Laberge, 2004).

The Food and Drug Administration's Center for Devices and Radiological Health (CDRH) states that technique errors are the most common type of error in diabetes care (Kaye & Chenault, 2002). The CDRH recommends that human factors (HF) principles be involved in the initial steps of device design, and that the special needs of the intended user be kept in mind at all times. Also stated by this department is the belief that as medical technology becomes more reliable, more errors will be caused by user actions, increasing the importance of HF research (Kaye & Chenault, 2002).

The purpose of the proposed research is to help clarify the usability needs of individuals with diabetes regarding medical equipment and training. Numerous studies have been conducted showing the decay of training for blood glucometers, and the errors that result from their inappropriate use (Bergenstal, Pearson, Cembrowski, Bina, Davidson, & List, 2000; Jones 1994; Raine 2002). To this point, no material has been found on the rate of occurrence or effects of near and far transfer errors in blood glucometers. Near transfer errors, or errors committed with familiar functions or technologies (Jamieson, Cabrera, Mead, & Rousseau, 1995) will continue to occur as professional and national guidelines are established. Far transfer errors, or errors committed with new unfamiliar functions or technologies (Jamieson, Cabrera, Mead, & Rousseau, 1995), will continue to occur as glucometers evolve. Human factors research provides an opportunity to reduce the risk of committing these errors and their subsequent influences on quality of life and mortality. More specifically, differences in the occurrence of near and far transfer errors between age groups will be the focus of the proposed research.

The contents of this paper begin by providing background information on aging, training, and medicine. More in-depth information concerning HF design considerations for the elderly and methodological concerns, and older adults and training will be provided next. Finally, a synthesis of the training, aging, and medical literature will take the form of an experiment that follows a rigorous procedural methodology to test specific hypotheses via a plan of statistical analysis that should inform the literature and provide guidance for future research efforts.

Aging Research

Until recently, it was commonly accepted that individuals would begin to constantly lose mental function a short time after achieving adulthood. However, in many cases, researchers treated mental ability as a single process, generalizing that a deficit in one area equaled an “across the board” deficit in all areas (Howard & Howard, 1997). While age-related change is still accepted, it is no longer considered a ubiquitous phenomenon. One of the most eloquent descriptions of aging comes from the work of Sterns and Alexander (1977). Entropy, the natural process of increasing chaos, forces a system to break down. Organisms prevent increasing entropy through the biological process of homeostasis. The interaction of entropy and homeostasis will be unique for each organism, but the overall progression of chaos and disorder allows researchers to understand the general process of aging.

Previous research documents areas of decline, stability, and improvement in the abilities of older adults. For instance, semantic memory is typically referred to as memory involving general or world knowledge (DeFreitas, Dunaway, & Torres, 2009). Semantic memory is the type of memory responsible for learning, manipulating, and remembering language and concepts without tying them to specific occurrences, and it frequently shows no decay later in life. It has even been shown that semantic memory can improve as a function of time (Howard & Howard, 1997). Crystallized knowledge, skills and information gained throughout a person’s entire life, stays intact except for the oldest old or in certain medical conditions (McLaughlin, Rogers, & Fisk, 2004). By contrast, decreases in episodic memory, the type of memory associated with remembering personal experiences, have been linked to increasing age. Working memory is a type of memory involved in keeping long term memories available for processing or preventing the decay of information that is currently being processed (Salthouse & Babcock, 1991). Age-related declines in working memory have been well documented. The decline in working memory results in slower processing of information for older individuals. For designers, this means avoiding task requirements that force older adults to keep multiple pieces of information in working memory at the same time (Howard & Howard, 1997; Mayhorn, Stronge, McLaughlin, & Rogers, 2004; McLaughlin, Rogers, & Fisk, 2004). Older adults would also benefit from more complete instructions that do not require manipulating information (Howard & Howard, 1997; McLaughlin, Rogers, & Fisk, 2004). For example, older adults should benefit more from video instructions where all procedural steps are acted out rather than from static step-by-step photographs requiring the user to visualize the actions to be performed (McLaughlin, Rogers, & Fisk, 2004; Mykityshyn, Fisk, & Rogers, 2002). The learning of new skills in older adulthood has been found to be more difficult as well (Howard & Howard, 1997). Learning new motor tasks is especially more difficult if the task is unfamiliar (Howard & Howard, 1997). If individuals are diagnosed with diabetes later in life, or if older diabetics encounter equipment unlike anything from their previous experience, designers will need to develop more effective training methods to ensure the safe use of their equipment.

Accompanying these changes in cognition, several age-related changes in perceptual and motor ability have been noted in the literature as well. Older adults experience more difficulty seeing in suboptimal light, at near distances, and when reading fine print (Scialfa, Ho, & Laberge, 2004). In audition, the ability to hear higher frequency sounds is compromised, especially when using pure tones instead of natural speech. Changes in motor abilities include restricted range of motion, loss of sensation in the fingers, and increased difficulty in coordinated movements. Age-related changes are often significant enough to affect the performance of older adults when using different technologies.

Older Adults and Interfaces

Ellis, Joo, and Gross (1991) studied the effects of age, gender, and interface during the use of a health risk appraisal program. Participants using a mouse for input took longer to perform the task, and more participants who used a mouse instead of a keyboard declined to wait for a printout of their final results. Interface style did not influence user intent to change lifestyle habits or the subjective helpfulness of the system, adding evidence to the idea that interface style alone and not pre-existing bias was responsible for performance declines. The authors state that instead of defining a specific input device as preferable for a certain task, researchers should examine the individual task, as well as the hand-eye coordination and dexterity of the user.

A more recent investigation into input device selection (McLaughlin, Rogers, & Fisk, 2009) agrees that rather than classifying input devices as preferable for a single type of task, designers should instead choose an input device based on how well a device matches the inputs required by the interface. A dual task methodology was used with younger and older adult participants. Participants completed tasks with input devices either matching or mismatching the input requirements of the tasks. Researchers found significant differences in participant performance using age and level of attention as variables. Older participants assigned to a reduced attention condition caught fewer dots in an experimental game. When assessing response time, a main effect of age was found along with an interaction between age and level of attention and an interaction between age, level of attention, and the level of match between input device and input requirements. Older adults were more strongly affected by multitasking and mismatches between tasks and devices than younger adults.

Czaja, Hammond, Blascovich, and Swede (1989) examined performance and task time while novices learned how to use a text editor. The experiment used three age groups (young, middle, and older) and three instructional methods (instructor, computer, and manual-based) to examine their influence on task time and error rate. No significant differences in typing ability were found between age groups, but as age increased the number of errors performed and the total task time increased. Middle aged (40-54 years of age) and older adults (aged 55-70 years) performed significantly more format errors, errors requiring participants to manipulate the look of a printed page. It was suggested that declines in spatial memory and visualization skills may be the cause. A non-significant increase in the rate of format errors was experienced by the participants who received computer based training. The authors found no significant differences between age groups in terms of attitudes toward computers, and training had no effect on attitudes toward computers during a post test. Due to the increased error rate of computer trained participants, the authors suggest a more active method of instruction. Researchers noted the difficulty older participants had in predicting what functions were contained in which menus, leading them to recommend future research in finding menu titles more intuitive for older and novice users.

One answer to Czaja et al.'s (1989) call for research on menu design indicates labels given to menu items have more impact on user performance than the overall organizational scheme of the menu itself (Resnick & Sanchez, 2004). Experimentally controlled websites were constructed using either a product-centered or task-centered design. Researchers also differentiated the quality of labels used into high, medium, and low conditions. Participants were recruited to perform a card sorting task to determine where products were placed inside each organizational scheme. Participants then used a Likert scale to determine the quality of the labels used in each scheme. After constructing 6 different websites, new participants completed an experimental shopping task and completed a satisfaction questionnaire. The product-centered organizational scheme was found to produce fewer errors and require fewer clicks to navigate correctly than the task-centered scheme. Satisfaction scores were significantly higher for the product-based scheme as well. While organizational scheme had many significant findings associated with it, researchers noted that significant differences usually occurred only when lower quality labels were presented. Label quality was associated with significantly faster response time, fewer errors, user satisfaction, number of required clicks, number of products found, and an interaction with organizational scheme in that higher quality labels resulted in preferable participant performance. The results from this experiment suggest that labels produce a greater effect on performance than overall structure. The fact that participants determined the label quality categories also indicates that a user-centered method to select labels results in better labels than simple benchmarking or using labels more familiar to developers.

Research into the usability of personal digital assistants (PDAs) (Mayhorn, Lanzolla, Wogalter, & Watson, 2005) resulted in findings similar to previous research on performance using personal computers (Czaja et al., 1989). Younger and older adults entered medication schedules into a PDA during 3 separate trial periods. Participants were timed and errors were recorded, with a distinction being made between cognitive and motor control errors. An illustrated, step by step instruction manual was created for participants, and was available for reference during each trial. Older adults performed more slowly and committed more cognitive and motor control errors than younger adults. In addition to these findings, participants were asked to give feedback about positive and negative qualities of the PDA used in the study. Older and younger participants produced a fairly uniform critique of the PDA. However older adults made more suggestions about how to improve the usability of the PDA, demonstrating that adding older adults to testing samples adds information that could not be obtained otherwise. The data from this study suggests that age-related trends in performance with personal computers can be expected when using portable computers. Age-related changes in older adults have implications for interface design, as well as any training provided to teach the new technologies these interfaces are designed for.

Training Research

Blaiwes, Puig, and Regan discussed the state of transfer research in a 1973 review. Their article calls for evaluation through training, the idea that the effectiveness of a training program can be measured by the effect it has on trainee performance. If a training program is long and intensive while performance is inadequate afterward, the method of training is ineffective. The optimum level of fidelity is still disputed today, and multiple reasons against perfect fidelity are given, such as cost or safety (Blaiwes, Puig, & Regan, 1973). The fact that some tasks show better performance after being trained through a transfer task over a higher fidelity simulator provides another reason to argue against high fidelity. The technological limitations available to researchers also determine the level of fidelity to adopt (Blaiwes, Puig, & Regan, 1973). One of the major problems in transfer research at this time was that percent transfer did not allow for comparing training programs with different lengths. To solve this problem the transfer effectiveness ratio (TER) was proposed (Blaiwes, Puig, & Regan, 1973). The TER takes the length of training into account and produces a ratio of hours saved on the job compared to hours trained. For example a ratio of 2 would indicate that 2 hours of on the job learning are saved for every hour a participant is trained. A related problem with determining training effectiveness is that evaluators' objective scores for a trainee often indicate failure while the evaluator's personal opinion is that adequate performance is being obtained (Spears, 1985). Blaiwes, Puig, and Regan (1973) suggest task analysis to aid in determining which tasks are critical for measuring performance to avoid this. Task analysis is a method used to understand the flow of operations or information used to complete a task, usually by breaking a large task into a series of subtasks (UsabilityNet, 2006). Additionally, measures that can be recorded by equipment or extracted from data are preferred. Lastly, intermediate measures should be used to evaluate individuals. These measures are subtasks that lead a user closer to the completion of a task (Blaiwes, Puig, & Regan, 1973).

Methods for determining the cost effectiveness of training are often problematic as well. Increasing fidelity increases cost but increasing fidelity does not increase transfer at a similar rate. For this reason the tradeoff between fidelity and transfer should be optimized to reduce costs (Blaiwes, Puig, & Regan, 1973). Designers should also seek to balance the amount and sequence of academic, on the job, and simulator training.

Investigation into finding an optimal level of fidelity has received much attention in the field of medicine (Chandra, Savoldelli, Joo, Weiss, & Naik, 2008; Grady, Kehrer, Trusty, Entin, Entin, & Brunye, 2008; Lamata, Gómez, Sánchez-Margallo, Lamata, Pozo, & Usón, 2006; O'Connor, Schwaitzberg, & Cao, 2008). An investigation into training proper nursing procedures (Grady et al., 2008) trained participants on both a low and high fidelity training mannequin. Participants randomly learned one of two skills (nasogastric tube or urinary catheter insertion) with a low fidelity mannequin, and learned the remaining skill with a high fidelity mannequin. The low fidelity mannequin merely provided a human shape, while the high fidelity mannequin provided auditory and haptic feedback such as a pulse rate. Recruited were 39 nursing students completing a standard introductory nursing class. Participants were tested on nasogastric tube insertion 9 weeks after training the skill, and tested on urinary catheter insertion 4 weeks after training the skill. All tests were conducted using the higher fidelity mannequin rather than real patients. Significantly better performance was noted when participants completed the skill learned under the higher fidelity condition, and participants reported more confidence after training with a higher fidelity mannequin.

Investigation into the most effective way to train respiratory therapists in oral intubation examined differing levels of fidelity (Chandra et al., 2008). Registered respiratory therapists trained with either a high fidelity virtual reality simulator (including haptic feedback) or a low fidelity wooden box that allowed repetition of training movements. Participants were allowed 1 hour with their simulator and were then given 2 attempts to perform fiberoptic orotracheal intubation on real world patients. There was a variable delay between training and testing lasting no longer than a single week. Participants were not given feedback in between intubation trials. Statistical analyses revealed no significant differences between participants trained with high or low fidelity simulators in terms of intubation accuracy or time to task completion. A main effect of trial number was found, in that participant accuracy increased regardless of training method.

A study of laparoscopic surgical (also known as minimally invasive surgery or MIS) training (Lamata et al., 2006) sought to determine which kinds of sensory information were most informative during simulator training. Researchers recruited 29 surgeons and asked them to rate the consistency of up to 10 different tissue samples using an 11 point rating system. Tissue samples were identified using only text information, only visual information, only tactile information, or using a combination of visual and tactile information. Participants were given no time constraints inside a stress free environment. Results indicated that tactile information was the single most useful type of sensory input, but overall the combination of visual and tactile information resulted in the most accurate ratings of tissue consistency. Due to the fact that participant ratings using an 11 point scale showed a low degree of overlap, researchers concluded that a simulator reporting a tissue's absolute consistency would be not be justified. Instead a simulator with no more than 5 levels of tactile feedback was deemed most appropriate.

Simon and Roscoe (1984) advocated for expanding the scope of individual transfer experiments. This was the first transfer experiment to use so many devices, training simulators, and transfer devices (Simon & Roscoe, 1984). Stating traditional methods that examined a single variable at a time were not cost effective, the authors compared their multivariate method to the traditional methods used previously. The multivariate approach was found to save a significant amount of time and money. To provide the same amount of data, traditional experiments would require over five times as many participants as the multivariate approach. Traditional data was found to be less accurate, less generalizable, and unable to examine interactions. Overall, the multivariate approach proved to require less participant data to provide a more thorough understanding of the task (Simon & Roscoe, 1984). This experiment refuted the belief that examining multiple variables at once was too time consuming and expensive to be worth a researcher's time.

Recommendations guiding the application of feedback in transfer experiments can be found in Lintern’s 1991 article. Lintern (1991) suggests that feedback in transfer experiments be limited to situations when a participant performs incorrectly. This technique serves to draw attention to the critical invariants in a task, while praising proper performance can distract attention (Lintern, 1991).

Similar to the quest for an optimal level of fidelity, research has been undertaken to determine the optimal amount of feedback during training (O’Connor & Schwaitzberg, 2008). Medical students participated in a study that trained them to improve their suturing and intracorporeal knot tying skills. Three levels of feedback were used: a condition with no feedback, a condition where participants were given knowledge of results (also known as KR, or whether a motor task was successfully performed or time to task completion), and a condition where participants were given KR as well as knowledge of performance (also known as KP, or corrective instruction regarding the movements during a trial). Participants were given a total of 24 hours of training spaced out over 4 weeks. After training sessions participants completed a NASA-TLX survey to measure the perceived workload of their tasks. Workload is a complicated construct that has been defined in many different ways (Hart, 2006). A common definition for workload is any mental or physical cost that must be paid while performing a task (O’Connor & Schwaitzberg, 2008). Analysis of participant learning curves found significantly greater performance in the groups receiving any feedback than the group that received no feedback. However, no significant differences were observed between the two feedback conditions. The same trends were observed for performance on the suturing and knot tying tasks. Analysis of NASA-TLX results demonstrated that the group receiving KR and KP feedback reported significantly less workload than the other groups. The KR and KP feedback condition also resulted in significantly fewer errors than the remaining conditions. Based on these results, authors recommended that KR only feedback be used to train medical students, as errors during training were seen as acceptable. Requiring fewer instructors would reduce costs, and allow students to train without scheduling supervision.

One particular study in the realm of KR (Anderson, Magill, & Seklya, 2001) found implications similar to those of previously mentioned interface research (McLaughlin, Rogers, & Fisk, 2009). Researchers told participants to draw a single continuous line from a starting point to a target midpoint and back to the starting point with a stylus. The goal of the task was to come as close to the target points as possible while drawing. The experimental task was carried out blindly, in the sense that participants never viewed the tabletop on which they were drawing. Instead participants viewed trials through a special display that only showed a single point onscreen at one time. At the start of a trial the starting point was displayed, and as soon as participants moved the stylus the display changed to show the midpoint until the trial's completion. In addition, lines were drawn using a participant's non-dominant hand. KR feedback was manipulated on two levels, with participants either receiving KR after a trial's completion or with a 2 trial delay (KR was given after two additional trials were completed). Intrinsic feedback was manipulated by placing a rubber spring on the end of the stylus used by half of the participants. Two retention tests were administered: the first 1 minute after training and the second 24 hours after training. During retention testing no KR information was given. Statistics performed on the training data found giving immediate KR resulted in higher performance. Analysis of the first retention test revealed that participant accuracy was significantly lower as testing blocks were administered. The 24 hour retention test found once again that performance declined as testing went on, and that the unmodified stylus condition resulted in significantly higher performance than the spring tipped stylus condition. Interestingly, at the 24 hour mark, the delayed KR group proved to be significantly more accurate than other participants. During all testing blocks participant accuracy dropped as time went on, with participants using the spring tipped stylus experiencing a significantly greater performance decrement than those using an unmodified stylus. The results of this study show that while a training method may show superior performance during a skill's acquisition, or after a short retention interval, it may not remain superior in the long run. In agreement with other feedback research (O'Connor, & Schwaitzberg, 2008), this study shows that participant error during training is acceptable and may be preferable. The findings of this study also support previous interface design work (McLaughlin, Rogers, & Fisk, 2009) in that it demonstrates assigning feedback schedules to broadly defined families of tasks can be just as ineffective as trying to determine "optimal" input devices without examining a task's input requirements.

Research examining feedback schedules has not been limited to summarizing performance after a task's completion. Augmented feedback, information not available in a task's environment, can be delivered while a task is being performed. Results from a series of experiments investigating augmented feedback while training table tennis skills (Todorov, Shadmehr, & Bizzi, 1997) indicated that the use of augmented feedback can provide better retention of skills with less practice than conventional training. Participants were given 50 attempts to hit a target with a table tennis ball using a standard table tennis paddle. Participants in the control group were given feedback from an experienced table tennis player during this time. Participants in the augmented feedback condition viewed a virtual reality (VR) display during training. Feedback from the VR setup modeled an expert swing for the experimental task. VR trained participants wore an electromagnetic sensor, allowing researchers to display the orientation and trajectory of the user's paddle during the trial. This allowed participants to compare their attempt to that of an expert in real time. Participants then attempted another 50 shots without any human or VR feedback. The accuracy and overall improvement of the VR trained group was significantly higher than the control group in the second block of target practice. In a second experiment, participants trained and tested on a more difficult table tennis shot over a period of 3 days. Control participants were allowed 50% more practice shots than the experimental condition. Results showed that VR trained participants once again made significantly more accurate shots than control participants. Accuracy improvements were present on each of the 3 days of testing, suggesting that the benefits of VR training did not level off (reach a point of inflection) quicker than traditional methods.

Several of the findings reviewed in this section demonstrated a concern for finding training methods that led to better performance over time, regardless of any decrements observed during actual training (Anderson, Magill, & Seklya, 2001; O'Connor, Schwaitzberg, & Cao, 2008; Todorov, Shadmehr, & Bizzi, 1997). Given the trend toward increasing self-care in medical practice, and the increasing demand for new training schedules it will require, the findings from these studies should be especially relevant in the medical arena (Bogner, 2004; Kaye & Chenault, 2002; Scialfa, Ho, & Laberge, 2004).

Medical Research

A study conducted by Raine in 2002 examined the occurrence of errors in self-monitored blood glucose (SMBG) testing. Conducted at a clinic where patients are routinely asked to bring their equipment for SMBG testing, Raine (2002) checked to see if patients had properly coded their glucometers to the indicated testing strip lot code indicated on the testing strip container. Researchers found that out of their 201 study patients, 84% had mismatched codes. Accurate meter/strip coding is a standard aspect of SMBG testing procedure. The clinic routinely requires patients to bring supplies, so no priming effects could have occurred. Also, patients without their equipment were not allowed back into the study, to prevent testing effects. The odds of having incorrect strip codes stored in a glucometer increased if a patient had type 2 diabetes, and increased again if that patient was following an oral medication regimen (not taking insulin). There was no effect of age or sex, suggesting that treatment plans and training are important for ensuring adherence to a care regimen. The study concluded that practitioners should be more involved in preventing errors in SMBG testing practice. In a list of seven common errors decreasing the accuracy of test results, four were related to the process of control solution calibration in blood glucometers (Raine, 2002). Though they did not differ significantly, HbA1c (Hemoglobin

A1c) test results were higher for those with incorrect strip codes. HbA1c testing is a blood test performed in a laboratory that measures the number of glucose molecules attached to the hemoglobin in the blood and is the standard of care for long-term estimates of glycemic control (over the past 6-12 weeks). While the difference in those levels may not have been significant (8.2% for incorrect coding vs. 7.7% for correct coding), the long term effects of those levels can be. Diabetes has been estimated to reduce life expectancy by up to 15 years. However, advances in the treatment of diabetes showed up to a 30% (up to 4.5 years) reduction in this figure from 1980-2000 (McKinlay & Marceau, 2000). While normally non-significant findings are just that, when a patient's life is involved an argument can be made for an effect size at the individual level.

A study by Bergenstal et al. in 2000 examined participant's SMBG testing procedure and compared their personal glucometer test results to those of clinical tests. The study used ambulatory patients not in emergency situations. Participants were recruited during routine diabetes education visits. Trials began with a questionnaire developed by experts asking for information about how a participant's current meter was obtained, the training that was provided for the current meter, and use of the meter. The content validity of the questionnaire was verified by outside experts. After the questionnaire, participants performed an SMBG test in front of a practitioner. A second sample of blood from the same site was used to obtain a value for an immediate laboratory test. A third sample of blood from the same site was used for an HbA1c measure. The participant then chose another finger and repeated the procedure (obtaining all 3 samples over again). The practitioner then completed a questionnaire about the participant's technique. If the participant's immediate test results were within 15% of the immediate laboratory test results, the participant was released. If either test was outside the acceptable range, the participant was given a quick re-education on proper technique and then repeated the testing procedure again (2 new fingersticks obtaining 3 samples from each fingerstick). After the initial testing was completed, 19% of the 280 participants (52) failed to achieve results within 15% of the laboratory test. After the re-education testing, the practitioner repeated the technique questionnaire. Even after re-education, roughly 31% of the remaining participants failed to achieve acceptable SMBG test results within 15% of laboratory results. The study also found that while 58% of participants reported being taught to use control solutions, 62% of participants used incorrect techniques in control testing procedures. This was the second most commonly observed error in the study. Incorrect control solution tests were found to be one of the most significant barriers to obtaining accurate SMBG test results. At the same time, it was found that regular use of control solution testing was not performed by participants in the first place. When participants did perform control tests at home, they commonly used incorrect techniques.

Research examining blood glucometers has also investigated required maintenance (Mykityshyn, Fisk, & Rogers, 2002). Younger and older participants having no experience with blood glucometers were asked to perform 3 calibration and control testing tasks with a single glucometer. Participants were initially trained using 1 of 2 methods developed by the researchers, a manual or video. After performing each task twice, participants did not return to the laboratory until a two week delay period had passed. At this point participants completed the calibration battery twice and were released. Overall, younger adults performed more accurately and faster than older adults, and were not influenced by their method of training. Participant performance improved over time, but displayed a decrement between training sessions. Older adults showed a significantly larger decrement in accuracy and task time after the delay period than the younger adults. Interestingly, older adults in the video training group performed as accurately as either younger adult group on the first day of testing (during trial 2). Older adults reported significantly higher workload for the calibration battery, as measured using the NASA-TLX. Older adults also reported significantly lower levels of confidence than younger adults for 3 out of the 4 total trials, and scored significantly lower when tested on their knowledge of the glucometer used in the study. However, older adults exposed to the video training reported higher confidence, lower workload, and higher knowledge scores than older adults exposed to the researcher constructed user manual.

Medical systems are designed with the intent of saving lives. Good intentions however, are no guarantee for success. In fact, there are times when medical systems themselves promote errors (Bogner, 2004, Linden, 2004). Linden (2004) provides a detailed account of how the procedures related to storing, transporting, and administering blood transfusions leave several openings where errors could easily be committed. While many may consider the risk of blood borne illnesses to be the greatest threat from a transfusion, patients are actually much more likely to receive the wrong blood type (Linden, 2004).

Hospital procedures commonly assign an identifier to unidentified patients admitted for treatment. The identifiers follow a logical pattern, such as UV8888888 followed by UW8888889. These identifiers are often assigned to many patients in rapid succession, such as when a large number of unconscious patients enter the emergency room at once, or when a number of unrelated incidents bring unrelated individuals to the hospital in a short period of time. Family members may be assigned similar identifiers regardless of when they enter the hospital because of their names (e.g., Richard Smith Senior and Junior). Without careful attention employees can easily confuse packets of blood with similar labels. In general, complex situations invite error (Linden, 2004).

While interface design (McLaughlin, Rogers, & Fisk, 2009) and feedback (Anderson, Magill, & Seklya, 2001) researchers have found a need for a finer level of granularity in their variables of interest, medical systems researchers may be too narrowly focused in their examination of error. The tendency to assume that an error was caused solely by the practitioner who committed it is myopic and in need of correction. In reality, the environment of a health care provider may impose constraints that lead to errors. One such theory (Bogner, 2004) of the environment's influence on error production describes the health care environment as a collection of 5 systems (from the outside in: Legal regulatory reimbursement national or culture factors, Organizational factors, Social environment, Physical environment, and Ambient conditions) enveloping the provider and health care recipient. Changes in one layer of the model cause changes in every layer underneath it. For this reason, focusing solutions at the practitioner level may be impossible or inefficient. The system can also encourage incorrect responses to errors. Linden (2004) gives the example of doctors who misdiagnose obvious signs of a transfusion error as other conditions because of lack of familiarity. Adequate training was not utilized because more likely conditions are first considered. Linden's (2004) writing on blood transfusions demonstrates a need for ensuring adequate transfer of training to real life situations. It also shows that errors must be studied to determine their underlying causes, and thus create safer systems.

McLaughlin, Rogers, and Fisk (2004) assert that older adults need special attention from designers and educators of equipment used to manage diabetes. The article addresses many age-related changes that can make the practice of SMBG testing more difficult for older patients, such as decreased contrast sensitivity (the ability to detect differences in areas based on the amount of lightness and darkness), poorer fine motor control, and more difficulty in visualizing actions. To increase the usability of equipment for diabetes management, cognitive walkthroughs, heuristic analysis, and user testing with older populations are suggested. The authors also suggest more direct training methods such as videos over reading a set of instructions (McLaughlin, Rogers, & Fisk, 2004).

Relevant Human Factors (HF) Research and Methodology Concerns

HF research has yielded recommendations for encouraging the use of technology by older adults, as well as advice for designing devices and training programs for the older adult user. In general, older adults are slower to adopt technology than their younger counterparts. They also require more training than younger adults to reach asymptotic performance. However, older adults are interested in new technology, and they even request training on devices younger adults consider intuitive (Rogers, Mayhorn, & Fisk, 2004). Older adults cite maintaining functional independence as a top reason for using technology, and they are more likely to use technology when training is available (Morrell, Mayhorn, & Echt, 2004). Older adults compare their current methods of performing actions to any new technology instead of instantly adopting new methods. If an older adult sees no added benefit from a technology (no reduction in response time using email versus sending a letter), or sees costs that outweigh potential benefits (the loss of emotional content when sending email, even with the use of emoticons), the new technology will not be embraced (Rogers, Mayhorn, & Fisk, 2004).

When designing technology intended for use by older adults, a user centered design process is recommended (Mayhorn, Rogers, & Fisk, 2004). Because of the stability or possible improvement of semantic memory, the use of metaphor is highly recommended. Metaphors allow individuals to learn a new idea utilizing information with which they are already familiar (Shakeri & Funk, 2007). Thus, metaphors can enable older adult users to capitalize on their unchanging (or improving) semantic memory. A common example of the use of metaphor is the computer “desktop”, which like a physical desktop can hold and organize “files” and “folders”. An iterative process of gathering data from the older adult population of interest is suggested as a method of obtaining the best selection of metaphors by Mayhorn, Rogers, and Fisk (2004). Indeed, employing an ineffective metaphor may complicate a user’s learning process instead of simplifying it, as the user must now first understand the metaphor before understanding the technology of interest (D. Kaber, personal communication, Fall 2007). Designers must also be cautious of overextending a metaphor to the point that it is no longer usable. For example, the desktop metaphor has been furthered in at least one previous operating system to the point that ejecting a diskette involved dragging the diskette's icon into a virtual trashcan. Users may have preferred breaking the metaphor for a simpler menu option to eject portable media.

Research in HF investigates issues related to both medicine and aging, but the methods for neither area are perfected. According to Nichols, Rogers, and Fisk (2003), the changes associated with chronological age effect a majority of mental and physical abilities. However, the authors criticized HF research for reporting poor information on age. The authors criticized the fact that studies commonly defined "adult" as any individual older than 20 years of age. Important information about age-related trends was lost when an adult age condition included 20 year olds and 80 year olds. Studies also commonly defined older age groups as age sixty five and above. In their 2003 article, Nichols, Rogers, and Fisk call for increased attention to the age groups used in research. If expected sample sizes are small, they recommend using extreme-group, cross-sectional designs. An example of an acceptable design here would be to compare a younger group of 20-30 year olds with an older group of 65-75 year olds. By keeping age ranges small, researchers gain more information on specific age-related effects. This boosts the inferential ability of a study, and aids future meta-analyses that would examine age-related effects. Another reason for using smaller age ranges was that this method makes a sample more representative.

Training and Older Adults

The construction of training programs for any technology should seek to capitalize on the strengths of the older adult population. The systems approach to creating training programs, advocated by Mayhorn, Stronge, McLaughlin, and Rogers (2004) aids designers in this process. The systems approach is a user centered design process that involves the input of participants at every step. Researchers start the process by performing a needs analysis. Once it has been determined that a training program is warranted, identifying what participants do and do not want to learn occurs. A task analysis is performed for every element of the curriculum, with special attention dedicated to identifying areas of potential error. A person analysis is conducted to determine the strengths and weaknesses of the target population to ensure the effectiveness of the training. Specific training methods are chosen with input from participants and relevant literature. Finally, the training program is evaluated through qualitative in-depth interviews with participants and quantitative statistical measures of usability, retention, and transfer.

At a glance, blood glucometers and automated teller machines (ATMs) appear to be very different devices. Blood glucometers are small, portable, commonly used on a flat tabletop, and designed for a medical purpose. ATMs by comparison are large, stationary, interfaced with by using a display in front of the user, and are designed for banking transactions. However, many kinds of ATMs exist today, just as many different kinds of blood glucometers are currently on the market. While it is possible for a consumer to use only one type of ATM or glucometer for a period of time, both devices are routinely upgraded. Indeed, blood glucometers contain 2 batteries: one that is replaceable and another internal battery that cannot be replaced (M. King, personal communication, June 10, 2009). Internal batteries have an average lifespan of 5 years, and a glucometer must be replaced when its' internal battery fails. During this 5 year period, a particular model of glucometer can be expected to go through at least 2 upgrades. ATMs and glucometers also share the property of feature creep (accumulation of features over time) as they are upgraded (Jamieson, Cabrera, Mead, & Rousseau, 1995; Jamieson & Rogers, 2000). The risk of developing diabetes increases as one ages (Centers for Disease Control and Prevention, 2007), meaning newly diagnosed older adults have little or no experience with blood glucometers. Research focusing on training older adults to use ATMs has been performed with samples that lack experience with ATMs (Jamieson, Cabrera, Mead, & Rousseau, 1995; Jamieson & Rogers, 2000). For these reasons, findings from research examining how older adults learn to use ATMs should be applicable to studies of how older adults deal with transferring to novel glucometers.

While studies on technique errors can be found, little research is available on near and far transfer errors in older participants. A study conducted in 2000 addressed the ability of older participants to learn how to use automated teller machines or ATMs (Jamieson & Rogers, 2000). A group of younger and a group of older participants were tested in this study. To ensure similar groups across age, participants were given visual screenings and asked about medications that can interfere with attention. Participants were also given multiple cognitive ability tests, and asked about their average computer use and ability.

All participants performed practice trials on a simulated ATM, run on a laboratory computer. Participants were trained in how to operate the mouse on the computer before trials began to prevent time delay due to unfamiliarity with computers. Participants in both age groups were randomly trained with either a blocked or random practice method on ATM simulator 1. Blocked practice routines performed the same type of transaction over and over before proceeding to a new transaction type. Random practice routines randomized transaction types, with no apparent pattern to presented transactions. The random practice condition was counterbalanced to prevent presentation effects. Participants were then asked to use a second simulation. In ATM simulation 2, the layout of features was different from ATM simulation 1, and new features were added to the ATM. The trials on ATM 2 used a random presentation method. This was done to make the task more difficult to stress the differences between the training styles, as well as to simulate real life conditions when one would actually use an ATM.

The study found that, regardless of age, random practice participants had better completion rates than blocked participants. Completion rate was defined as a participant's ability to finish a requested transaction. However, older adults benefited less than younger adults from random practice presentation. No effect of practice method was shown on the speed of a transaction. It was suggested that familiarity with computers could play a part in why younger adults were more accurate when performing transactions.

Another study by Jamieson, Cabrera, Mead, and Rousseau in 1995 investigated training older adults to use ATMs. Older adults that had never used an ATM before were recruited for the study. The age of the participants ranged from 65 to 80 years. There was no group of younger participants to study age-related effects. Two groups were constructed at random, differing in the type of instruction they received. One group received the basic written instructions that could be expected from a bank on how to use an ATM, while the other received hands-on training on a simulator built by the researchers. The study took a total of three days, with the first day obtaining demographic information, ATM use information, and training on how to use a computer mouse. The second day consisted of giving participants their instructional material (either the handout alone or the handout with the computer simulation), giving participants time to review their materials, and then having participants perform transactions on the ATM simulator. Blocks of trials were constructed to assess participant ability. Blocks were always performed in sequential order (block 1, block2, block 3, and so on), where blocks 1 and 3 were exactly the same in requested procedure, dollar amount, and account number. Blocks 2 and 4 had the same number and order of transactions as 1 and 3, but the account numbers and dollar amounts were allowed to differ. On day 3, participants completed two additional blocks with the first ATM simulation. Block 5 had the same exact content as block 1, and block 6 was similar to blocks 2 and 4 in that the dollar and account numbers were allowed to vary. Participants were then switched to a second ATM simulation. This simulation retained all of the original features of the first ATM, and added several additional features (transaction types). The layout of the device was different as well, with the old familiar functions being in new places. Block 7 contained requests identical to block 5, while block 8 was a perfect match to block 6. These identical blocks of practice were used as a measure of near transfer, as participants were asked to perform familiar tasks on a new device (Jamieson et al., 1995). The last two blocks, 9 and 10, contained new transaction types that participants were not previously exposed to. Participants had no warning that performing new transaction types would be required of them. Training blocks 9 and 10 were used to measure far transfer, as participants were asked to perform unfamiliar tasks on a device that was familiar to them (Jamieson et al., 1995).

Researchers measured the amount of time each transaction took, as well as if participants were able to complete the transaction. Participant mistakes were not recorded, a participant was simply allowed unlimited time and was allowed to go back and repeat steps as long as they wanted. The sole criterion for completeness or correctness was if the participant was able to complete all of the required steps. The study found that older adults who received online training completed more of the transactions than those who were not trained online. This was true for the initial simulator, as well as the near and far transfer tasks on the second simulator. The same was true for the amount of time required to complete a transaction. While not all of the blocks were significantly different in completeness or time, any significant differences were always in favor of the group trained online.

The study showed that older adults trained with a real-time method performed better than a group of older adults who simply read a list of steps. Only the first few blocks of testing favored the computer trained participants, but their accuracy was better throughout the whole experiment. This supports online training with the same system that is about to be used when the population is between 65-80 years of age.

Conceptual Integration

While the reviewed literature provides a unique perspective regarding transfer in elderly populations or errors committed by diabetic participants, no single study has attempted to examine these two issues together. While previous studies have examined technique errors when a patient used his or her current blood glucometer (Bergenstal et al., 2000; McLaughlin, Rogers, & Fisk, 2004; Raine 2002) or a single novel glucometer (Mykityshyn, Fisk, & Rogers, 2002; Rogers, Mykityshyn, Campbell, & Fisk, 2001), no study measured errors as a patient transferred to using a second, novel experimental glucometer. At the same time, while it was repeatedly found that control solution testing was a large source of error in the results of SMBG testing, no single study has been found that focuses on transfer and errors during the control solution testing process. Thus, the research reported here helps to close that gap by measuring near and far transfer in control solution testing. Near and far transfer is of special interest in the older segments of the population because of the previously mentioned cognitive changes associated with aging (Howard & Howard, 1997; McLaughlin, Rogers, & Fisk, 2004). Semantic memory and crystallized knowledge remain mostly intact, allowing patients to capitalize on previously learned concepts and metaphors. Working memory and general processing speed declines may result in difficulty learning unfamiliar technologies and procedures. As new glucometers introduce new features unfamiliar to older users, these patients may have more difficulty using them. For example, some newer glucometers are using “testing disks” rather than individual testing strips. These disks contain multiple strips inside of them, and instead of inserting a new strip into the glucometer to start each test, the disk must be rotated to expose an unused blood application area.

Empirical Questions

The completed research compared transfer across age ranges concerning a participant's ability to accurately perform a control solution test with a blood glucometer. Empirical questions of interest included: How would the characteristics of each glucometer influence a participant's ability to perform an accurate test when transferring to a new glucometer? Would age be an important variable in determining accuracy or time to complete a test? How would working memory, semantic memory, inferential ability, and perceptual speed relate to a participant's performance? Would overall workload or NASA-TLX subscale measures differ between meters or age groups?

Research Hypotheses

H_1 - It was hypothesized that younger adults would perform significantly fewer errors than older adults in all error categories (fewer errors when using a first glucometer, fewer near transfer errors, and fewer far transfer errors with a second glucometer) (Charness, Schumann, & Boritz, 1992; Gist, Rosen, & Schwoerer, 1988; Howard & Howard, 1997; Howell, 1997; Jamieson & Rogers, 2000; Kline & Scialfa, 1997; Mayhorn, Lanzolla, Wogalter, & Watson, 2005; McLaughlin, Rogers, & Fisk, 2004; Mykityshyn, Fisk, & Rogers, 2002; Sharit, Czaja, Nair, & Lee, 2003).

H_2 - It was hypothesized that younger adults would perform significantly faster than older adults in the control solution testing task (Charness, Schumann, & Boritz, 1992;

Howell, 1997; Mayhorn, Lanzolla, Wogalter, & Watson, 2005; McLaughlin, Rogers, & Fisk, 2004; Mykityshyn, Fisk, & Rogers, 2002).

H_3 – It was hypothesized that both age groups would perform significantly more far transfer errors than near transfer errors (Jamieson & Rogers, 2000; Rogers, Mykityshyn, Campbell, & Fisk, 2001).

Chapter 2

Method

Participants

The reported research used a design of age (2: younger and older adults) by order of meter type presentation (2: OneTouch UltraMini and Nova Max Link) between subjects factorial. A total of 84 participants were recruited. One age group consisted of young adults from ages 18-28 ($M=19.62$, $SD=1.40$), and the other consisted of older adults from ages 65-75 ($M=70.23$, $SD=3.29$). Participants without diabetes were recruited to minimize the effects of previous experience with the equipment and task. Consistent with the recommendations of Nichols, Rogers, and Fisk (2003) the study incorporated an extreme-group, cross-sectional design. Specifically this design was used to examine the presentation of different meter types (in different orders) to differing age groups, while recording reaction time, errors with the training meter, near transfer errors, far transfer errors, and NASA-TLX scores as dependent variables. Counterbalancing the presentation of glucometers prevented any presentation effects from biasing the results of the study. Due to the differing cognitive and motor requirements between differing glucometers, the design of a participant's first glucometer could be important in determining future ability to use a new glucometer correctly. Variables influencing participant transfer accuracy were expected to be especially relevant for older adults, as declining working memory capacity and difficulties learning novel unfamiliar motor tasks have been established as age-related trends (Howard & Howard, 1997). The specified ages were determined from reviewing the relevant literature

(McLaughlin, Rogers, & Fisk 2004; Jamieson, Cabrera, Mead, & Rousseau 1995; Nichols, Rogers, & Fisk 2003; Rogers, Cabrera, Walker, Gilbert, & Fisk 1996). When conducting research with older participants, the age of 65 is an acceptable minimum, and it is also close to the age at which the Centers for Disease Control and Prevention reports a large increase in diagnoses of diabetes (Centers for Disease Control and Prevention, 2007; McLaughlin, Rogers, & Fisk 2004). The ten year range for each participant age group was determined from the recommendation in Nichols, Rogers, and Fisk (2003) that age ranges be small enough to capture age-related effects. The minimum of 18 for the younger adult population was set to maximize potential participants without the use of specialized underage consent techniques.

ANOVA testing performed on the results of the cognitive battery and other demographic variables using experimental conditions as IVs revealed several significant differences. Overall significant differences in ability scores were found such that young adults (YAs) used significantly more pieces of handheld technology per day, reviewed instructional materials provided for significantly less time, progressed significantly further in the DSS (Digit Symbol Substitution) test (Wechsler, 1997), remembered significantly more symbols from the DSS, scored significantly higher via the absolute method of grading on the Computation Span (Salthouse & Babcock, 1991), had achieved significantly less education, scored significantly lower on the Shipley Vocabulary test (Shipley, 1986) than older adults (OAs), rated composite NASA-TLX workload scores as significantly lower than older adults, and rated each subscale except for temporal demand on the NASA-TLX significantly lower than older adults. The only significant difference noted on abilities by order of meters used was that individuals using the Nova MaxLink followed by the OneTouch UltraMini scored significantly higher than other groups via the simple scoring method of the Computation Span (see Tables 1-3 and Figures 1-4 for more detailed cognitive battery and demographic information, as well as a comparison of overall and analyzed sample). No significant interactions between age and meter order were discovered as a result of this testing.

Table 1

Summary of Cognitive Battery Results by Age Group

Test	F	$p<.05$
Tech used	9.26	yes
Review time	6.12	yes
DSS speed	37.17	yes
DSS memory	21.49	yes
Shipley	70.74	yes
Inference	0	no
Simple WM	1.7	no

Table 1 Continued

Absolute WM	8.29	yes
Education achieved	103.01	yes
Overall NASA-TLX	15.53	yes
Mental TLX rating	8.9	yes
Physical TLX rating	6.28	yes
Temporal TLX rating	1.19	no
Performance TLX rating	9.64	yes

*Note: DSS speed=number of symbols matched with correct numeral (out of 100), DSS memory=number of recalled symbols (out of 9), Shipley=number of vocabulary items correctly answered (out of 40), Inference=number of correct inferences made (out of 20), Simple WM=number of rounds where at least 2 of 3 trials are answered in complete serial recall followed by .5 if the round after failing this criterion contains at least 1 serially recalled trial (out of 7), Absolute WM=number of trials successfully serially recalled (out of 21), .Education=highest education completed where high school diploma=2, bachelor's degree=4, Masters=6, and M. D., J. D., Ph.D. or other advanced degree=7.

Table 2

Summary of Cognitive Battery Results by Meter Use

Test	F	*p*<.05
Tech used	0.14	no
Review time	1.1	no
DSS speed	0.37	no
DSS memory	2.63	no
Shipley	1	no
Inference	0.12	no
Simple WM	7.57	yes
Absolute WM	2.25	no
Education achieved	0.05	no

Table 3

Comparison of Overall and Analyzed Sample

DV	IV	Overall Sample M(SD)	Analyzed Sample M(SD)
T Errors	YA	0.04(0.02)	0.04(0.02)
NT Errors	YA	0.03(0.03)	0.03(0.03)
FT Errors	YA	0.51(0.27)	0.49(0.24)
Time	YA	67.76(24.14)	69.20(26.42)
T Errors	OA	0.08(0.03)	0.07(0.03)
NT Errors	OA	0.07(0.04)	0.07(0.04)
FT Errors	OA	0.55(0.18)	0.52(0.20)
Time	OA	148.68(62.73)	148.68(62.73)
T Errors	M1	0.06(0.03)	0.06(0.03)
NT Errors	M1	0.03(0.02)	0.03(0.02)
FT Errors	M1	0.70(0.15)	0.68(0.17)

Table 3 Continued

Time	M1	99.29(51.23)	406.03(52.76)
T Errors	M2	0.05(0.03)	0.05(0.03)
NT Errors	M2	0.07(0.04)	0.07(0.04)
FT Errors	M2	0.34(0.14)	0.35(0.13)
Time	M2	109.15(70.22)	111.85(71.32)

*Note:T=training, NT=near transfer, FT=far transfer, M1=use of OneTouch followed by Nova, M2=use of Nova followed by OneTouch, Time measured in seconds, Errors reported as averaged percentages for all applicable trials.

Recruitment of both groups required some effort. Young adults were recruited via NCSU's Introductory Psychology pool, and by fliers posted in Poe Hall. Older adults were recruited from a list of older adults who previously participated in research by NCSU students and faculty.

All trials occurred in a laboratory setting on campus. The lab resembles a normal office setting, with a conference table and 6 personal computers. A consistent setting was desired to prevent unique environmental effects from clouding the results of the research (which could occur if all older participants were tested at a familiar senior center while all younger participants were tested in an unfamiliar laboratory setting).

Materials

Two different blood glucometers were used for this experiment, the OneTouch UltraMini and the Nova Max Link. While it is possible that a participant may have previous experience with one or both of the chosen meters due to exposure from diabetic friends and family, it would be desirable for both meters to be new and unfamiliar to each participant. Recruiting participants with minimal exposure to glucometers enhanced the external validity of the current research in that participant behavior should be similar to that of individuals recently diagnosed with diabetes. To ensure the desired sample was recruited, a survey was constructed to assess each participant's level of experience with either meter used in the study. Potential participants found to have previous experience with blood glucometers were excluded from the study. This prevented any previous experience from affecting the results of the reported research. The glucometers used in this study were donated and recommended by the principal investigator's personal endocrinologist at the time, Dr. Glenn Stall. The two models were both popular and currently in use by the public. Not only were the glucometers used available in pharmacies nationwide, free units are frequently made available to diabetic patients with a small number of testing strips, allowing users to "test drive" new glucometers before they upgrade. The principal investigator has been unable to find any evidence that variables such as cost or number of units sold are indicators of a glucometer's usability, and thus these factors were not considered during the selection process.

The manufacturer's instructions for performing a control solution test were provided to participants as a teaching tool. Participants were also given the materials required to perform a control solution test with the glucometer. These materials vary, but commonly include two different bottles of solution, standard testing strips, and a cleaning cloth or napkin of some sort in addition to the glucometer. During testing, a stopwatch was used to time participant trials and a standard hard drive video camera was used to record participant trials.

The NASA-TLX (task load inventory) was used to measure participant perception of task-related workload. The NASA-TLX is a standard workload inventory, having been used for over 20 years in over 550 studies (Hart, 2006). Participants completed the full scale after every trial, and completed weightings after the fourth, sixth, and tenth trials. Moreover, a battery of cognitive tests was administered to participants to measure individual differences in cognition that might later be used to assess correlations between cognition, age, time to task completion, and error rates. Semantic knowledge was measured by the Shipley Vocabulary test (Shipley, 1986). Working memory was assessed by administering the Computation Span test (Salthouse & Babcock, 1991). Participant perceptual speed was measured with the Digit Symbol test (Wechsler, 1997).

A behavioral checklist was used by observers to measure the rate of errors in the testing procedure. This checklist was constructed from a cognitive task analysis (CTA) and hierarchical task analysis (HTA) performed using the instructions for each glucometer, as well as behaviors mentioned in Bergenstal et al. (2000) and Jones (1994). The CTAs were conducted using the manufacturer’s instructions for the glucometer and Gordon and Gill’s

(1997) chapter on the technique. The HTAs were conducted using the manufacturer's instructions for the glucometer and Sheperd's 2001 book on the subject.

Measurement

Test time (in seconds) was a simple measure of how long a participant took to perform a complete control solution test. Timing began when a participant indicated they were ready, and ended when a participant indicated that they were finished.

Error rate was a measure of deviation from standard testing procedure. To ensure accurate data capture all experimental sessions were videotaped. The behavioral checklist used by observers listed errors the researchers expected to see participants commit. These errors were listed in order of their expected occurrence to ease coding. Adequate space was provided to record any unexpected errors observed during trials. A fresh checklist was used for each control test a participant performed. For each test, the sum of all observed errors was totaled. In all, error was measured with 3 variables, errors committed with the first meter, near transfer errors committed with the transfer meter (such as inserting a test strip backwards), and far transfer errors (errors committed with new options such as a food log) committed with the transfer meter.

The presentation of the meters was counterbalanced (as referenced by the variable of meter type) to combat presentation effects. Recording meter type allowed researchers to investigate the influence of glucometer complexity. Two groups were recorded (A and B) to signify which meter was presented first. Meter type group A was trained on the OneTouch UltraMini, and used the Nova Max Link as a transfer meter. Meter type group B was trained on the Nova Max Link, and used the OneTouch UltraMini as a transfer meter.

Procedure

This study required participants to schedule and attend two separate experimental sessions, with the second session occurring on the next consecutive business day. While diabetic patients may go longer than a single day before recalibrating with control solutions, this decision was made to minimize participant attrition. Participants used a blocked presentation method. While evidence suggests that a randomized practice method increases learning and decreases transfer errors (Jamieson & Rogers, 2000), a patient in the real world is not required to change control testing procedures until a new meter is purchased. Ultimately, the decision was made in the interest of external validity.

Before the start of a trial, participants were randomly assigned to either meter type group A or B, ensuring that presentation effects would not be an issue with the study. The randomization process was forced, alternating participants between groups A and B as trials were run. This forced process was used as participants were likely to trickle in, and alternating trials would reduce history and maturation effects on the data analysis.

Upon arriving at the laboratory site, all participants were given a copy of the informed consent document. The information on the document was reviewed orally, and two copies of the document were signed by both parties. Participants then filled out a brief demographic form that requested background information before the cognitive battery was administered to each participant.

After completion of the cognitive battery, participants were given the written instructions for performing a control solution test for their first glucometer. The experimenter gave the least amount of feedback possible to prevent the biasing of experimental results. Answers to expected questions included encouraging participants to try their best, ensuring participants that there are no “wrong” decisions, and telling participants that frustration is acceptable as the study was designed to be difficult. Initially the experimenter was not allowed to give feedback concerning whether an action was right or wrong. During initial testing with older adults the Nova MaxLink glucometer required replacing due to participants allowing solution to enter the glucometer. For this reason the experimenter was allowed to instruct participants not to allow control solution to enter the glucometer. Minimal feedback is recommended in the literature covering fidelity and feedback (Anderson, Magill, & Seklya, 2001; Chandra et al., 2008; Lamata et al., 2006; Lintern, 1991; O’Connor, Schwaitzberg, & Cao, 2008; Todorov, Shadmehr, & Bizzi, 1997). Participants were given unlimited time to review the instructions, and upon indicating they were ready to proceed, were presented with a glucometer and all of the materials necessary for testing. When a participant again indicated readiness or began the testing process, an observer started a timer, and began watching for errors to be marked on the behavioral checklist. Whenever a participant indicated that he or she was finished, the observer stopped the timer and recorded the testing time. After each trial participants completed a NASA-TLX survey. The observer obtained a new checklist, and the participant began again, until 4 control tests were performed. The participant was then released.

The next day, the participant returned for a follow-up. The same overall procedure was used, with minor differences. The participant was not presented with instructions for performing a control solution test. Instead, the participant was required to use the glucometer from the previous session and perform two control solution tests from memory. At this point, the second glucometer and all of its materials were presented. The observer then used the checklist for the second glucometer. These trials allowed the examination of near and far transfer errors. As an example, a near transfer error would be putting a testing strip into the new glucometer backwards. A far transfer error would be a participant's inability to enter an accurate strip code into the glucometer. The participant was given no instructions on how to perform a control solution test with this glucometer. Relying on previous knowledge, participants were required to perform four control solution tests with the second (transfer) glucometer. After completing the last control solution test, the participant was debriefed and released.

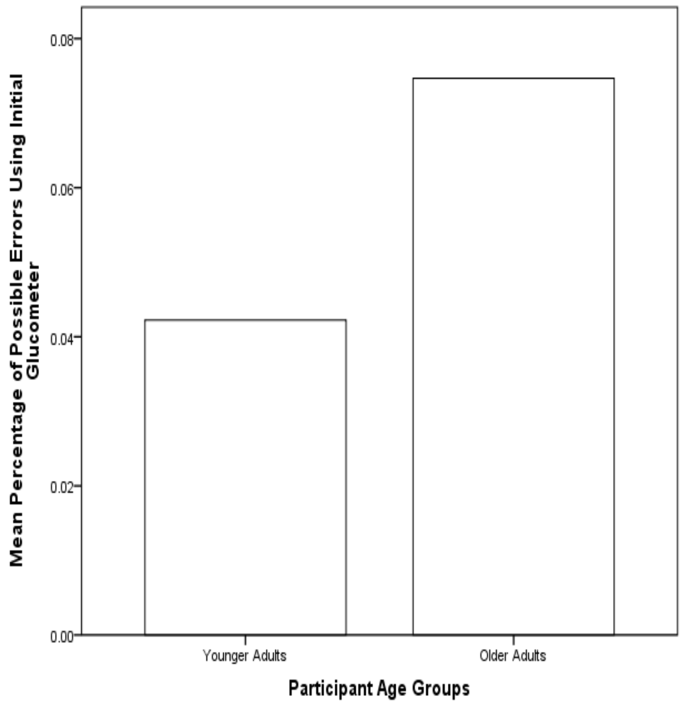

Figure 1 Initial Errors by Age Group

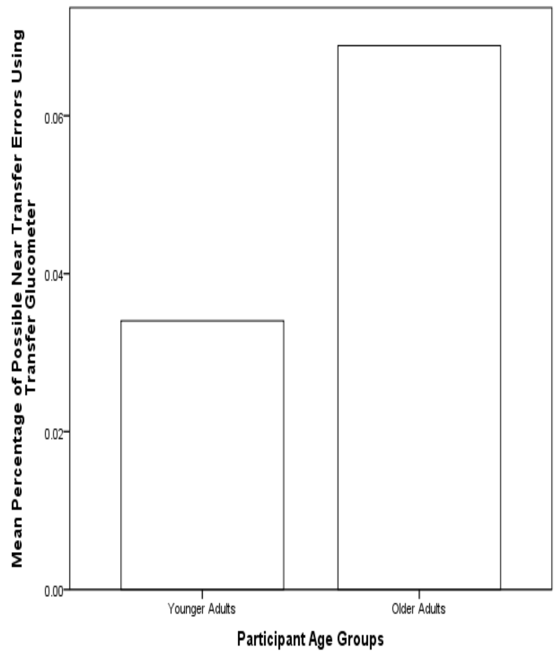

Figure 2 Near Transfer Errors by Age Group

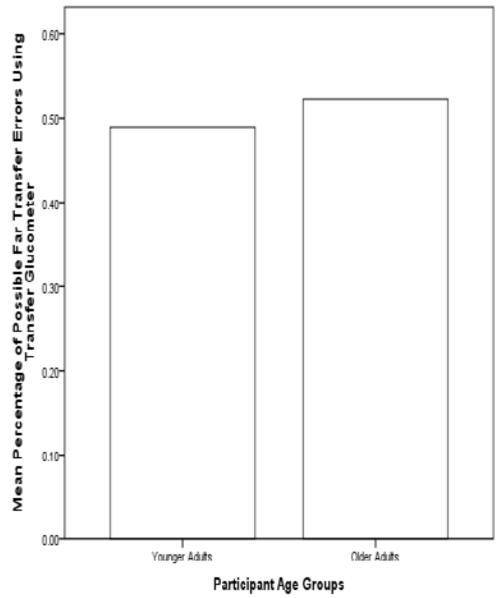

Figure 3 Far Transfer Errors by Age Group

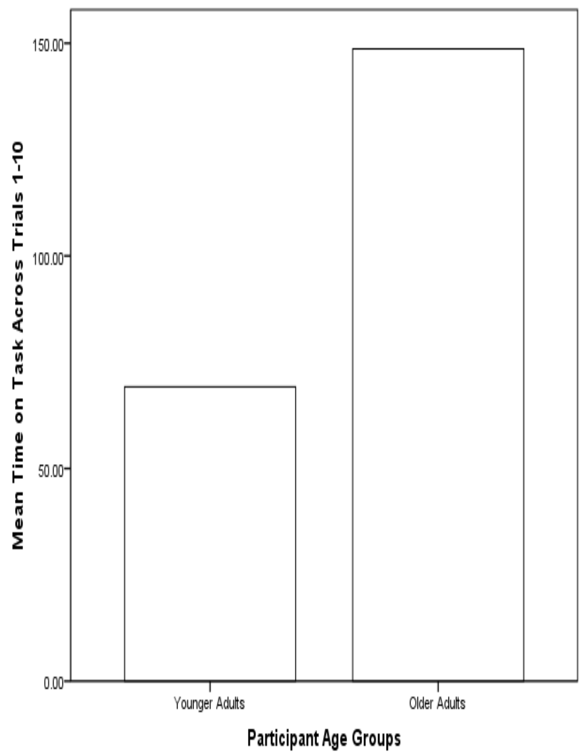

Figure 4 Time on Task by Age Group

Chapter 3

Results

Multiple analyses were conducted to examine significant differences between participant groups and identify unique predictors for multiple aspects of performance. MANOVA analysis was employed to test Hypotheses 1-2 by investigating the influence of participant age group and presentation of glucometers on the dependent variables of task time, initial errors, near transfer errors, and far transfer errors. Due to Box's test of equality of covariance matrices being violated with $p<.001$ when testing was conducted using all participant data, participants were dropped from analysis at random such that a subset of sixty-four participants were included in the analyses discussed below. Violation of Box's test indicates that the covariance matrices between the dependent variables for the dataset were not homogenous, a requirement of MANOVA testing. Accepted practices for MANOVA in this situation are to randomly drop participants from the larger groups until either Box's M yields a *p* value of greater than .001, or to assume MANOVA to be robust against violations of homogeneity of covariance matrices when group sizes are equal (Field, 2005). For this reason all analyses reported will only reflect a subset (n=64) of all data collected. Multiple regression was used to examine the predictive value of cognitive battery results and demographic information on the criterion of task time and number of session 1 (trials 1-4), retention (trials 5-6), initial (trials 1-6), transfer (trials 7-10), and overall (trials 1-10) errors committed.

MANOVA Results

To test hypotheses 1 and 2 and determine if participant age and the order of meters used influenced the amount of time on task, number of errors committed with an initial meter, number of near transfer errors, or number of far transfer errors committed with a second meter, a multivariate analysis of variance (MANOVA) was conducted. Time was measured in seconds and averaged between all 10 trials. Due to differing numbers of possible errors for each meter, each type of error (training, near, or far transfer) was calculated into a percentage of possible errors for each trial, and all trials were averaged together. Results revealed a significant main effect for age, $F(4, 61) = 18.26$, $p < .05$, $\eta^2 = .55$ and order of meter use $F(4, 61)=53.46$, $p<.05$, $\eta^2=.78$ (see Table 4). Correlation analysis was performed to test for problematic relatedness between dependent variables (see Table 5). MANOVA results should be interpreted with caution as far transfer errors and task time were not normally distributed in some conditions. Follow-up ANOVA's revealed a significant main effect of age for task time $F(1, 64)=45.15$, $p<.05$, initial errors $F(1, 64)=32.23$, $p<.05$, and near transfer errors $F(1,64)=29.92$, $p<.05$, as well as a significant main effect of order of meter use for near transfer $F(1,64)=40.46$, $p<.05$ and far transfer $F(1, 64)=100.76$, $p<.05$ errors committed. Results indicated that YAs took significantly less time to complete control testing, and performed significantly fewer initial and near transfer errors than OAs. Individuals using the OneTouch UltraMini first performed significantly fewer near transfer errors while those using the Nova MaxLink first performed significantly fewer far transfer errors. Overall these results lend partial support to hypothesis 1 that older adults would commit significantly more initial, near transfer, and far transfer errors than younger adults. Results also supported research hypothesis 2, that older adults would require significantly more time to complete control solution testing than younger adults.

Table 4

Error and Time on Task by Age Group and Glucometer Counterbalance

DV	IV	M(SD)
T Errors	YA	0.04(0.02)
NT Errors	YA	0.03(0.03)
FT Errors	YA	0.49(0.24)
Time	YA	69.20(26.42)
T Errors	OA	0.07(0.03)
NT Errors	OA	0.07(0.04)
FT Errors	OA	0.52(0.20)
Time	OA	148.68(62.73)
T Errors	M1	0.06(0.03)

Table 4 Continued

NT Errors	M1	0.03(0.02)
FT Errors	M1	0.68(0.17)
Time	M1	106.03(52.76)
T Errors	M2	0.05(0.03)
NT Errors	M2	0.07(0.04)
FT Errors	M2	0.35(0.13)
Time	M2	111.85(71.32)

*Note:T=training, NT=near transfer, FT=far transfer, M1=use of OneTouch followed by Nova, M2=use of Nova followed by OneTouch, Time measured in seconds, Errors reported as averaged percentages for all applicable trials.

Table 5

Correlation Matrix for MANOVA Dependent Variables

Variable	Task Time	Errors 1-6	NT Errors	FT Errors
Task Time	1	0.45	0.45	0.02
Errors 1-6	0.45	1	0.42	0.18

Table 5 Continued

1NT Errors	0.45	0.42	1	-0.22
FT Errors	0.02	0.18	-0.22	1

To test hypothesis 3 which suggested that both age groups would commit a higher incidence of far transfer than near transfer errors, a repeated measures ANOVA (RM ANOVA) was conducted. The between subjects factor was age, and the within subjects factor was error type (near or far). Results revealed a significant main effect of error type ($F(1,74)=286.91$, $p<.05$, $\eta^2=0.80$) such that the incidence of far transfer errors was significantly higher than that of near transfer errors. Thus hypothesis 3 was supported by the data.

Regression Results

Multiple regression testing was performed to discover significant predictors of time on task. All regressions used the variables of participant gender, age, scores for speed and memory from the Digit Symbol Substitution (DSS) Test of information processing, semantic memory score taken from the Shipley Vocabulary Test, inferential ability score, both simple and absolute scores of working memory capacity from the Computation Span Test (simple scores involve finding the first point at which participants are unable to reach a specified threshold in recall, while absolute scores examine the entire test for accuracy) and averaged

TLX scores for the trials investigated in each model. Correlations run between these predictors after analyses were completed led to hierarchical regression testing described later (see Table 6).

Results of testing for predictors of participant time on task using the averaged trial time across overall trials (all ten experimental trials) found a statistically significant model ($F(9, 61)=5.63$, $p<.05$) explaining 49% of the variance in the dataset (see Table 7). While a participant's age, memory for DSS symbols, semantic memory score, absolute WM score, and averaged NASA-TLX scores were all positively correlated with a participant's time on task, only age was uniquely predictive ($b=0.67$, $t=3.77$, $p<.05$). All other variables from the cognitive battery were negatively correlated with time on task; however, none were statistically significant in their predictive ability. A multiple regression examining the averaged time on task for initial trials (trials 1-6) resulted in a model ($F(9, 65)=7.39$, $p<.05$) explaining 54% of the variance in the dataset (see Table 8). While participant age, memory for symbols on the DSS, averaged NASA-TLX scores, and absolute WM score were positively correlated with time on task, only age was significantly predictive ($b=0.64$, $t=4.00$, $p<.05$). While gender, speed on the DSS, semantic memory, inferential ability, and simple WM scores were negatively correlated to time on task, none of these scores were uniquely predictive of time on task. Testing on average time for trial for a participant's session 1 trials (first four trials) with instructions provided resulted in a model ($F(9, 70)=5.70$, $p<.05$) that explained 46% of the variance in the dataset (see Table 9). While age, memory for DSS symbols, averaged NASA-TLX scores, and absolute measures of WM capacity were all positively correlated with time on task for these trials, only age (b=0.45, t=3.47, p<.05) was significantly predictive. While gender, speed on the DSS, semantic memory, inferential ability, and simple measures of WM capacity were all negatively correlated with time on task, none of these variables was uniquely predictive. Testing based on the averaged time on task for retention of training (trials five and six) where participants were required to test with their initial glucometer from memory resulted in a model ($F(9, 69)=5.96$, p<.05) that explained 47% of the variance in the dataset (see Table 10). While participant age, speed on the DSS, memory for DSS symbols, semantic memory score, NASA-TLX scores, and absolute WM scores were positively correlated to time on task, only age (b=0.45, t=2.72, p<.05) and NASA-TLX scores (b=0.43, t=3.75, p<.05) were significantly predictive. While participant gender, inferential ability, and memory for symbols on the DSS, inferential ability, and simple WM scores were negatively correlated with time on task, none of these variables were statistically significant in their predictive ability.

Examination of Time on Task During Transfer Trials

Testing completed using averaged time on task for transfer trials (trials seven through ten) resulted in a model lacking predictive ability ($F(9, 69)$=, p<.05) that explained 32% of the variance in the dataset (see Table 11). Positively correlated variables in this model were age, DSS speed, inferential ability, simple WM scores, and averaged NASA-TLX scores. Negatively correlated variables were gender, DSS memory, semantic ability, and absolute WM scores. The only predictive variable in this model was the averaged NASA-TLX scores (b=0.48, t=4.17, p<.05).

Regression Analyses of Session 1, Initial, Transfer, and Overall Trials for Errors

Multiple regression testing was performed to discover significant predictors of error rates. All regressions used the variables of participant gender, age, scores for speed and memory from the Digit Symbol Substitution (DSS) Test of information processing, semantic memory score taken from the Shipley Vocabulary Test, inferential ability score, and both simple and absolute scores of working memory capacity from the Computation Span Test. Results of testing on raw numbers of errors committed during session 1 trials (trials one through four) resulted in a model ($F(9, 72)=5.62$, $p<.05$) that explained 45% of the variance in the dataset (see Table 12). While participant age, speed on the DSS, memory for symbols on the DSS, averaged NASA-TLX scores and absolute WM score were positively related to higher rates of errors, only age ($b=0.74$, $t=4.33$, $p<.05$), memory for the DSS symbols ($b=0.22$, $t=2.09$, $p<.05$), and averaged NASA-TLX scores ($b=0.24$, $t=2.06$, $p<.05$) were significantly predictive. While participant gender, semantic memory score, inferential ability score, and simple WM score were negatively correlated with the occurrence of errors, only simple WM scores ($b=-0.42$, $t=-2.26$, $p<.05$) were uniquely predictive. Results of testing on raw numbers of errors committed during initial trials (trials one through six) resulted in a model ($F(9, 70)=6.32$, $p<.05$) that explained 48% of the variance in the dataset (see Table 13). While participant age, speed on the DSS, memory for symbols on the DSS, averaged NASA-TLX scores, and absolute WM score were positively correlated to error rate, only age ($b=0.68$, $t=4.04$, $p<.05$), memory for DSS symbols ($b=0,24$, $t=2.25$, $p<.05$), and averaged NASA-TLX scores ($b=0.36$, $t=3.13$, $p<.05$) were significantly predictive. While participant gender, semantic memory score, inferential ability score, and simple WM score were all negatively correlated to error rate, only simple WM scores (*b*=-0.38, *t*=-2.06, *p*<.05) were significantly predictive. Results of testing on raw numbers of errors committed during retention trials (trials five through six) resulted in a model (F(9, 72)=6.37, *p*<.05) that explained 48% of the variance in the dataset (see Table 14). While participant age, speed on the DSS, memory on the DSS, averaged NASA-TLX scores, and absolute WM score were all positively correlated to error rate, only age (*b*=0.57, *t*=3.47, *p*<.05) and averaged NASA-TLX scores (*b*=0.45, *t*=4.13, *p*<.05) were significantly predictive. While participant gender, inferential ability, semantic memory, and simple WM scores were all negatively correlated with committing errors, none of these variables was significantly predictive. Results of testing on raw numbers of errors committed during transfer trials (trials seven through ten) resulted in a model (F(9, 71)=3.73, *p*<.05) that explained 35% of the variance in the dataset (see Table 15). While participant age, speed on the DSS, averaged NASA-TLX scores, and simple WM scores were all positively correlated to error rate, only age (*b*=0.48 *t*=2.59, *p*<.05) and averaged NASA-TLX scores (*b*=0.26, *t*=2.38, *p*<.05) were statistically predictive. While participant gender, memory for symbols on the DSS, semantic memory, inferential ability, and absolute WM scores were negatively associated with errors occurring none of these variables were uniquely predictive. Results of testing on raw numbers of errors committed during trials one through ten resulted in a model (F(9, 68)=5.78, *p*<.05) that explained 47% of the variance in the dataset (see Table 16). While participant age, speed on the DSS, memory for symbols on the DSS, averaged NASA-TLX scores, and absolute WM scores were positively correlated to committing errors, only age (*b*=0.80, *t*=4.62, *p*<.05) was significantly predictive. While participant gender, semantic memory, inferential ability, and simple WM scores were negatively correlated to performing errors none of these variables were uniquely predictive.

Regression Analysis of Near and Far Transfer Errors Committed

Results of testing on raw numbers of near transfer errors committed during transfer trials resulted in a model (F(9, 71)=4.28, *p*<.05) that explained 38% of the variance in the dataset (see Table 17). Positively correlated variables for this model were age, DSS speed, simple WM scores, and averaged NASA-TLX scores. Negatively correlated variables were gender, DSS memory, semantic memory, inferential ability, and absolute WM scores. Of all the positively correlated variables in this model, only age (*b*=0.56, *t*=3.09, *p*<.05) and averaged NASA-TLX scores (*b*=0.24, *t*=2.21, *p*<.05) were significantly predictive.

Results of testing on raw number of far transfer errors committed during transfer trials resulted in a non-significant model (F(9, 71)=1.11, *p*>.05) that explained 14% of the variance in the dataset (see Table 18). Positively correlated variables for this model were age, DSS speed, DSS memory, semantic memory scores, inferential ability, and averaged NASA-TLX scores. Negatively correlated variables for this model included gender, simple WM scores, and absolute WM scores. Of all the negatively correlated variables in this model only gender (*b*=-0.25, *t*=-2.06, *p*<.05) was significantly predictive, where men made more errors than women.

Hierarchical Regression Analysis of Experimental Trials

In order to discover whether age added significant predictive value to regression models for the experimental trials hierarchical regression analysis was performed. Different variables were used during hierarchical regression then during multiple regression due in part to a colinearity analysis and partly due for theoretical reasons. Block 1 for all tests contained the variables of gender, speed on the DSS, inferential ability score, absolute working memory score, number of miniaturized pieces of technology used per day, and the averaged overall NASA-TLX scores for each relevant trial. Block 2 consisted of a participant's age group which was dummy coded as 0 for younger adults or 1 for older adults (Field, 2005). Memory for the symbols on the DSS test were removed due to the use of several other more differentiated and theoretically interesting measures of memory already being incorporated in to regression models. Reviewing bivariate correlations revealed a correlation of at least 0.73 between a participant's semantic memory score and their age group on all models tested, and a correlation of at least 0.85 between simple and absolute working memory scores for all models tested. For this reason measures of semantic memory and simple working memory were removed from all later models. Absolute working memory scores were retained over simple working memory scores due to the more sensitive grading criteria used in the absolute method. The averaged number of pieces of miniaturized technology used per day by participants was included in these models due to theoretical interest. All hierarchical models reported contained no bivariate correlations higher than 0.70 or variables sharing more than 0.60 load on a single factor.

Hierarchical Regression Analysis of Errors for Session 1

Regression results for the number of errors committed during the first four trials found a significant model ($F(6, 72)=4.01$, $p<0.05$) explaining 27% of the variance in the dataset inside Block 1 (see Table 19). Only averaged NASA-TLX scores for trials 1-4 were positively correlated with the number of errors committed, and this predictor was significant ($b=0.38$, $t=2.99$, $p<.05$). While gender, speed on the DSS, inferential ability, absolute WM scores, and number of miniaturized items used per day were all negatively correlated with the number of errors committed, none of these variables was significantly predictive.

The amount of variance explained during Block 2 was 34% ($F(7,72)=4.78$,$p<0.05$), a significant increased in predictive ability ($p<0.05$). While speed on the DSS, NASA-TLX scores, and participant agegroup were positively correlated with the number of errors committed, only TLX scores ($b=0.27$, $t=2.06$, $p<.05$) and agegroup ($b=0.37$, $t=2.67$, $p<.05$) were significantly predictive.

Hierarchical Regression Analysis of Near Transfer Errors

Regression results for the number of near transfer errors committed during trials 7-10 found a significant model ($F(6, 71)=3.21$, $p<0.05$) explaining 23% of the variance in the dataset inside Block 1 (see Table 20). While absolute WM scores and TLX scores were positively correlated to number of near transfer errors, only TLX scores ($b=0.37$, $t=3.24$, $p<.05$) were significantly predictive. While gender, speed on the DSS, inferential ability, and number of miniaturized devices used were all negatively correlated to the number of errors committed, none of these variables was uniquely predictive.

The amount of variance explained during Block 2 was 36% ($F(7,71)=5.15, p<0.05$), a significant increased in predictive ability ($p<0.05$).While speed on the DSS, absolute WM scores, TLX scores, and agegroup were all positively correlated to the number of errors committed, only TLX scores ($b=0.26$, $t=2.38$, $p<.05$) and agegroup (b=, t=, p<.05) were uniquely predictive. While gender, inferential ability,and use of miniaturized technology were all negatively correlated to the criterion, none of these predictors were significant.

Hierarchical Regression Analysis of Far Transfer Errors

Regression results for the number of far transfer errors observed during trials 7-10 found a non-significant model ($F(6, 71)=1.17$, $p>.05$) explaining 10% of the variance in the dataset in Block 1. The model was again non-significant during Block 2 ($F(7, 71)=1.30$, $p>.05$), explaining 13% of the variance in the dataset, with a non-significant increase in the amount of variance explained ($p>.05$).

Hierarchical Regression Analysis for Time on Task During Session 1

Regression results for time on task during the first four trials found a significant model ($F(6, 70)=6.45$, $p<0.05$) explaining 38% of the variance in the dataset (see Table 21). While absolute WM scores and averaged NASA-TLX scores were both positively correlated with time on task, only averaged NASA-TLX scores were significantly predictive ($b=0.29$, $t=2.41$, $p<0.05$). While participant gender, speed on the DSS, inferential ability, and reported number of miniaturized devices used daily were all negatively correlated with time on task, only speed on the DSS ($b=-0.27$, $t=-2.12$, $p<0.05$) and number of miniaturized devices used ($b=-0.29$, $t=-2.80$, $p<0.05$) were significantly predictive.

The amount of variance explained during Block 2 was 46% (F(7,70)=7.57,*p*<0.05), a significant increased in predictive ability (*p*<0.05). While absolute WM scores, TLX scores, and agegroup were all positively correlated to the criterion, only agegroup (b=0.39, t=3.05, p<.05) was uniquely predictive. While gender, DSS speed, inferential ability, and use of miniaturized technology were all negatively correlated to the criterion, none of these predictors were significant.

Hierarchical Regression Analysis for Time on Task During Session 2

Regression results for time on task during trials 7-10 found a significant model (F(6, 69)=4.47, p<.05) explaining 30% of the variance in the dataset in Block 1 (see Table 22). While inferential ability and TLX scores were both positively correlated with time on task, only TLX scores (*b*=0.55, *t*=4.97, *p*<.05) were significantly predictive. While gender, DSS speed, absolute WM scores, and number of miniaturized items used were all negatively correlated with time on task, none of these predictors was uniquely significant.

The amount of variance explained in Block 2 was 32%, a non-significant increase in explanatory power (*p*>.05). Block 2 was significant in predictive ability (F(7, 69)=4.13, *p*<.05). While DSS speed, inferential ability, TLX scores, and agegroup were all positively correlated with time on task, only TLX scores (*b*=0.51, *t*=4.45, *p*<.05) were significantly predictive. While gender, absolute WM scores, and use of miniaturized technology were all negatively correlated to time on task, none of these variables were uniquely predictive.

Table 6

Correlation Matrix for Initial Regression Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	-0.57	-0.5	0.7	-0.06	-0.15	-0.29	0.07	0.44	0.38	0.44	0.32	0.45
2	-0.57	1	0.39	-0.33	0.28	0.33	0.42	0.08	-0.41	-0.28	-0.39	-0.15	-0.34
3	-0.5	0.39	1	-0.27	0.28	0.15	0.17	-0.04	-0.34	-0.39	-0.35	-0.29	-0.39
4	0.7	-0.33	-0.27	1	0.18	-0.01	-0.15	0.03	0.23	0.19	0.27	0.17	0.27
5	-0.06	0.28	0.28	0.18	1	0.13	0.13	-0.14	-0.32	-0.3	-0.31	-0.14	-0.27
6	-0.15	0.33	0.15	-0.01	0.13	1	0.87	-0.01	-0.12	-0.05	-0.08	0.01	-0.04
7	-0.29	0.42	0.17	-0.15	0.13	0.87	1	-0.05	-0.17	-0.11	-0.14	-0.08	-0.13
8	0.07	0.08	-0.04	0.03	-0.14	-0.01	-0.05	1	0.29	0.18	0.27	0.23	0.3
9	0.44	-0.41	-0.34	0.23	-0.32	-0.12	-0.17	0.29	1	0.72	0.97	0.37	0.85
10	0.38	-0.28	-0.39	0.19	-0.3	-0.05	-0.11	0.18	0.72	1	0.86	0.38	0.78
11	0.44	-0.39	-0.35	0.27	-0.31	-0.08	-0.14	0.27	0.97	0.86	1	0.4	0.88

Table 6 Continued

12	0.32	-0.15	-0.29	0.17	-0.14	0.01	-0.08	0.23	0.37	0.38	0.4	1	0.79
13	0.45	-0.34	-0.39	0.27	-0.27	-0.04	-0.13	0.3	0.85	0.78	0.88	0.79	1

Note: 1=Age, 2=Speed on DSS, 3=Memory on DSS, 4=Shipley Vocabulary score, 5=Inferential Ability, 6=Simple WM score, 7= Absolute WM score, 8=Gender, 9=NASA-TLX trials 1-4, 10=NASA-TLX trials5-6, 11=NASA-TLX trials 1-6, 12=NASA-TLX trials 7-10, 13=NASA-TLX trials 1-10.

Table 7

Regression for Time on Task during trials 1-10

Predictor	B	SE	β
Gender	-0.02	15.03	0
Age	1.66*	0.44	0.67
DSS Speed	-0.15	0.66	-0.03
DSS Memory	0.21	3.34	0.01
Semantic Memory	-1.11	2.41	-0.07
Inferential Ability	-3.2	2.4	-0.15
Simple WM score	-6.13	9.23	-0.13

Table 7 Continued

Absolute WM score	0.53	0.77	0.14
Averaged TLX scores	0.6	0.53	0.15
R^2		0.49	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 8

Regression for Time on Task during trials 1-6

Predictor	B	SE	β
Gender	-0.7	17.73	0
Age	2.14*	0.54	0.64
DSS Speed	-0.43	0.82	-0.06
DSS Memory	2.61	3.62	0.08
Semantic Memory	-0.09	2.95	0
Inferential Ability	-5.6	2.9	-0.2
Simple WM score	-13.94	10.87	-0.22

Table 8 Continued

Absolute WM score	1.09	0.94	0.21
Averaged TLX scores	0.8	0.54	0.17
R^2		0.54	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 9

Regression for Time on Task during trials 1-4

Predictor	B	SE	β
Gender	-2.69	24	-0.01
Age	2.58*	0.74	0.6
DSS Speed	-0.97	1.14	-0.12
DSS Memory	1.39	4.8	0.03
Semantic Memory	-2.43	4.11	-0.09
Inferential Ability	-6.43	3.95	-0.18

Table 9 Continued

Simple WM score	-10.41	14.95	-0.13
Absolute WM score	1.15	1.27	0.18
Averaged TLX scores	0.74	0.68	0.13
R^2		0.46	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 10

Regression for Time on Task during trials 5-6

Predictor	B	SE	β
Gender	-11.78	12.02	-0.1
Age	1.03*	0.38	0.45
DSS Speed	0.36	0.6	0.08
DSS Memory	0.38	2.58	0.02
Semantic Memory	0.66	1.99	0.05
Inferential Ability	-0.76	2.1	-0.04

Table 10 Continued

Simple WM score	-11.55	7.67	-0.28
Absolute WM score	0.86	0.67	0.24
Averaged TLX scores	1.33*	0.35	0.24
R^2		0.47	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 11

Regression for Time on task during trials 7-10

Predictor	B	SE	β
Gender	-19	15.24	-0.14
Age	0.62	0.5	0.24
DSS Speed	0.51	0.74	0.1
DSS Memory	-1.68	3.76	-0.06

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 12

Regression for Error Rate during trials 1-4

Predictor	B	SE	β
Gender	-3.8	14.26	-0.16

Table 12 Continued

Age	0.33*	0.08	0.74
DSS Speed	0.02	0.12	0.03
DSS Memory	1.02*	0.51	0.22
Semantic Memory	-0.66	0.42	-0.23
Inferential Ability	-0.59	0.4	-0.16
Simple WM score	-3.42*	1.51	-0.42
Absolute WM score	0.24	0.13	0.36
Averaged TLX scores	0.14*	0.07	0.24
R^2		0.45	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 13

Regression for Error Rate during trials 1-6

Predictor	B	SE	β
Gender	-6.12	3.33	-0.18
Age	0.42*	0.11	0.68

Table 13 Continued

DSS Speed	0.06	0.16	0.05
DSS Memory	1.57*	0.7	0.24
Semantic Memory	-0.7	0.58	-0.18
Inferential Ability	-0.72	0.56	-0.14
Simple WM score	-4.34*	2.11	-0.38
Absolute WM score	0.28	0.18	0.3
Averaged TLX scores	0.33*	0.11	0.36
R^2		0.48	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 14

Regression for Error Rate during trials 5-6

Predictor	B	SE	β
Gender	-2.14	1.15	-0.18
Age	0.13*	0.04	0.57
DSS Speed	0.01	0.06	0.03

Table 14 Continued

DSS Memory	0.49	0.25	0.21
Semantic Memory	-0.16	0.2	-0.12
Inferential Ability	-0.2	0.2	-0.11
Simple WM score	-1.04	0.75	-0.26
Absolute WM score	0.07	0.07	0.21
Averaged TLX scores	0.14*	0.03	0.45
R^2		0.48	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 15

Regression for Error Rate during trials 7-10

Predictor	B	SE	β
Gender	-5.35	3.39	-0.17
Age	0.29*	0.11	0.48
DSS Speed	0.05	0.17	0.04
DSS Memory	-0.36	0.78	-0.06

Table 15 Continued

Semantic Memory	-0.44	0.61	-0.12
Inferential Ability	-0.46	0.59	-0.09
Simple WM score	2.96	2.28	0.28
Absolute WM score	-0.13	0.19	-0.15
Averaged TLX scores	0.20*	0.08	0.26
R^2		0.35	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 16

Regression for Error Rate during trials 1-10

Predictor	B	SE	β
Gender	-8.31	5.34	-0.16
Age	0.77*	0.17	0.8
DSS Speed	0.06	0.25	0.03
DSS Memory	1.23	1.17	0.12

Table 16 Continued

Semantic Memory	-1.26	0.91	-0.21
Inferential Ability	-1.54	0.89	-0.19
Simple WM score	-1.08	3.42	-0.06
Absolute WM score	0.16	0.29	0.11
Averaged TLX scores	0.25	0.18	0.16
R^2		0.47	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 17

Regression for Near Transfer Errors Committed during trials 7-10

Predictor	B	SE	β
Gender	-3.27	2.7	-0.13
Age	0.27*	0.09	0.56
DSS Speed	0.02	0.13	0.02
DSS Memory	-0.27	0.63	-0.05

Table 17 Continued

Semantic Memory	-0.45	0.49	-0.15
Inferential Ability	-0.43	0.47	-0.1
Simple WM score	1.86	1.82	0.21
Absolute WM score	-0.05	0.16	-0.08
Averaged TLX scores	0.15*	0.07	0.24
R^2		0.38	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 18

Regression for Far Transfer Errors Committed during trials 7-10

Predictor	B	SE	β
Gender	-1.78*	0.87	-0.25
Age	0.03	0.03	0.25
DSS Speed	0.04	0.04	0.14
DSS Memory	0.19	0.2	0.14

Table 18 Continued

Semantic Memory	0.04	0.16	0.05
Inferential Ability	0.06	0.15	0.05
Simple WM score	-0.01	0.58	0
Absolute WM score	-0.02	0.05	-0.08
Averaged TLX scores	0.03	0.02	0.17
R^2		0.14	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only

Table 19

Regression for Errors Committed Trials 1-4

	Block 1			Block 2		
Predictor	B	SE	β	B	SE	β
Gender	-4.73	2.75	-0.2	-4.6	2.63	-0.2
DSS Speed	-0.09	0.12	-0.11	0.04	0.12	0.04
Inferential Ability	-0.19	0.42	-0.05	-0.53	0.42	-0.14
Absolute WM score	-0.06	0.08	-0.09	-0.04	0.08	-0.05

Table 19 Continued

Technology	-1.72	1.06	-0.18	-0.81	1.07	-0.09
Averaged TLX scores	0.22	0.08	0.38	0.16	0.08	0.27
Agegroup				8.36	3.13	0.37
R^2		0.27			0.34	
ΔR^2					0.07	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only, Technology= measure of average pieces of miniaturized technology used by participants per day.

Table 20

Regression for Errors Committed Trials 7-10

		Block 1			Block 2	
Predictor	B	SE	β	B	SE	β
Gender	-3.2	3.02	-0.12	-3.27	2.77	-0.13
DSS Speed	-0.21	0.13	-0.22	0.01	0.13	0.01
Inferential Ability	-0.14	0.47	-0.04	-0.56	0.45	-0.14
Absolute WM score	0.04	0.09	0.05	0.07	0.08	0.1

Table 20 Continued

Technology	0.1	1.2	-0.14	-0.19	1.16	-0.02
Averaged TLX scores	0.22	0.07	0.37	0.16	0.07	0.26
Agegroup				12.09	3.33	0.49
R^2		0.23			0.36	
ΔR^2					0.13	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only, Technology= measure of average pieces of miniaturized technology used by participants per day.

Table 21

Regression for Time on Task Trials 1-4

	Block 1			Block 2		
Predictor	B	SE	β	B	SE	β
Gender	-11.22	25.53	-0.05	-11.85	24.02	-0.05
DSS Speed	-2.3	1.09	-0.27	-0.95	1.12	-0.11
Inferential Ability	-2.72	3.84	-0.08	-5.91	3.76	-0.16
Absolute WM score	0.02	0.74	0	0.2	0.7	0.03
Averaged TLX scores	1.66	0.69	0.29	0.99	0.68	0.17

Table 21 Continued

Technology	-26.71	9.61	-0.29	-16.92	9.6	-0.18
Agegroup				86.62	28.44	0.39
R^2		0.38			0.46	
ΔR^2					0.08	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only, Technology= measure of average pieces of miniaturized technology used by participants per day.

Table 22

Regression for Time on Task Trials 7-10

	Block 1			Block 2		
Predictor	B	SE	β	B	SE	β
Gender	-21.22	15.47	-0.16	-21.32	15.38	-0.16
DSS Speed	-0.02	0.65	0	0.43	0.73	0.09
Inferential Ability	1.77	2.52	0.08	1.01	2.56	0.05
Absolute WM score	-0.15	0.45	-0.04	-0.08	0.45	-0.02
Technology	-7.71	6.08	-0.14	-5.16	6.34	-0.09

Table 22 Continued

Averaged TLX scores	1.76	0.35	0.55	1.63	0.37	0.51
Agegroup				24.42	18.34	0.19
R^2		0.3			0.32	
ΔR^2					0.02	

Note. *p<.05, NASA-TLX scores used are for appropriate trials only, Technology= measure of average pieces of miniaturized technology used by participants per day.

Chapter 4

Discussion

The current set of results is consistent with previous findings from the aging and technology literature. For instance, these research findings are in line with previous research that suggests that older adults require more time on task and perform more errors than younger adults during initial training and retention testing for technology-related tasks such as word processing (Czaja, Hammond, Blascovich, &Swede, 1989) software, or smaller devices such as PDAs (Mayhorn, Lanzolla, Wogalter, & Watson, 2005) or blood glucometers (Mykityshyn, Fisk, & Rogers, 2002). Moreover, the current results also support findings that older adults perform more slowly and less accurately than younger adults when transferring between pieces of larger equipment such as ATMs (Jamieson, Cabrera, Mead, & Rousseau, 1995; Jamieson & Rogers, 2000).

Within the context of medical device training and aging, MANOVA results supported the hypotheses that YAs take less time to complete CSTs, and commit fewer training and near transfer errors. These results suggest that future meters should be designed with more input concerning the needs of OAs to minimize differences due to age. Results did not support the hypothesis that YAs would commit significantly fewer far transfer errors. Failure to support this hypothesis may be due to the nature of the possible far transfer errors in the current research. The Nova MaxLink is capable of sending glucose test results to an insulin pump. This feature can be avoided by marking a test as a CST or by turning the feature off completely. Participants transferring to the Nova MaxLink were likely equally unaware this feature existed due to lack of feedback from the device. Participants successfully starting a test with the OneTouch UltraMini were confronted with a prompt of “C” followed by a number between 1 and 50. At this point participants in either age group were probably capable of inferring a match was needed between the code number in the meter and on the vial of testing strips. Findings concerning the order of meter use were unexpected. While meters differed in the number of possible errors in a trial, all errors were analyzed in terms of percentages. Examination of the trends suggests that fewer far transfer errors may have occurred when using the OneTouch UltraMini than the Nova MaxLink due to the visibility of features associated with possible far transfer errors. Future research should examine how to best make users aware of possible far transfer issues in the absence of instructional materials. Future investigation should also examine how best to prevent these errors once users are aware of them. Examination of the trends for near transfer errors suggests that fewer near transfer errors occurred when using the Nova MaxLink than the OneTouch UltraMini due to the design of testing materials, flexibility of testing procedure, or the presentation of information in the instruction manuals. Future research might examine the physical design of testing equipment and interface prompts on usability. Additionally, research should identify what elements of initial training influence rates of transfer errors.

The use of a repeated measures ANOVA to investigate hypothesis 3 produced evidence that regardless of age group, participants committed a significantly higher rate of far transfer errors than near transfer errors. This finding suggests far transfer errors are a class of error not preventable simply by possessing higher scores on cognitive tests. Rather, far transfer errors may occur due to being hidden via a lack of feedback, or due to a participant's inability to troubleshoot novel features when they are aware of them. To improve overall usability, future glucometers and accompanying instructions should make novel features more visible as well as understandable.

Results of multiple regression testing showed several trends. Of the variables tested, speed on the DSS, semantic memory score, inferential ability score, and absolute measures of WM capacity were never significantly predictive regardless of which variable was used or which trials were focused on. Gender was only significantly predictive when examining far transfer errors within a model that was not significantly predictive. Participant age was a significant predictor of time on task or errors committed in ten out of twelve models calculated, and it was always positively correlated with the criterion. Memory for the symbols on the DSS was only predictive for two models examining error rate for trials one through four and trials one through six, where it was positively correlated to the number of errors committed. The simple score for WM capacity was only predictive for two error models, examining trials one through four and trials one through six yet it was always negatively associated with the criterion. Averaged NASA-TLX scores were significant predictors of errors committed or time on task in seven out of twelve models calculated, and was always positively correlated to the criterion. Taken together, these results clearly demonstrate the usefulness of participant age and NASA-TLX scores as predictors for time on task and number of errors committed when performing control solution tests with a blood glucometer. Also clearly demonstrated was the lack of a gender effect found in this sample as evidenced by the single non-predictive model that contained gender as a significant predictor of the criterion.

While different age groups may have differential needs from designers to prevent errors (refer to results and discussion of MANOVA testing), results here suggest no special needs based on gender. The lack of significance for the measure of semantic memory suggests that glucometer instructions may be written at an appropriate level of complexity for the sample tested. While word and sentence length or reading level may have been appropriate, future investigation of how to reduce confusion related to jargon should be performed. Participants were able to differentiate the parts of speech in the instructions but had trouble with the specialized vocabulary used to discuss the testing procedure. For example, the area of the testing strip that should be inserted into the OneTouch UltraMini has “contact bars” which are inserted into the “test port”. The area of the testing strip to which control solution is applied is covered with a clear plastic coating to allow the user to see if an adequate amount of solution has been applied to the strip. Rather than apply solution to the top of the strip, users are instructed to use the strip “edge to apply sample” using the “confirmation window”. Perhaps more interesting is the lack of significance of speed on the DSS and absolute measures of WM capacity. Here the data suggest that control solution testing may be a task that is too complicated for a simple test like the DSS to be predictive, even for time on task. The DSS is a task requiring participants to use a legend to match a sheet of numbers with appropriate symbols. Participants have little trouble understanding how to perform this task, rarely ask questions during the task, and the legend easily fits on the same page as the test stimuli. By contrast, during control solution testing, many participants were unclear about how to perform the task even after reading the instructions. Participants routinely had questions for the experimenter during trials, and frustration may have occurred when the experimenter was unable to provide them with answers. Additionally the instructions for completing tests were spaced out over multiple pages of material, with some participants referring to the instructions for almost every step and others avoiding reading the instructions whenever possible. It is possible that speed on the DSS would be predictive after participants had more experience with control solution testing, or after participants were trained to the point that they did not need to reference the instruction manual.

The finding that absolute WM scores were not predictive of time on task or errors committed while simple WM scores were (for trials 1-4 and 1-6) suggests that WM capacity may not be as central to control solution testing as the principal experimenter first thought. The absolute method of grading for the Computation Span Test may be unnecessarily complicated for assessing performance during control solution tests. While grading for the simple span requires participants to recall two of three trials for a given number of stimuli verbatim to receive credit, the absolute method grades trials on a trial by trial basis. While the simple span method stops scoring as soon as participants fail to attain 2/3 recall, the absolute method scores the entire test of 21 separate trials. While many participants are unable to recall many complete trials after they fail to recall 2/3 trials for a given number of stimuli, the absolute grading method would allow two participants recalling differing numbers of stimuli to obtain similar scores. Future investigation using different weighting methods for the absolute score of the Computation Span should be performed. Findings that simple WM scores were only predictive for errors committed in trials 1-6 suggest that when first using a blood glucometer participants may benefit from higher WM scores. The benefit of higher WM scores may disappear during trials 7-10 due to having some information about blood glucometers transferred to long term memory. Alternatively, as trials 7-10 used a novel glucometer with no instructions, it is possible that WM would not be as useful for preventing errors as participants had little useful information to retrieve or hold in working memory. Future analysis of the relationship between WM scores and the types of errors committed should be performed. It may be that higher WM scores were beneficial for preventing errors in trials 1-6 by allowing participants to mentally keep steps for testing in the correct order. Control solution testing with the chosen meters was not a flexible procedure, meaning that the order of steps was critical for completely a trial correctly. The finding that a participant's inferential ability was never predictive for time on task or errors committed suggests that participants may not have had enough information given to them during the experimental trials to allow higher inferential ability to be an asset.

That memory for the symbols on the DSS was positively correlated to the number of errors committed in trials 1-4 and 1-6 was surprising. A possible explanation for this finding is that participants who were better at retaining information from the instructions may have suffered from information overload. Lacking familiarity with glucometers and testing procedure, these participants may have forgotten the proper sequence of steps in memory, or may have erroneously combined information from several steps together to create completely new (and incorrect) actions.

NASA-TLX scores were always positively correlated to the criterion during regression testing. Along with participant age, this variable was one of the most consistent predictors of participant performance. This finding provides evidence that increased subjective workload can be taken as an indicator of impaired performance. A possible explanation for this result is that increased frustration and demands of mental or physical resources led to situations where participants were unable to perform optimally, resulting in errors or slower performance. ANOVA testing found that older adults' composite NASA-TLX ratings as well as ratings for each subscale except temporal demand were significantly higher than those of younger adults (see Participants section above). This finding is mostly in agreement with previous work (Mykityshyn, Fisk, & Rogers, 2002) which determined older adults reported significantly greater workload on all aspects of the NASA-TLX than younger adults when using a blood glucometer. Previous work (Mykityshyn, Fisk, & Rogers, 2002) also found that older adults using a manual versus a properly designed instructional video experienced increased feeling of workload, suggesting that glucometers released in the future could improve usability by changing their format of instruction. Failure to find significant differences in temporal demand scores based on age was interesting as this was in conflict with previous findings (Mykityshyn, Fisk, & Rogers, 2002). This finding was also surprising as older adults took significantly longer to complete control solution testing than younger adults, and as older adults were more likely to experience a glucometer turning off unexpectedly in the middle of a test due to a power saving feature. Indeed, investigators saw several older adult participants restart a control solution test multiple times due to being unable to complete a test within the preprogrammed 2 minute time limit. While older adults may not have experienced feeling rushed or concluded that faster performance was necessary when this occurred, it likely contributed to their higher rate of error when compared to younger adults. Future models of glucometers could improve usability by removing the time limit before glucometers turn off, by adding a prompt that users can respond to extend the time allowed for a test, or by emphasizing the importance of speed during control solution testing in future documentation. Future research should investigate exactly what contributes to the perception of temporal demand amongst older adults.

Taken as a whole, results of multiple regression testing suggest that participants may not have had enough information explicitly given to them to allow differences in the abilities measured by the cognitive battery to influence participant ability after transferring to a new glucometer. Alternatively, it could be that the abilities measured by the cognitive battery are not useful for predicting behavior with a novel piece of technology. The predictive usefulness of the NASA-TLX suggests future research should examine participant attitudes toward technology in addition to observational and cognitive testing data. Whether NASA-TLX results are reflective of an unmeasured cognitive deficiency (or unanalyzed interaction between abilities) or the importance of user attitude is uncertain at this point.

Results of hierarchical regression testing found several trends. The variables of gender, inferential ability, and absolute WM scores were never significantly predictive in the significant models run regardless of which Block was examined. While gender was predictive when examining far transfer errors inside Block 2, this model was non-significant. While DSS speed and use of miniaturized technology were found to be significantly predictive, both variables were only significant for estimating time on task for trials 1-4 before accounting for the variance explained by age category. The only variables with repeated predictive value were relevant NASA-TLX scores and participant agegroup. Indeed, NASA-TLX scores were significantly in both Block 1 and Block 2 numerous times. Accounting for the variance explained by a participant's agegroup also resulted in significant increases in explanatory ability. Some variables remained consistently correlated to time on task and number of errors across tests. These included gender, use of miniaturized technology, NASA-TLX scores, and agegroup. Other variables were inconsistent, being positively correlated to the criterion during some blocks and negatively correlated in others. DSS speed, inferential ability, and absolute WM scores were all inconsistent predictors. Overall, the two most useful variables for predicting participant error count and time on task were NASA-TLX scores and agegroups, as both variables were revealed a consistent direction of correlation and were repeatedly significant at predicting the criterion. The failure of DSS speed to be predictive after removing memory for symbols on the test suggests that the DSS is not adequate for predicting performance on tasks such as CST.

Participants needed to refer to glucometer instructions numerous times during testing, whereas the DSS has participants perform a simple task once. DSS speed may be more successful at predicting the performance of expert testers who can perform testing as one smooth procedure with no need to refer to instructions. The failure of absolute WM scores to be predictive of either criterion suggests working memory as measured by the Computation Span may not be appropriate for predicting performance during novice CST. The repeated referral to instructions along with the availability of instructions during trials 1-4 may have negated the advantage of increased WM capacity, while the lack of usable information during trials 7-10 may have negated the advantage of increased WM capacity during Session 2. The failure of use of miniaturized technology to be repeatedly predictive of either criterion suggests something other than familiarity with small devices is responsible for differences in participant performance. Future investigation may reveal familiarity with specific kinds of technology such as medical devices or devices with small delicate parts may be more predictive of performance during CST. It is possible that familiarity with devices with a similar function such as other home medical devices would be more predictive of performance than familiarity with miniaturized devices.

While these findings are potentially interesting, methodological shortcomings must be considered. Diabetics may receive live training with corrective feedback when learning to use their first glucometer. However diabetics are not guaranteed live training with their initial meter or any other meter. OAs in this study were individuals capable of coming to a laboratory setting. Indeed, self-selection was an issue for this study in that multiple prospective older adults declined to participate upon learning the requirement to park on campus and walk to the laboratory, even after being told parking would be free and that experimenters would be available as escorts from the parking facilities to the laboratory. While diabetic individuals may have greater experience with CSTs, diabetes is associated with increased risk of several cognitive and motor impairments (Christman, Vannorsdall, Pearlson, Hill-Briggs, & Schretlen, 2010; Lin, Northam, Rankins, Werther, & Cameron, 2010; McNally, Delamater, Rohan, Drotar, & Pendley, 2010). Future research should incorporate diabetics and those having experience with meters such as diabetes educators to verify that these results generalize. ANOVA testing on the results of demographic and cognitive battery testing revealed several significant differences between participants based on age (YA using more pieces of handheld miniaturized technology on average per day, scoring higher in both measures from the DSS, reviewing instructional materials for less time, scoring higher when using the absolute method of grading for the Computation Span test, scoring lower when using the simple method of grading for the Computation Span, and scoring lower than OAs on the Shipley Vocabulary test). While these results may be expected based on previous studies of age-related trends (Howard & Howard, 1997; Salthouse & Babcock, 1991; Scialfa, Ho, & Laberge, 2004; Mayhorn, Stronge, McLaughlin, & Rogers, 2004; McLaughlin, Rogers, & Fisk, 2004) future research should attempt to construct age groups with more similar ability scores to establish a true effect of age rather than ability differences.

Even in the face of methodological limitations the results of this research warrant further consideration amongst researchers interested in aging, transfer of training, and medical human factors. It has been mentioned before that adults aged 65 and over are more at risk for developing diabetes than the rest of the population and that this demographic makes up over a third of those with diabetes in the U. S. population (Centers for Disease Control and Prevention, 2007). Results from the present research show that this demographic performs significantly more initial and near transfer errors than their younger counterparts while requiring significantly more time to complete control testing procedures. That participants most likely to use blood glucometers in the future showed significantly poorer performance than any other group suggests future glucometers should be designed with a greater emphasis on the needs of older adults. While the myth that older adults are resistant to using new technology is persistent (Morrell, Mayhorn, & Echt, 2004; Rogers, Mayhorn, & Fisk, 2004), the domain of device design for managing diabetes is one where it cannot be used as an excuse for less than optimal usability.

References

Anderson, D. I., Magill, R. A., Seklya, H. (2001). Motor learning as a function of KR schedule and characteristics of task-intrinsic feedback. *Journal of Motor Behavior*, 33(1), 59-66.

Bergenstal, R., Pearson, J., Cembrowski, G. S., Bina, D., Davidson, J., & List, S. (2000). Identifying variables associated with inaccurate self-monitoring of blood glucose: proposed guidelines to improve accuracy. *The Diabetes Educator,* 26(6), 981-989.

Blaiwes, A. S., Puig, J. A., & Regan, J. J. (1973). Transfer of training and the measurement of training effectiveness. In R.W. Swezey & D.H. Andrews (Ed.), *Readings in training and simulation: A 30-year perspective* (pp. 2-12). Santa Monica, CA: Human Factors and Ergonomics Society.

Bogner, M. S. (2004). Series foreward. In M. S. Bogner (Ed.), *Misadventures in health care: Inside stories* (pp. ix-x). Mahwah, NJ: Institute for the Study of Human Error, LLC.

Centers for Disease Control and Prevention. (2007). Crude and age-adjusted prevalence of diagnosed diabetes per 100 population, United States, 1980-2005. Retrieved March 12, 2008 from http://www.cdc.gov/diabetes/statistics/prev/national/figage.htm

Centers for Disease Control and Prevention. (2007). Number (in millions) of persons with diagnosed diabetes, United States, 1980-2005. Retrieved March 12, 2008, from http://www.cdc.gov/diabetes/statistics/prev/national/figpersons.htm

Chandra, D. B., Savoldelli, G. L., Joo, H. S., Weiss, I. D., & Naik, V. N. (2008). Fiberoptic oral intubation: The effect of model fidelity on training for transfer to patient care. *Anesthesiology*, 109, 1007-1013.

Charness, N., Schumann, C E., & Boritz, G. M. (1992). Training older adults in word processing: Effects of age, training technique, and computer anxiety. *International Journal of Technology and Aging*, 5(1), 79-106.

Christman, A. L., Vannorsdall, T. D., Pearlson, G. D., Hill-Briggs, F., & Schretlen, D. J (2010). Cranial volume, mild cognitive deficits, and functional limitations associated with diabetes in a community sample. *Archives of Clinical Neuropsychology*, 25, 49-59. DOI:10.1093/arclin/acp091

Czaja, S. J., Hammond, K., Blascovich, J. J., & Swede, H. (1989). Age related differences in learning to use a text-editing system. *Behaviour and Information Technology*, 8(4), 309-319.

DeFreitas, C. M., Dunaway, L. A., & Torres, I. J. (2009). Preferential semantic fluency impairment is related to hallucinations, but not formal thought disorder. *Schizophrenia Research*, 107, 307-312.

Ellis, L. B. M., Joo, H., & Gross, C. R. (1991). Use of a computer-based health risk appraisal by older adults. *The Journal of Family Practice*, 33(4), 390-394.

Field, A. (2005). Chapter 14 Multivariate Analysis of Variance (MANOVA). In A. Field (Ed.) *Discovering Statistics Using SPSS*. (571-618). London, England: SAGE Publications Ltd.

Gist, M., Rosen, B., & Schwoerer, C. (1988). The influence of training method and traineage on the acquisition of computer skills. *Personnel Psychology*, 41, 255-265.

Gordon, S. E., & Gill, R. T. (1997). Chapter 13: Cognitive task analysis. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp.131-140). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Grady, J. L., Kehrer, R. G., Trsuty, C. E., Entin, E. B., Entin, E. E., & Brunye, T. T. (2008). Learning nursing procedures: The influence of simulator fidelity and student gender on teaching effectiveness. *Journal of Nursing Education*, 47(9), 403-408.

Hart, S. G. (2006). *NASA-Task load index (NASA-TLX); 20 years later*. Retrieved June 1, 2009, from http://humansystems.arc.nasa.gov/groups/TLX/tlxpublications.html

Hart S. G., & Staveland, L. E. (n.d.). *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.* Retrieved June 1, 2009, from http://humansystems.arc.nasa.gov/groups/TLX/downloads/NASA-TLXChapter.pdf

Howard, J. H., & Howard, D. V. (1997). Chapter 2: Learning and memory. In A. D. Fisk & W. A. Rogers (Eds.), *Handbook of Human Factors and the older* adult (pp.7-26). London, United Kingdom: Academic Press Limited.

Howell, W. C. (1997). Chapter 1: Foreward, perspectives, and prospectives. In A. D. Fisk & W. A. Rogers (Eds.), *Handbook of Human Factors and the older* adult (pp.1-6). London, United Kingdom: Academic Press Limited.

Jamieson, B. A., Cabrera, E. F., Mead, S. E., and Rousseau, G. K. (1995). Training new technology: Automatic teller machines and older adults. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, 153-157.

Jamieson, B. A., Rogers, W. A. (2000). Age-related effects of blocked and random practice schedules on learning a new technology. *Journal of Gerontology: Psychological Sciences*, 55B(6), 343-353.

Jones, H. (1994). Common errors associated with capillary blood glucose monitoring. *British Journal of Nursing*, 3(19), 1021-1022.

Kaye, R., and Chenault, V. M. (2002). An overview of human factors engineering at CDRH in the safety and effectiveness of blood glucose meters. *Diabetes Technology & Therapeutics*, 4(2), 247-251.

Kline, D. W., & Scialfa, C. T. (1997). Sensory and perceptual functioning: Basic research and human factors implications. In A. D. Fisk & W. A. Rogers (Eds.), *Handbook of Human Factors and the older* adult (pp.27-54). London, United Kingdom: Academic Press Limited.

Lamata, P., Gómez, E. J., Sánchez-Margallo, F. M., Lamata, F., del Pozo, F., Usón, J. (2006). Tissue consistency perception in laparoscopy to define the level of fidelity in virtual reality simulation. *Surgical Endoscopy*, 20, 1368-1375.

Lin, A., Northam, E. A., Rankins, D., Werther, G. A., & Cameron, F. J. (2010). Neuropsychological profiles of young people with type 1 diabetes 12 yr after disease onset. *Pediatric Diabetes*, 11, 235-243. DOI:*10.1111/j.1399-5448.2009.00588.x*

Linden, J. V. (2004). The trouble with blood is it all looks the same: Transfusion errors. In M.S. Bogner (Ed.), *Misadventures in health care Inside stories* (pp.13-25). Mahwah, NJ: Institute for the Study of Human Error, LLC.

Lintern, G. (1991). An informational perspective on skill transfer in human-machine systems. In R.W. Swezey & D.H. Andrews (Ed.), *Readings in training and simulation: A 30-year perspective* (pp. 56-71). Santa Monica, CA: Human Factors and Ergonomics Society.

Mayhorn, C. B., Lanzolla, V. R., Wogalter, M. S., & Watson, A. M. (2005). Personal digital assistants (PDAs) as medication reminding tools: Exploring age differences in usabilty. *Gerotechnology*, 4(3), 128-140.

Mayhorn, C. B., Rogers, W. A., & Fisk, A. D. (2004). Designing technology based on cognitive aging principles. In D. C. Burdick & S. Kwon (Ed.), *Gerotechnology: Research and practice in technology and aging* (pp.42-53). New York, NY: Springer Publishing Company, Inc.

Mayhorn, C. B., Stronge, A. J., McLaughlin, A. C., & Rogers, W. A. (2004). Older adults, computer training, and the systems approach: A formula for success. *Educational Gerontology*, 30, 185-203.

McKinlay, J., & Marceau, L. (2000). U.S. public health and the 21st century: Diabetes mellitus. *Lancet*, 356, 757-761.

McLaughlin, A. C., Rogers, W. A., & Fisk, A. D. (2009). Using direct and indirect input devices: Attention demands and age-related differences. *ACM Transactions on Computer-Human Interaction*, 16(1), 2:1-2:15.

McNally, K., Delamater, A., Rohan, J., Drotar, D., & Pendley, J. S. (2010). Executive functioning, treatment adherence, and glycemic control in children with type 1 diabetes. *Diabetes Care*, 33(6), 1159-1162. DOI:10.2337/dc09-2116

Morrell, R. W., Mayhorn, C. B., & Echt, K. V. (2004). Why older adults use or do not use the Internet. In D. C. Burdick & S. Kwon (Ed.), *Gerotechnology: Research and practice in technology and aging* (pp.71-85). New York, NY: Springer Publishing Company, Inc.

Mykityshyn, A. L., Fisk, A. D., & Rogers, W. A. (2002). Learning to use a home medical device: Mediating age-related differences with training. *Human Factors*, 44(3), 354-364.

National Aeronautics and Space Administration. (n. d.) NASA Task Load Index. Retrieved June 1, 2009 from http://humansystems.arc.nasa.gov/groups/TLX/downloads/TLXScale.pdf

Nichols, T. A., Rogers, W. A., and Fisk, A. D. (2003). Do you know how old your participants are? *Ergonomics in Design*, 11(3), 22-26.

O'Connor, A., Schwaitzberg, S. D., & Cao, C. G. L. (2008). How much feedback is necessary for learning to suture? *Surgical Endoscopy*, 22, 1614-1619.

Raine, C. H. (2003). Self-monitored blood glucose: A common pitfall. *Endocrine Practice,* 9(2), 137-139.

Rogers, W. A., Mayhorn, C. B., & Fisk, A. D. (2004). Technology in everyday life for older adults. In D. C. Burdick & S. Kwon (Ed.), *Gerotechnology: Research and practice in technology and aging* (pp.3-17). New York, NY: Springer Publishing Company, Inc.

Rogers, W. A., Mykityshyn, A. L., Campbell, R. H., & Fisk, A. D. (2001). Analysis of a “simple” medical device. *Ergonomics in Design*, 6-14.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, 27, 763-776.

Scialfa, C. T., Ho, G., & Laberge, J. (2004). Perceptual Aspects of Gerotechnology. In D. C. Burdick & S. Kwon (Ed.), *Gerotechnology: Research and practice in technology and aging* (pp.18-41). New York, NY: Springer Publishing Company, Inc.

Shakeri, S., & Funk, K. (2007). A comparison of human and near-optimal task management behavior. *Human Factors*, 49(3), 400-416.

Sharit, J., Czaja, S. J., Nair, S., & Lee, C. C. (2003). Effects of age, speech rate, and environmental support in using telephone voice menu systems. *Human Factors*, 45(2), 234-251.

Sheperd, A. (2001). *Hierarchical Task Analysis*. New York: Taylor & Francis.

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

Simon, C. W., & Roscoe, S. N. (1984). Application of a multifactor approach to transfer of training research. In R.W. Swezey & D.H. Andrews (Ed.), *Readings in training and simulation: A 30-year perspective* (pp. 13-34). Santa Monica, CA: Human Factors and Ergonomics Society.

Spears, W. D. (1985). Measurement of learning and transfer through curve fitting. In R.W. Swezey & D.H. Andrews (Ed.), *Readings in training and simulation: A 30-year perspective* (pp. 35-50). Santa Monica, CA: Human Factors and Ergonomics Society.

Sterns, H. L., & Alexander, R. A. (1977). Chapter 5 cohort, age, and time of measurement: Biomorphic considerations. In N. Datan & H. W. Reese (Ed.), *Life-span developmental psychology: Dialectical perspectives on experimental research* (pp.105-120). London, UK: Academic Press.

Todorov, E., Shadmehr, R., & Bizzi, E. (1997). Augmented feedback in a virtual environment accelerates learning of a difficult motor task. *Journal of Motor Behavior*, 29(2), 147-159.

UsabilityNet. (2006). Task analysis. Retrieved June 12, 2009 from http://www.usabilitynet.org/tools/taskanalysis.htm

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale III. (3rd Ed.).* San Antonio, TX: The Psychological Corporation.