2007

Effects of input device on different age groups performing goal-oriented tasks

James Frakes Kravitz
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation

Kravitz, James Frakes, "Effects of input device on different age groups performing goal-oriented tasks" (2007). *Master's Theses*. 3396.
DOI: <https://doi.org/10.31979/etd.utej-ua2r>
https://scholarworks.sjsu.edu/etd_theses/3396

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

EFFECTS OF INPUT DEVICE ON DIFFERENT AGE GROUPS PERFORMING
GOAL-ORIENTED TASKS

A Thesis

Presented to

The Faculty of the Department of Industrial and Systems Engineering

Program in Human Factors and Ergonomics

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

James Frakes Kravitz

May 2007

UMI Number: 1445244

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 1445244

Copyright 2007 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© 2007

James Frakes Kravitz

ALL RIGHTS RESERVED

APPROVED FOR THE DEPARTMENT OF
INDUSTRIAL AND SYSTEMS ENGINEERING

Dr. Kevin Jordan

Dr. Mark Van Selst

Arnold M. Lund, Ph.D., Microsoft Corporation

APPROVED FOR THE UNIVERSITY

ABSTRACT
EFFECTS OF INPUT DEVICE ON DIFFERENT AGE GROUPS PERFORMING
GOAL-ORIENTED TASKS

by James Frakes Kravitz

Convertible tablet PCs can use a pen or a mouse for input. The pen is better suited than the mouse for some tasks because of its interaction properties, and research has shown it may ameliorate age-related decrements in performance. This study compared the pen and mouse on a series of realistic tasks for older (55-69) and younger (18-30) adults. Precision tasks were better served by the mouse, while ballistic tasks with strong analogs to real-world actions were served equally well by the pen or the mouse. Older adults were slower than younger adults on both devices, but contrary to the research hypothesis, no benefits were observed specifically for older adults with the pen. This study reinforces findings regarding the importance of task demands when selecting input devices. Younger adults seemed more willing than older adults to embrace the pen. Design recommendations are provided for future pen computing systems.

Acknowledgements

I owe a debt of gratitude to many individuals for their contributions to the successful completion of this thesis. I would first like to thank my thesis advisor, Dr. Kevin Jordan, for patiently reviewing my many ideas and providing guidance, thoughtful feedback, and friendly encouragement along the way, and for keeping me focused once the thesis path was chosen. I also must thank my thesis committee members, Drs. Arnie Lund and Mark van Selst, for their time, enthusiasm, and valuable contributions. Additional thanks are due to my mentor, employer, and friend Dr. Tony Andre for bouncing ideas, contributing materials and facilities to the study, and accommodating my need for time as the defense drew near.

Many friends and colleagues were also a great source of support and inspiration. Thanks to José-Luis Garcia-Chico and Jim Beno who were part of the thesis salon where our ideas percolated, and to Eliana Martella and Adriana Manago for listening to me and being supportive friends. Thanks are also due to my friend Sebastian Uijtdehaage for helping me keep it all in perspective.

Finally, none of this work would have been possible without my family – my brother and sister-in-law, David Kravitz and Majie Zeller, and my parents, Edward and Kathryn Kravitz. Thanks to them for reading drafts, providing inspiration, support (financially, emotionally, and intellectually), and love.

Table of Contents

Introduction.....	1
Research problem	1
The mouse as an input device	1
Diversity of tasks performed on today's PCs.....	2
User capabilities in an aging workforce	3
Summary	3
New input technology options: the tablet PC	4
Tablet PC description and uses	4
Similarities to other technologies	6
Input device properties – dimensions of interaction	9
Direct vs. indirect devices.....	9
Absolute vs. relative devices.....	12
Physical manipulation.....	14
Summary of device properties.....	16
Task effects	17
Elemental task differences	17
Goal-oriented task differences	18
Mediating the adverse effects of aging.....	21
Summary	24
User preferences.....	24
The present study.....	27

Summary	27
Hypotheses	29
Method.....	31
Design	31
Independent variables	31
Dependent variables.....	31
Experimental order	31
Participants.....	32
Materials	34
Experimental instruments	34
Data collection instruments.....	35
Procedure	38
Task descriptions.....	41
Excel dragging task	41
PowerPoint dragging task	43
PowerPoint manipulation task.....	45
Tracing task.....	47
Cross-referencing “form filling” task	49
Cross-referencing “document editing” task	54
Results	59
Data preparation – overview	59
Error analysis.....	59

Testing for order effects.....	60
Effect size.....	60
Excel dragging task	61
Error analysis and data preparation	61
Descriptive statistics	62
Inferential statistics.....	66
PowerPoint dragging task	70
Error analysis and data preparation	70
Descriptive statistics	71
Inferential statistics.....	75
PowerPoint manipulation task	81
Error analysis and data preparation	81
Descriptive statistics	82
Inferential statistics.....	86
Tracing task.....	92
Error analysis and data preparation	92
Descriptive statistics	93
Inferential statistics.....	97
Cross-referencing “form filling” task	104
Error analysis and data preparation	104
Descriptive statistics	105
Inferential statistics.....	109

Cross-referencing “document editing” task.....	117
Error analysis and data preparation	117
Descriptive statistics	118
Inferential statistics.....	122
Usability ratings of input devices.....	129
Descriptive statistics	131
Inferential statistics.....	133
Device preference.....	135
Post-trial interviews.....	139
Discussion.....	143
Review of hypotheses.....	143
Discussion of quantitative results.....	145
User preferences and qualitative indicators.....	152
Limitations	153
Suggestions for future research.....	155
Design recommendations.....	156
References.....	159
Appendix A Participant Recruitment Materials.....	163
Recruitment Script.....	163
Recruitment flyer used for undergraduate Psychology students.....	164
Recruitment website link from flyer.....	165
Recruitment advertisement used on craigslist.org for older adults	166

Appendix B Experimental Materials.....	167
Agreement to participate in research (consent) form	167
Digit Span Test.....	168
Borg RPE CR-10 Scale.....	169
Appendix C Procedural Scripts.....	174
General Instruction Script.....	174
Borg CR-10 Participant Instructions	176
Borg CR-10 Test items.....	177
Appendix D Task Analysis.....	178
Appendix E Omitted Tasks.....	182
Windows task	182
Excel scrolling task.....	184
Appendix F Letter of Approval from Human Subjects Institutional Review Board	185

List of Figures

Figure 1. Tablet PC with pen.	5
Figure 2. Pocket PC.....	6
Figure 3. Touch screen and light pen.	7
Figure 4. Graphic tablet.	8
Figure 5. The tablet PC configured for two devices: (A) pen, and (B) mouse.....	34
Figure 6. Excel dragging task.	42
Figure 7. Excel dragging practice task.	42
Figure 8. PowerPoint dragging task.	43
Figure 9. PowerPoint dragging practice task.	44
Figure 10. PowerPoint manipulation task.....	45
Figure 11. PowerPoint manipulation practice task.....	46
Figure 12. Tracing task.....	47
Figure 13. Tracing practice task.....	48
Figure 14. Cross-referencing “form filling” task - first page of paper survey.....	50
Figure 15. Cross-referencing “form filling” task - electronic equivalent.....	51
Figure 16. Cross-referencing “form filling” practice task.....	52
Figure 17. Drop-menu items in the cross-referencing “form filling” task	53
Figure 18. Cross-referencing "document editing" task - first page.....	55
Figure 19. Cross-referencing "document editing" task - second page	56
Figure 20. Cross-referencing "document editing" practice task.	57
Figure 21. Slip errors for Excel dragging task.....	61
Figure 22. Excel dragging task measures	67
Figure 23. Slip errors for PowerPoint dragging task.....	71

Figure 24. PowerPoint dragging task measures.....	76
Figure 25. PowerPoint dragging: time as a function of order.....	78
Figure 26. PowerPoint dragging: time, limited to first device.....	79
Figure 27. Slip errors for PowerPoint manipulation task	81
Figure 28. Text selection error (mistake) in PowerPoint manipulation task	82
Figure 29. PowerPoint manipulation task measures	87
Figure 30. PowerPoint manipulation: RPE-mental as a function of order	91
Figure 31. PowerPoint manipulation: RPE-mental, limited to first device	92
Figure 32. Examples of tracing images	93
Figure 33. Tracing task measures.....	98
Figure 34. Tracing: RPE-mental - 3-way interaction of device, age, and order.....	102
Figure 35. Tracing: RPE-mental, limited to first device	103
Figure 36. Examples of slip errors in the cross-referencing “form filling” task.....	104
Figure 37. Cross-referencing "form filling" task measures	110
Figure 38. Cross-referencing "form filling": time as a function of order.....	112
Figure 39. Cross-referencing "form filling": time, limited to first device.....	113
Figure 40. Cross-referencing "form filling": RPE-physical as a function of order	115
Figure 41. Cross-referencing "form filling": RPE-physical, limited to first device	116
Figure 42. Examples of slip errors in the cross-referencing “document editing” task...	117
Figure 43. Cross-referencing "document editing" task measures.	123
Figure 44. Cross-referencing "document editing": time as a function of order.....	125
Figure 45. Cross-referencing "document editing": time, limited to first device.....	126
Figure 46. Mean scores of usability dimensions by device – both age groups.....	134
Figure 47. Device preference by task, younger adults only	136

Figure 48. Device preference by task, older adults only	137
Figure 49. Overall ratings of device by age group, limited to first device.....	139

List of Tables

Table 1. Time hypotheses.	29
Table 2. Hypotheses for ratings of perceived effort.	30
Table 3. Experimental design matrix.	32
Table 4. Participant demographics and screening information.	33
Table 5. Study protocol summary.	40
Table 6. Descriptive statistics for Excel dragging task	65
Table 7. Effects summary for Excel dragging task	66
Table 8. Descriptive statistics for PowerPoint dragging task	74
Table 9. Effects summary for PowerPoint dragging task	75
Table 10. Descriptive statistics for PowerPoint manipulation task.	85
Table 11. Effects summary for PowerPoint manipulation task	86
Table 12. Descriptive statistics for tracing task	96
Table 13. Effects summary for tracing task	97
Table 14. Descriptive statistics for cross-referencing “form filling” task	108
Table 15. Effects summary for cross-referencing “form filling” task	109
Table 16. Descriptive statistics for cross-referencing “document editing” task	121
Table 17. Effects summary for cross-referencing “document editing” task	122
Table 18. Dimensions of usability survey breakdown	129
Table 19. Descriptive statistics of usability dimensions survey results	132
Table 20. Summary of effects for usability dimensions survey	135
Table 21. Device preference by task	136
Table 22. Time and error hypotheses	144
Table 23. Ratings of perceived physical and mental effort hypotheses	145

Introduction

Research problem

The mouse as an input device

An integral component of the graphic user interface (GUI) is an input device for manipulating it, and since the GUI's introduction to personal computing, the mouse has been the device of choice. There are good reasons why the mouse has dominated computer use: it affords an intuitive extension of the hand into the virtual workspace; it effectively performs the elemental GUI tasks of pointing, clicking, and dragging (MacKenzie, Sellen, & Buxton, 1991); it is stable – requiring little support from the user to maintain its position (Hinckley, 2003); it allows the user to traverse the interface space efficiently; it is also inexpensive and easily designed in various ways to accommodate additional functions, hand shapes, and other user needs.

Yet, despite its ubiquity, even from the early days of the personal computer, the mouse has been a poor tool for certain tasks. Early digital painting programs that came with GUI operating systems (e.g., MacPaint) allowed the user to create freehand illustrations, yet the mouse does not support freehand drawing tasks well, as it lacks the tactile feedback and hand grip necessary for precise finger movements that drawing requires (Meyer, 1995). It has long been established that some tasks are better performed with alternate input technologies, and some specialized communities of users (e.g., Computer-Aided Design) have adopted them.

Twenty-five years later, computers are used for a wide spectrum of office and creative tasks, yet the mouse still remains the default, “one size fits all” tool that is

typically provided with a computer. An important question for computing today is what types of input devices best support users in performing the broad spectrum of tasks for which personal computers are being used.

Diversity of tasks performed on today's PCs

The personal computer today is no longer just a word processor or spreadsheet application. In addition to such typical office tasks, people use computers for banking, creating presentations, gaming, archiving and playing digital media (e.g., photos, music, videos), home design visualization, and a host of other applications. Creative industries utilize computers in many capacities, from handling office work and accounting, to desktop publishing, generating artwork, and running specialized applications designed for the particular industry. Finally, of course, the mass-market entry to the internet through email and the world wide web has made computers the principal gateway to information and communications in an information society.

Such a diversity of tasks associated with personal computing necessarily broadens the user base from a smaller, specialized group to practically any person actively participating in the information society. Thus, along with the earlier question of device suitability for task performance, a second question arises: how does the wide user base interact with the input device? The “one size fits all” mouse may not be appropriate for every task. Furthermore it may not be appropriate for all users in the context of specific tasks.

User capabilities in an aging workforce

There are countless ways to look at the population of “all users” using personal computers. However, the factor of age is particularly important for several reasons. First, aging is inevitable and universal. Second, the time is quickly approaching when, for the first time, large numbers of people will be entering more advanced age groups and using computers. Third, there are well-documented issues associated with older adults on two levels: age-related motor and psychomotor decline, and learning issues for people who have to learn new technological skills later in life. Finally, research on older adults and input devices suggests that older adults may have more difficulty with the mouse than younger users, and that alternate input devices may mitigate these differences (e.g., Smith, Sharit, & Czaja, 1999; Walker, Millians & Worden, 1996; Chaparro, Bohan, Fernandez, Choi, & Kattel, 1999; Charness, Holley, Feddon, Jastrzembski, 2004).

Summary

While the mouse has been a reliable and effective input device for use with the GUI since its inception, the explosion of use of the personal computer into the daily lives and activities of a broad cross-section of people performing a wide diversity of tasks raises the question: is a “one size fits all” device suitable for all tasks across all users, or should personal computing evolve into more specialized configurations? Additionally, with the known effects of aging and their impact on human-computer interaction, along with the inevitability of aging for all computer users, a vision of computing for the 21st century must include consideration of the changing needs and capabilities of users as they age. Therefore, research addressing these questions is vital to clarify and enhance the

existing literature, as well as flesh out uncharted areas created by new and evolving technology.

New input technology options: the tablet PC

Tablet PC description and uses

A central assumption of the present study is that no one input device is optimal for all use contexts, and thus identifying cases where one device performs better or is preferred when compared to another is an important goal. Some devices are very similar to each other in interaction capabilities (e.g., a mouse and a trackball, which is essentially a mechanical mouse turned upside down) and differ only in physical manipulation, whereas other devices have fundamentally different interaction functionality (e.g., a mouse versus a pen, which have different ways of “clicking” as well as dramatically different physical interaction). Nevertheless, with the exception of highly specialized input devices (e.g., bar-code readers), all input devices have to accomplish the same kinds of tasks in the computer interface.

Figure 1. Tablet PC with pen.

The introduction to the market of the convertible tablet PC provides an opportunity to look at a computer that affords both traditional mouse input as well as the more novel pen-based input. These computers can function as a regular laptop computer with keyboard and mouse or touchpad input, but with a flip of the screen can be converted into a tablet PC with pen input. They have been successfully implemented in domains such as hospitals and teaching environments where the ability to capture analog sketches and diagrams, as well as mobility and computing power are valued (Anderson, et al., 2004; Machado, 2004; Keely, 2005). From a methodological perspective, they provide an opportunity to compare computer input devices while controlling other aspects of the computer such as screen size, resolution, and performance.

Similarities to other technologies

Figure 2. Pocket PC.

The tablet PC and pen is most analogous in form and function to a personal digital assistant (PDA), or pocket PC, except that it is larger. The interaction with regard to text input is different, but the fundamental user experience is very similar. The advantage of a pen interface in a mobile computing context is that no additional surface space is needed for the input device.

Another similar device is a digital pen and tablet seen in grocery or department store checkouts, where the user points and selects choices with the pen and signs their name on the screen itself. In addition to the advantage of having only one interaction space, in this case the natural analogy to signing a paper slip makes the interface easy for the user and expedient for the busy check-out line.

Figure 3. Touch screen and light pen.

Touch screens, which use the finger as the input “device,” and light pens, are also similar to the tablet PC with pen, the main difference being that with these devices the display is usually fixed, and often in a vertical orientation. Touch screens are best used in contexts where the requirement of an external input device is inappropriate, e.g., kiosks, ATMs (Hinckley, 2003). Use of the finger alone for interaction, however, while providing the most natural “direct” interaction, does not afford the precision control or movement scaling found in mechanical input devices. (Bullinger, Kern, & Braun, 1997).

Figure 4. Graphic tablet.

Graphic tablets are input devices in which the user manipulates a pen on a pressure-sensitive tablet, to control the cursor on the display. Graphic tablets come in various sizes, ranging from small, mouse-pad size up to large full-display size. Each point on a graphic tablet maps directly to a point on the display; thus, graphic tablets are said to be absolute positioning devices. They also have the precise tactile control afforded by the pen not found in other indirect devices like a mouse, and are popular for sketching and drafting tasks (Greenstein, 1997).

Because tablet PC with pen computing is relatively new, there are few published studies that look at the current device technology, and studies of similar devices, while informative, may not be directly applicable or up to date. This is an ongoing problem in the literature of input device evaluation in general; as the technology evolves, results from earlier experiments may no longer be relevant. However, properties shared between earlier technologies and tablet PC computing can provide a basis for deriving insights

from earlier studies. The following section defines these properties of input devices, and reviews the relevant literature associated with each.

Input device properties – dimensions of interaction

Direct vs. indirect devices

Definition. Input device interaction may be characterized as either direct or indirect. With direct interaction, the input surface is also the output surface, or display (e.g., a touch screen). With indirect interaction, the input surface is different from the output surface (e.g., a mouse pad, or graphic tablet). Direct interaction is most like using an analog tool (e.g., pencil and paper, hammer and nail, mechanical buttons) and is therefore more natural to use for analog-type tasks (e.g., drawing) and cases where a computer has replaced a mechanical interface panel (e.g., ticket vending machine, ATM). A tablet PC with pen is characterized by direct interaction (as are touch screens, light pens, and voice control), whereas a mouse is indirect, with the user moving the mouse control on a mouse pad, and observing the results of their actions on the display.

Issues of eye-hand coordination, stimulus-response compatibility, and learnability are minimized with direct interaction (Greenstein, 1997). However, technological limitations on the recognition and interpretation of writing and speech remain problematic and limit the kinds of tasks that can be accomplished with direct interaction alone (i.e., without a keyboard). Parallax is a stimulus-response mismatch that results from a spatial separation between the display and an interaction with the display. Because the tablet PC interaction surface is raised a number of millimeters above the actual LCD output, parallax problems can occur and interfere with the benefit of direct

interaction, particularly as the size of the interaction target decreases. Finally, different ergonomic issues (e.g., fatigue) are associated with direct interaction because of the more physically involved nature of the interaction. Various sources also cite occlusion of part of the display (by the hand and arm) as a problem with direct interaction (e.g., Bullinger, Kern, & Braun, 1997; Greenstein, 1997; Hinckley, 2003), but this concern remains largely anecdotal.

Several early studies found advantages to using direct instead of indirect interactions. Albert (1982) found that participants performed better with a light pen and touch screen than with indirect positioning devices, although he did not include a mouse in the study. He attributed the results to “direct eye-hand coordination” present with the light pen and touch screen. Haller, Mutschler, and Voss (1984) found that a light pen resulted in better performance than a host of other devices, including a mouse, for a cursor positioning task. Both of these studies were conducted very early in the history of personal computer use, however, so their sample may no longer be relevant to today’s computer-savvy public.

Results from more recent studies that primarily compare a direct to an indirect device suggest that task demands are a strong influence on which type of device is better suited to the task. For elemental pointing tasks, light pens have been found to result in faster performance than the mouse (Charness, Holley, Feddon, & Jastrzembski, 2004), and touch screens have been found to result in superior or equivalent performance compared to a mouse on a tracing task (Cohen, Meyer, & Nilsen, 1993; Sears & Schneiderman, 1991). However, in a follow-up study using a more complex task, Meyer,

Cohen, and Nilsen (1994) found that the same touch screen that had resulted in better performance on the tracing task resulted in the worst performance and low subjective user ratings when the task involved scrolling and activating menus. The direct interaction devices appear to better support simpler ballistic tasks (e.g., pointing) or tasks akin to actual physical manipulation (e.g., tracing).

When the factor of age is introduced, however, the task demands appear less influential. In a study that compared a touch screen to a rotary encoder (an indirect input device consisting of a rotary knob and push button) for a series of typical tasks on an entertainment system simulator (Rogers, Fisk, McLaughlin, & Pak, 2005), it was hypothesized that participant performance on ballistic, pointing, and discrete tasks would be better with the direct device, whereas when executing precision and repetitive tasks their performance would be better with the indirect device. While the hypotheses for the direct device were generally confirmed for younger users, in all but one pointing task there was no difference between devices for older users. Hypotheses for the indirect device were only confirmed for 1 of 5 conditions, for both age groups, and in once case the opposite result was found for the younger group. The odd choice of rotary encoder as an indirect device for comparison may have affected the outcome of this study; although it eliminates the problem of participant experience with a mouse, the limited interaction capabilities of the rotary encoder may have contributed to performance differences when compared to the touch screen.

Older adults are often identified as potential beneficiaries of direct interaction devices because these devices are easier to understand and require no mapping or

translation, cognitive functions that are known to be slower in older adults (Proctor, 2005; Vercruyssen, 1997). However, two recent elemental task studies comparing touch screens to a mouse for older users present starkly contrasting results. Phillips (2003) found that older adults not only performed better with the mouse, but preferred it as well. The tradeoff between physical exertion and benefits of direct interaction in this study clearly favored the less physically challenging task of mousing. Murata and Iwase (2006), however, found that pointing on a touch screen was faster for older adults than mousing. They did not report user preference or other behavioral data. Both studies were methodologically similar, although they used different testing software and variations on target size and distance. It is possible that participants in contemporary samples, who likely have experience with the mouse, do not consistently show better performance with direct devices with which they have less experience. Also, because within-group variability in motor performance of older adults can be quite dramatic (Spirduso & MacRae, 1990, as cited in Vercruyssen, 1997), the results from such studies should be interpreted with the particular characteristics of the sample taken into account.

Absolute vs. relative devices

Definition. Devices may be described as operating with absolute positioning or relative positioning. With absolute positioning, every point on the input device surface maps to a point on the display. Thus a pen or graphic tablet, which both have a fixed surface area, operate in absolute mode, whereas a mouse, trackball, or rotary encoder, which have no fixed size surface to work with, only operate in relative mode. A tablet PC with pen operates in absolute mode.

Absolute versus relative positioning properties provide a tradeoff between the cognitive effort involved in mapping and translation, and physical effort involved in manipulating the screen pointer with the device. An absolute positioning device like the tablet with pen requires no mapping or translation, but involves more physical effort to move the cursor on the screen. A relative positioning device like the mouse requires mapping and translation, but also requires much less physical effort to manipulate the pointer.

Absolute positioning devices like a graphic tablet are generally assumed to be better for drafting or sketching tasks (Bullinger, Kern, & Braun, 1997), which has been affirmed empirically (Apte & Kimura, 1993; Elliott & Hearst, 2002). Additionally, in a comparison of graphic tablet, mouse, and trackball, participants using the graphic tablet were found to perform as well as when using the mouse on pointing tasks, and performed faster with the tablet than with the mouse on dragging tasks (MacKenzie, Sellen, & Buxton, 1991). However, in a study of goal-oriented interaction tasks (including common GUI interface actions such as selecting menu items, zooming, and dragging) user performance on absolute devices (a graphic tablet and a touch screen) was worst when compared with relative devices (a mouse and a mousepen, a stylus-like device with a mechanical ball in the tip) (Meyer, Cohen, & Nilson, 1994). As with direct and indirect interaction, it appears that simple ballistic tasks in isolation or tasks that are analogous to real manipulation (e.g., dragging, drawing) are better served by absolute device, while relative devices are more useful for complex and varied tasks.

However, in a pointing and dragging study with older adults (aged 65 and older), participant performance with the mouse was superior to the graphic tablet in both pointing and dragging tasks (Alvarado, 2004), although in this study the participants had on average 7 years experience with the mouse, and no prior experience with the tablet or trackball. The trackball, also a relative device, was statistically equivalent to the absolute devices in the Meyer (1994) study. However, when participants used the trackball in the simpler pointing and dragging studies, their performance was poorest. User experience with the various devices may have clouded the findings, particularly with the later studies where participants generally were skilled mouse users. Additionally, the mechanics and physical manipulation of the input device may be more important than whether it is an absolute or relative device.

Physical manipulation

The pen is held between the thumb and finger, affording the precise finger manipulation that makes it superior to the mouse for drafting. The mouse, held by a looser grip of the whole hand, is more evenly manipulated with all the fingers and the wrist, thus alleviating pressure on the fingers while still allowing enough control of the cursor for interacting with the computer. Actual physical effort was found by EMG to be greater for the mouse than the trackball (Karlqvist, et al. 1999), although in a separate study, older users were found to have the same amount of actual exertion while perceiving greater exertion with the mouse (Chaparro, Bohan, Fernandez, Choi, & Kattel, 1999). No studies comparing actual physical effort involved in manipulating a pen compared to a mouse were found.

Positioning the pointer with a pen involves active use of the full arm, just as an analog pen does, whereas with the mouse a flick of the wrist is adequate to move the pointer across the screen. The active involvement of the arm has implications on two levels: the physical effort involved in accomplishing the task, but also any potential cognitive assistance that limb action may provide to the execution of a task. Thus for moving around the GUI the mouse may be a better tool, whereas for tasks that are analogous to physical manipulation (e.g., slide sorting) the pen may be preferred, although this has only partially been demonstrated empirically (Elliott & Hearst, 2002). It is likely that for a well-practiced task with a traditional pen or pencil (e.g., drawing, tracing, or handwriting), users should be able to positively transfer their skills from an analog pen to a digital pen, particularly if they can orient the digital input surface in a natural manner.

Issues related to the orientation of the input surface and the associated requirement of the hand and arm have been shown more clearly to be a problem. Not surprisingly, when a user is obliged to hold their hand up to interact with a vertically oriented device, be it a monitor (in the case of touch screens or light pens) or a vertically mounted device, they experience fatigue and dislike the interaction (Meyer, et al. 1994; Phillips, 2003). Mack and Montaniz (1991) found that changing the angle of the touch screen to a more horizontal position improved performance on typical office tasks so that performance with a pen was comparable to a mouse.

Summary of device properties

Direct versus indirect. Task demands are a factor in weighing the benefits of direct versus indirect interaction. Simpler tasks that are primarily ballistic in nature, or tasks which are analogous to real manipulation seem better suited for direct interaction. However, among older adults, task differences appear to be less important. The common wisdom that direct interfaces are better for older adults has been empirically supported, but also challenged.

Absolute versus relative. The clearest benefit for absolute positioning devices appears to be for drawing or sketching tasks, although absolute devices have also been demonstrated to be comparable to indirect devices and superior for dragging tasks. In more complex tasks, however, the indirect mouse performs well and is preferred by users, but it is difficult to identify this factor alone, as participants in recent studies often have years of experience using the mouse.

Physical manipulation. The physical affordances of a pen make it naturally superior for drawing tasks, and may make it beneficial for tasks that are analogous to physical manipulation. The orientation of the interaction surface affects the user's physical effort and preference, particularly with direct devices that involve interaction directly with the display; vertical interaction areas cause fatigue and discomfort in users. The mouse offers stability and requires less physical effort to manipulate it.

The tablet PC and pen shares the properties of many of the devices that have been studied: it has direct interaction, like a light pen or touch screen; it uses absolute positioning, like a graphic tablet or touch screen; it has physical affordances of a graphic

stylus that make it suitable for tasks that require precise hand control, and may be better suited for tasks that are analogous to physical manipulation. When compared to a mouse, these properties sometimes favor the mouse, and sometimes favor other devices. It has not been established how well a pen with tablet PC measures up, nor has the potential impact of task choice been explored.

Task effects

Elemental task differences

The majority of studies on input devices have measured movement time and errors in a barrage of repetitive elemental tasks, such as pointing, dragging, and clicking (e.g., Alvarado, 2004; Chaparro, 1999; Charness, 2004; Cohen, 1993; MacKenzie, 1991; Murata, 2006; Phillips, 2003; Smith, 1999; Walker, 1996). While these studies provide baseline information about human performance with input devices, they do not illuminate the user experience with devices. Furthermore, results are often difficult to compare across studies because of the range of devices, tasks, and testing methodology used (Soukoreff & MacKenzie, 2004). Vercruyssen (1997) warned of the limited ability of laboratory tasks to predict real-world performance for older adults. While older adults in general perform slower than younger adults, they may also adapt alternate strategies for accomplishing tasks, which may not be captured in simple task studies.

Elemental task studies have yielded some useful information about task effects. It has been demonstrated that dragging takes longer than pointing (Cohen, 1993; MacKenzie, 1991), and Chaparro (1999) found that this effect was consistent across older and younger age groups (i.e., both age groups experienced the difference). In a separate

study, older users were found to have more difficulty with clicking and double-clicking than younger users (Smith, 1999), presumably because of the additional fine motor requirements for clicking accurately. However, in studies with only older adults, Alvarado (2004) and Phillips (2003) found no difference between pointing and dragging. Older adults may be trading off speed for accuracy to compensate for motor skill deficiencies, resulting in less pronounced differences between tasks.

Goal-oriented task differences

Beyond elemental tasks, a handful of studies have looked at different input devices for more complex, goal-oriented tasks. This is an important area of research that has been largely neglected in the name of experimental control, but studies of complex tasks may provide a more meaningful user experience in the laboratory, and may shed greater light on actual user preferences (Vicente, 1997). Instead of subjecting participants to a barrage of repetitive tasks, participants experience something more akin to what they would do in a natural work environment. Goal-oriented task studies are not without problems either, as they suffer from less experimental control, and also are difficult to compare simply because the tasks are different (e.g., Apte and Kimura, 1993; Elliot and Hearst, 2002; Meyer, 1994; Rogers, 2005).

The advantage of goal-oriented task studies is that they provide insight into how the variables under consideration would affect performance in a more realistic context. Looking at the common architectural tasks of slide sorting and sketching, Elliott and Hearst (2002) compared a mouse and monitor, tablet PC, and large digital desk (the last two being both pen-based). Not surprisingly, participants performed better on the

sketching task with the pen-based devices, but the slide-sorting task produced more mixed results. For slide sorting 10 participants preferred the pen-based conditions, whereas 11 preferred the mouse. Performance data was not statistically significant, although in general poorest performance was observed on the large digital desk.

Reported parallax problems and difficulties encountered by shorter participants with the digital desk may have contributed to the findings. The tablet PC fared well, although participants reported that they thought the screen was not big enough. While this study did not show a significant preference for the more tactile pen for slide sorting as hypothesized, the results for the sketching task reinforced the established notion that optimal device selection is task dependent.

Rogers, Fisk, McLaughlin, and Pak (2005) studied a variety of common GUI tasks in the context of a simulated entertainment controller across two devices and age groups. Tasks were classified as discrete, pointing, ballistic, precision, and repetitive. Based on generalizations resulting from elemental studies, they hypothesized that the precision and repetitive tasks would be better served by an indirect device, and the other tasks would be better served by a direct device. Task demands were found to be more predictive of an optimal device for younger users; older users' performance, while slower than younger users for all tasks, was not significantly different between devices for 7 of the 9 tasks. Their results suggest that findings from elemental studies, which formed the basis for their hypotheses, may not carry over when a participant is working in a more goal-oriented environment.

How should goal-oriented tasks be developed? The Rogers study took the approach of creating a media simulator (i.e., a representative application) to look at a series of tasks executed in a meaningful context which, while not as simple as pointing, clicking, and dragging, could still be discretely isolated. The Elliott study took a domain specific approach by evaluating overall performance on complex tasks common to architects. Because of the discrete nature of the tasks (e.g., clicking a button) the former approach lacks external validity, whereas by looking at only an overall process, the latter approach lacks internal validity. In an effort to bridge this gap, identification of meaningful tasks that are employed by a wide range of individuals would be useful.

In an exploratory study of work-related reading, Adler, Gujar, Harrison, O'Hara, & Sellen (1998) looked at fifteen participants from a broad range of professions and identified some commonalities across domains. Among their findings, they identified reading for cross-referencing as the most frequently occurring task. In their taxonomy, "cross-referencing" involved reading from multiple sources or from one source in order to integrate information. Specific examples of cross-referencing were grouped into "extracting information" (e.g., referring to one document to fill out a form), "integrating information" (e.g., using information from one document to make a new one), "consistency checking" (e.g., checking to make sure documents are consistent), and "critiquing" (e.g., commenting on or referring to one document while writing another). By focusing on these types of goals, an investigator may be confident that the results will be meaningful to a broad cross-section of users.

Both elemental tasks and more goal-oriented tasks yield useful information in any comparison of input devices, although the majority of studies in the literature use elemental tasks. A goal-oriented approach may shed more light on computer interaction in real activities, especially when the task selected is of relevance to the target user group. Older users seem to be less affected by differences in task demands, perhaps because of the bias for older people to trade off speed for accuracy in computer interactions (Vercruyssen, 1997).

Mediating the adverse effects of aging

In many of the device comparison studies seen so far, age has been a complicating factor in the results (e.g., Chaparro, 1999; Murata & Iwase, 2006; Phillips, 2003; Rogers, 2005). The factor of age has been considered in device comparison studies because of the oft-cited “graying of America”, i.e., the increasing population of older adults who will remain active in work and computing, coupled with known decline in performance of motor and cognitive skills associated with aging. A handful of studies have looked at software-based interventions to help older users with computers (e.g., Walker, Millians, & Worden, 1996; Worden, Walker, Bharat, & Hudson, 1997), but relatively few have looked at how input devices themselves may improve older adults’ experience.

Following is a brief summary of demonstrated effects of aging related to computer input, and description of the few studies that have found devices improve performance for older adults.

Every study that has compared older adults to younger adults has found that older adults perform more slowly (Walker, Millians, & Worden, 1996; Chaparro, Bohan,

Fernandez, Choi, & Kattel, 1999; Smith, Sharit, & Czaja, 1999; Charness, Holley, Feddon, & Jastrzembski, 2004; Rogers, Fisk, McLaughlin, & Pak, 2005). The robust effect of slower performance among older adults is no surprise, as it is consistent with known effects of age-related declines in motor skills and cognitive functioning (Vercruyssen, 1997). Older adults have been found to have some difficulty using the mouse, particularly for tasks involving more precision, like clicking and double-clicking (Smith, Sharit, & Czaja, 1999), although adjusting mouse acceleration has been found to ameliorate some of the age-related differences (Walker, Millians, & Worden, 1996). The trackball has been suggested as possibly better for older adults because of the reduced perceived physical effort in using them (Chaparro, et al., 1999), yet use of the trackball has also resulted in poorer performance for older users when compared to the mouse and graphic tablet (Alvarado, 2004). Finally, the verdict is still out on whether touch screens are beneficial or not for older users. The previously mentioned studies by Phillips (2003) and Murata and Iwase (2006), which compared mouse and touch screens on pointing tasks for older adults, have contradictory results.

Charness, Holley, Feddon, and Jastrzembski (2004) did find a significant performance improvement for older adults compared to younger adults when using a light pen as opposed to a mouse for a menu pointing task. In their study, older users (aged 65-75) performed slower than younger users (aged 18-25) overall, but their decrement in performance was significantly reduced when using the light pen. While quite analogous to the tablet and pen, the light pen is typically used with a vertical monitor, so participants have to hold the light pen up to the screen, resulting in greater physical

effort. Not surprisingly, the study also found that participants reported higher workload with the pen, and preferred using the mouse, even though in all cases participants were faster with the pen than the mouse. Another finding of the study was that participants of all ages preferred using the light pen with the non-preferred hand, which the authors attributed to the lack of translation and mapping associated with using a direct device like a pen in a novel situation.

The promising results for the pen observed in the Charness et al. (2004) study were limited by the simple pointing tasks being performed. The study by Rogers, Fisk, McLaughlin, and Pak (2005) followed by comparing a direct interaction device (touch screen) to an indirect device (rotary encoder) on a variety of interaction tasks in an entertainment system simulator, across two age groups. As expected, the younger adults (aged 18-28) always performed faster than the older adults (aged 51-65), but for the most part, no device differences were seen for the older adults, even when a significant difference was seen for the younger group. In other words, device advantages observed between tasks for younger users became mostly irrelevant for older users. The only interaction between age and device was for a repetitive movement, in which older adults benefited more than younger adults when using the rotary encoder. In a pointing task analogous to the Charness (2004) study, participants performed better with the direct touch screen than the indirect rotary encoder, but no age interaction was observed. It should be noted that the older age group in this study was younger than older groups in other studies by as many as 25 years, so the relatively young “old” group may have minimized age effects.

It is not clear if a rotary encoder, the indirect device used in the Rogers et al. (2005) study, would shed much light on how use of the mouse compared to a direct device would affect performance. The rotary encoder has a more limited interaction capability than the mouse, as it has no active tracking ability. With a touch screen, while the user cannot track the pointer's movement on the screen either, "passive tracking" of the cursor position happens in the user's finger itself (Buxton, 1990). Also, the touch screen has problems of physical effort and potential fatigue that may not be evident with a tablet PC interaction. Thus, while these studies do show some device-related advantages for older users in the context of some tasks, no clear recommendations could be made to help older users outside of very specific interaction contexts.

Summary

The well established decline in motor skills and cognitive functioning associated with aging affects older adults' performance using computer input devices; older adults perform slower than younger adults. The input device's ability to reduce age-related decrements has only been demonstrated in two cases: a light pen for a pointing task (which was not found in a subsequent study) and a rotary encoder for a repetitive task. The evidence is still too sparse to make any meaningful recommendations, however, and in many cases users preferred the device that performed less well. This issue of user preference is explored in the next section.

User preferences

An interesting finding in the Charness et al. (2004) study is that even though all the participants performed faster with the light pen than with the mouse on a pointing

task, all users reported less workload with the mouse, and they all preferred the mouse. Given that the performance difference between devices, while statistically significant, is in some cases less than one second, the question of user preference and comfort with the device becomes relevant. However, not all device comparison studies collect user preference data or other subjective measures of user experience. Furthermore, it can be difficult to determine whether user preference is a reflection of the interaction experience with the device, or merely of familiarity with the device.

In numerous studies involving a variety of different instruments, it has been shown that users do not always express a preference for the device with which they performed best (Andre and Wickens, 1995). Two types of subjective user data are typically collected. Some studies ask their participants to rate devices with Likert-type scales or compare devices by a forced-choice questionnaire (Alvarado, 2004; Charness, 2004; Elliot and Hearst, 2002; Meyer, 1994; Phillips, 2003). A smaller set of studies collect psychophysical data in the form of workload scales or ratings of perceived effort (Chaparro, 1999; Charness, 2004). The Chaparro et al. (1999) study also collected physiological (EMG) measures to compare perceived physical effort with actual muscle activity. While subjective measures were often consistent with their objective counterparts, in some cases there was inconsistency.

When compared with performance – usually measured as movement time and errors – subjective ratings were found to be mostly consistent, but in some cases, participants preferred a device even though they performed less well with it. In the Charness et al. (2004) study, participants preferred the mouse to the light pen, even

though they performed better with the pen in all tasks and age groups. However, when using their non-preferred hand, participants preferred the light pen, which also led to better performance. Using the devices with the non-preferred hand is a novel situation, and as such the subjective ratings may have been less influenced by prior experience with the mouse.

The workload ratings collected in the Charness (2004) study indicated less workload for the mouse, which was also the preferred device, not surprising given the vertical orientation of the light pen. The Chaparro et al. (1999) study reported less perceived effort for the trackball than the mouse, even though performance was equivalent with both devices. When compared with actual physical effort, participants' perceived effort ratings were inconsistent with actual physical effort; participants perceived greater effort with the mouse than the trackball, even though the actual EMG readings showed no difference. The participants also preferred the trackball, perhaps because they perceived less effort in using it. Thus workload and perceived effort, while not always in congruence with performance or physiological measures of effort, appear to influence user preference.

It seems clear that a user's experience with the input device affects their personal preference. In most studies, user experience with the mouse is substantial, while with the devices used for comparison, user experience is often limited to whatever practice training they receive at the beginning of the study. Thus, preference ratings are likely biased in favor of the mouse. Psychophysical metrics like workload or perceived effort are perhaps somewhat more objective from the user perspective, as they tap more directly

into the interaction experience in which the user is engaged. Nevertheless, capturing user preference ratings is important in the evaluation of input devices because it reflects the user's willingness to use and purchase them.

The present study

Summary

Purpose. The purpose of this study is to investigate performance and user experience of the ubiquitous mouse and an alternate device in the execution of complex, goal-oriented tasks, across groups of older and younger users.

Input device. The introduction to the consumer market of the convertible tablet PC provides an opportunity to look at a computer that affords both traditional mouse input as well as the more novel pen-based input. The properties associated with the pen and tablet PC – direct interaction, absolute positioning, and physical manipulation – have been shown to be comparable to a mouse in many cases, and superior in certain contexts. No studies were found that compared input devices within the same convertible tablet PC, an effective control for various factors associated with the device technology.

Age. Cognitive slowing and decline in motor function associated with aging has been shown in some cases to be mitigated by the use of different input devices. However, the pattern is by no means clear. Older users may trade off speed for accuracy when using input devices, or adopt different coping strategies, which may ameliorate task effects and even device effects.

Task. As elemental tasks become more complex, (e.g., dragging, double-clicking), their execution takes longer and differences between age groups may appear.

However, with age, task demands may be less influential on the choice of an optimal device. Few comparative studies have been done on goal-oriented tasks, which may shed greater light on actual applied situations, and have greater external validity. Goal-oriented tasks may be thought of as discrete complex interactions or an entire process, and both approaches have value. Tasks involving cross-referencing have been found to be common across a broad range of domains, and may be a suitable process to examine.

User preference and perceived effort. Personal computer use is not only about movement time, and minor, if statistically significant, differences in performance of devices should not be the only consideration. User preference not only reflects the quality of the user's computing experience, but also the willingness of a user to purchase a product. Subjective preference surveys may be biased toward the mouse, as it is difficult to separate out familiarity with the device from user interaction experience. Subjective measures of perceived effort or workload are sometimes in conflict with objective performance and physiological measures, yet may be more influential on user preference than actual performance. Subjective psychophysical ratings therefore provide a metric for an important aspect of the user experience that is worth consideration.

Thus, this study aims to look at a substantially different alternative to the mouse – the pen with tablet PC – in a range of goal-oriented tasks reflecting both commonly performed discrete yet complex tasks as well as overall processes common to many domains of work. In addition to performance measures, user preferences and ratings of perceived effort will be recorded. Finally, different age groups will be compared to further investigate the potential benefits of alternative input devices for older adults.

Hypotheses

Across all the tasks, it is to be expected that performance differences will be observed between age groups – older adults will perform slower than younger adults. However, it is expected that a speed-accuracy tradeoff will be observed with older adults, resulting in no difference from younger adults with regard to errors. Depending on the task, it is further hypothesized that the magnitude of age effects will depend on the device being used, such that use of the pen will reduce age-related decrements in performance compared to the mouse.

Device effects are hypothesized to vary depending on the task demands. Tasks that require more complex, abstract, or precise activity are hypothesized to be better performed with the mouse because the mouse uses indirect interaction and relative positioning, and the physical stability of the mouse is better for minute work. The pen, on the other hand, is hypothesized to be superior for tasks that involve primarily simpler, ballistic movements, and for tasks that are closely analogous to their real-world counterpart, such as drawing or dragging. Age and device interactions are expected for the tasks that favor the pen, but not for those that favor the mouse, as the more difficult precision work is expected to be more difficult for older adults. Refer to Table 1.

Table 1. Time hypotheses.

	Age	Device	Age x Device
Precision task	Older > Younger	Pen > Mouse	None
Ballistic task	Older > Younger	Pen < Mouse	Older better w/pen
Analogous task	Older > Younger	Pen < Mouse	Older better w/pen

With regard to ratings of perceived effort, age differences are only hypothesized in the more difficult (i.e., precision) tasks, because of age-related performance decrements. In general, it is expected that the pen will have higher perceived effort than the mouse because of the more physically involved nature of using the pen, as well as the fact that most participants will be novices at using the pen but experienced at using the mouse. For tasks that have a strong analog to a real-world task (e.g., drawing, dragging), perceived effort is expected to be lower or equivalent for the pen compared to the mouse, as the positive transfer of a motor program should reduce physical demands as well as mental effort. No interactions are hypothesized for ratings of perceived effort. Refer to Table 2.

Table 2. Hypotheses for ratings of perceived effort.

	Age	Device	Age x Device
Precision task	Older > Younger	Pen > Mouse	None
Ballistic task	Older = Younger	Pen > Mouse	None
Analogous task	Older = Younger	Pen = Mouse	None

Finally, subjective user preferences are expected to be more consistent with perceived effort and subjective perceptions about the usability of the devices, than actual performance measures, in cases where there are discrepancies between the two.

Method

Design

Independent variables

The study employed a 2x2 mixed factorial design, with the within-subjects factor of input device (pen, mouse) and the between-subjects variable of age group (younger, older). The methodological factor of order (pen first, mouse first) was also tracked to measure any carryover effects. A variety of goal-oriented tasks were performed. However, as goal-oriented tasks incorporate various elements, they were analyzed separately, and task was not considered a separate factor.

Dependent variables

Performance data was collected in the form of time on task and errors in task execution. An error analysis was conducted for each task, specific to the task demands. Subjective measures of perceived effort were collected for each task, including ratings of perceived physical effort and perceived mental effort. Subjective ratings of usability dimensions were collected for each device, reflecting device comfort, efficiency, usefulness, appropriateness and satisfaction. After completing all tasks with both devices, participants' overall device preference for each task were recorded.

Experimental order

The device sequence was counterbalanced within experimental groups, resulting in two possible orders (pen first, mouse first). Participants within each group were randomly assigned to one of the orders with the constraint that there be equal number of participants assigned to each order. Additionally there were two task sets used with

minor variations so participants were not performing precisely the same tasks on the second device. The experimental design matrix is shown in Table 3.

Table 3. Experimental design matrix.

Group	P#	1st device	1st Task Set	2nd device	2nd Task Set
Younger 18-30	P1	Pen	A	Mouse	B
	P2	Pen	A	Mouse	B
	P3	Pen	A	Mouse	B
	P4	Pen	B	Mouse	A
	P5	Pen	B	Mouse	A
	P6	Pen	B	Mouse	A
	P7	Mouse	A	Pen	B
	P8	Mouse	A	Pen	B
	P9	Mouse	A	Pen	B
	P10	Mouse	B	Pen	A
	P11	Mouse	B	Pen	A
	P12	Mouse	B	Pen	A
Older 55-70	P13	Pen	A	Mouse	B
	P14	Pen	A	Mouse	B
	P15	Pen	A	Mouse	B
	P16	Pen	B	Mouse	A
	P17	Pen	B	Mouse	A
	P18	Pen	B	Mouse	A
	P19	Mouse	A	Pen	B
	P20	Mouse	A	Pen	B
	P21	Mouse	A	Pen	B
	P22	Mouse	B	Pen	A
	P23	Mouse	B	Pen	A
	P24	Mouse	B	Pen	A

Participants

A convenience sample of 24 participants (12 aged 18-30 with mean age of 21.17, 12 aged 55-70, with mean age 62.17) were recruited for the study. The target age ranges were based on typical ranges found in similar studies (e.g., Smith et al., 1999; Charness et al., 2004; Rogers et al., 2005). “Older adults” in the context of an age variable in human

factors experiments have been defined as individuals over age 60 (Rogers, 1997). While there is variability of individual differences in all age groups, it tends to increase with age (Czaja & Lee, 2003), so a larger age range for the older participants is appropriate to capture the diversity of individual differences in that population. See Table 4 for a summary of participant demographics and relevant information.

Table 4. Participant demographics and screening information.

Group		Age	Years Mouse	Years Alt Device	Digit Span
Younger	Mean	21.17	9.25	0.25	1.67
	SD	3.51	1.96	0.87	1.3
	Median	20.00	10.00	0.00	2.00
	Min	18	6	0	0
	Max	30	12	3	4
Older	Mean	62.17	13.75	2.08	0.67
	SD	4.45	7.81	3.12	1.15
	Median	61.50	11.00	0.50	0.00
	Min	55	3	0	0
	Max	69	28	10	3

Participants were screened for demographics, computer experience, and physical limitations, to assure a level of competence for performing two hours of computer work. Participants were required to fit in one of the designated age groups, and to have used

Microsoft Word and Microsoft Excel at a basic level of competence. An equal number of males and females were recruited for each age group.

Participants were recruited from undergraduate psychology courses at San José State University, from postings on computer bulletin boards such as craigslist.org, and from postings to email lists known to have members in the correct age demographic. The younger participants received college credit for participating in the study. Older participants were paid \$50 for 120 minutes of time.

See Appendix A – Participant Recruitment Materials for the participant screening script and examples of recruitment materials.

Materials

Experimental instruments

A Toshiba Satellite R25-S3503 convertible tablet PC with a 14.1” display (1440x900 screen resolution) was used for both input device conditions (see Figure 5).

A. Configured for use with the pen

B. Configured for use with the mouse.

Figure 5. The tablet PC configured for two devices: (A) pen, and (B) mouse.

For the pen condition, the PC was configured as a tablet, and the pen that comes supplied with the PC was used for input. For the mouse condition, the PC was configured as a regular laptop computer, with a Kensington Mouse-In-A-Box optical two-button mouse. The tablet display orientation was portrait for the pen condition, and landscape for the mouse condition. However, task windows were sized so the participant would be using the same size work space regardless of the display orientation.

A series of 6 goal-oriented tasks were performed using Windows XP Tablet Edition operating system, Microsoft Excel, Microsoft PowerPoint, Microsoft Digital Image, and Microsoft Word applications. These applications were pre-installed on the computer. Two additional tasks were included in the protocol but eliminated from the analysis because participants made too many mistakes with them. The last 2 tasks required the participant to cross-reference a printed document with an electronic one, and make changes to the electronic document based on the printed version.

Data collection instruments

The study was conducted in a usability laboratory equipped with video cameras and microphones. Camtasia screen capturing software was installed on the tablet PC and used to record each participant's actions. Screen recording files were saved as digital media files. Performance data was captured by both the experimenter using a stopwatch for time, and the screen-capturing software for time and errors. Observational data on the participant's manipulation of the input device, facial expressions, and body postures were collected while the study was being conducted. Video files depicting the participant

interacting with the computer from overhead was recorded to a hard disk and burned to DVD.

To assess if there were differences in cognitive capabilities between age groups, a brief digit span test was verbally administered. This consisted of repeating back to the experimenter four sequences of random numbers of 4, 5, and 6 digits in length each, for a total of 12 number sequences. All participants received the same digit sequences in the same order. Although randomly generated, no sequences contained multiple occurrences of a single number, or a series of numbers in order. To pass the test, participants were required to make no mistakes on the 4-digit sequences and no more than 4 mistakes overall. All participants successfully passed.

To capture ratings of perceived effort (RPE), the Borg CR-10 scale was used after each task. The Borg CR-10 is a ratio scale consisting of both a numeric scale and qualitative descriptors. Examples of the qualitative descriptors are “Nothing at all”, “very weak”, “weak”, “light”, “moderate”, “strong”, “heavy”. On the Borg CR-10 scale, the value of 10 is “extremely strong” and intended to be the highest perception the user has ever experienced (also referred to as “max P”). Values higher than 10 are possible to accommodate an unspecified range from “max P” to “absolute maximum” that the participant can imagine. The Borg CR-10 scale has been used for human performance experiments to measure effort, pain, and other psychophysical phenomena, as well as in ergonomic studies similar to the present study (e.g., Chaparro et al, 1999; Karlqvist et al., 1999). The reliability of the Borg CR-10 scale has been shown with a split-half correlation of .96, retest correlation of .98, and other reliability ratings in similar ranges

(Borg, 1998). The scale has also been shown to have good content and construct validity, as well as concurrent validity with physiological measures.

When using the Borg CR-10 scale, participants are asked to first match the qualitative descriptor to their perception of the dimension being measured, then assign a numeric value, which may have fractional values. For the present study, two dimensions of perceived effort were recorded. Both were derived from “effort,” one of the six dimensions of workload used in the NASA Task Load Index (TLX). The specific definitions participants were asked to rate are as follows:

- Rate your **perceived physical effort**, defined as “how hard you had to work physically to accomplish your level of performance”;
- Rate your **perceived mental effort**, defined as “how hard you had to concentrate, focus, and think to accomplish your level of performance”.

To capture usability dimensions of comfort, efficiency, usefulness, appropriateness, and satisfaction, a pencil-and-paper survey was administered after completing all the tasks for each device. Participants were asked to rate their level of agreement with the following statements using a 5-point Likert type scale:

- I think that I would like to use this device frequently
- This device fits my hand well
- I would recommend this device to a friend
- This device slowed me down
- I can imagine many cases where I’d want to use this device
- I would never use this device if I didn’t have to

- I would prefer to use this device for all my computer tasks
- I was able to work efficiently with this device
- It felt comfortable to use this device
- Some tasks seemed easier with this device

Additionally, participants were asked to rate the device using a 5-point semantic differential scale. They rated their impression of the device as being Natural or Awkward, Uncomfortable or Comfortable, and Universal or Task Specific.

Overall device preference by task was determined by a forced-choice pencil-and-paper survey which required participants to choose a favorite device for each task.

See Appendix B – Experimental Materials for the digit span data collection instrument, Borg CR-10 scale, usability dimensions and user preference surveys, experimenter data collection instrument, and the participant consent form.

Procedure

Prior to running any participants, the study as proposed was submitted to the Human Subjects Institutional Review Board of San José State University. The letter of approval for the study is included in Appendix F.

Upon arrival, participants were greeted and asked to sign a consent form to participate in research. Next, the digit span test was verbally administered. A brief description of the study was provided, followed by an explanation of the Borg CR-10 data collection instrument. Finally, several practice items using the Borg CR-10 scale were administered to make sure the participant understood how to use the scale.

Participants were then familiarized with the computer and both input devices so that they could demonstrate that they were comfortable using either device. Participants had the opportunity to adjust furniture and device position to their comfort. They were asked to set up and use the computer as they normally would, as well as can be reasonably managed in the artificial laboratory environment. For the tablet PC conditions, participants were encouraged to experiment with various positions of the device, including holding in their lap, placing flat, upright, etc. during the familiarization period, but asked to retain their chosen position during the study. Most participants chose to balance the tablet on the keyboard stand so that it rested at an angle. Some kept it flat on the desktop, and several used it in their lap.

Participants were all experienced with the mouse (median years experience with the mouse was 10 for younger adults, and 11 for older adults) and required little practice or training with it. However, for most participants, this was the first time they had used the pen or a tablet PC (median years experience with an alternative stylus-like input device was 0 for younger adults, and 0.5 for older adults), so all participants were given an introduction to tablet computing tutorial that comes with the computer. Following the tutorial, they were given practice time with the pen until they indicated that they felt comfortable using the pen.

After the familiarization period, participants performed the series of goal-oriented tasks with the first device in their predetermined sequence. Prior to starting each task, participants were given a simplified practice task that gave them practice on all of the interactions they would need for the actual task. After completing all the tasks with the

first device, the usability dimension survey was administered. A required 5 minute break followed use of the first device, then the same series of tasks were conducted using the second device. Tasks were modified slightly between devices so the participant was not executing exactly the same task, but task variations had the same degree of difficulty and requirements. Following use of both devices, the final overall choice survey was administered, and a debriefing interview was conducted. Participants were thanked for their time, and paid, if appropriate. The study took between 90 and 120 minutes to complete for each participant. A summary of the timeline for the study is shown in Table 5.

Table 5. Study protocol summary.

Run Time		Duration
0.0	Greet Participant and sign consent	3.0
3.0	Digit span	2.0
5.0	Introduction to study and questions	2.0
7.0	Borg CR-10 training	3.0
10.0	Mouse Training and Tasks	35.0
45.0	Post Device Survey	2.0
47.0	BREAK	5.0
52.0	Intro to tablet PC/pen	12.0
67.0	Tablet Training and Tasks	40.0
107.0	Post Device Survey	2.0
109.0	Post Protocol Preference Survey	2.0
111.0	Post - interview	8.0
119.0	Thanks (and payment)	1.0
120.0		

See Appendix C – Procedural Scripts for the general introduction script, the Borg CR-10 instructions, and the Borg CR-10 practice items.

Task descriptions

Following are descriptions of the tasks that were used in the study, in the order the participant was asked to do them. For each task a GOMS-type analysis was conducted for the purpose of gaining insights into the task demands, and predicting which tasks would be better served by which input device. While GOMS analyses are often used to provide quantitative predictions of task execution time in skilled users, they can also provide qualitative insights to aid in design (John & Kieras, 1996). A summary of the findings of the analysis is included in the task descriptions that follow. The complete analyses can be found in Appendix D – Task Analysis.

Two of the tasks that were included in the study protocol were dropped from the results because of the high number of mistakes that participants made on these tasks. These are included in Appendix E – Omitted tasks.

Excel dragging task

Description. In the Excel dragging task, the participant must select and drag numbered cells from one position to another on an Excel spreadsheet. The cells are randomly ordered in a row, and the participant must move them into an ordered vertical column (see Figure 6).

Figure 6. Excel dragging task.

Practice task. The practice task had three lettered cells and three red “target” cells in a random position. Participants used the practice to learn how to activate the move tool and how to drag the cells (see Figure 7).

Figure 7. Excel dragging practice task.

Analysis. There are 10 repetitions of the sub-goal to move an individual cell to its intended destination. Each repetition involves a ballistic move to select the cell, a precision positioning and selection action to activate the correct tool, a dragging motion

Figure 9. PowerPoint dragging practice task.

Analysis. There were 29 repetitions of the sub-goal of moving an individual shape into the target area. Dragging in PowerPoint is primarily a ballistic movement, as the participant can point, click, drag, and release in a single motion. The difficulty of performing the sub-goal depends primarily on the size of the target; smaller targets are more difficult to accurately select, including shapes that may be long but very thin. Because each shape had to fit in the grey box but could not touch the edges or another shape, some mental component in the form of planning both which shape to select next, and where to drop it, was involved as well.

Because the task involves primarily ballistic movements and dragging movements which are analogous to physical activities, it is hypothesized that the pen will be the better device for this task. Although some of the targets are small, thus requiring a degree of precision, they represent a small proportion of the overall task. An interaction

is hypothesized between age and device, such that age related decrements in performance are less with the pen than with the mouse.

PowerPoint manipulation task

Description. This task requires the participant to move “callout” shapes labeled as compass points to the appropriate spots on the screen, and then manipulate the “tail” of the shape so it connects to the center, smaller box (see Figure 10).

Figure 10. PowerPoint manipulation task.

Practice task. The practice task had two of the callout shapes with destinations clearly indicated, and a small center box as in the actual task. Participants were encouraged to practice the task as many times as was needed (see Figure 11).

Figure 11. PowerPoint manipulation practice task.

Analysis. There were three sub-goals of the task, each of which was performed 8 times (as there are 8 callout shapes): move the callout shape away from the center, position the callout shape in the correct compass point, and manipulate the callout shape so the tail lines up with the center box. The first two sub-goals were very similar to the PowerPoint dragging task, being primarily ballistic movements, with some mental effort involved in positioning the shape in the correct spot. The third sub-goal was similar to the Excel dragging task because it involved first selecting the shape, positioning the pointer very precisely so as to activate the manipulation tool (a small triangle), selecting and dragging a small point with the correct tool, and releasing it at the correct destination. The precision required for this task is higher than for the Excel dragging task, however, as both the target point for activating the tool and destination spot is much smaller.

Because of the amount of precision required to select the correct tool and manipulate the callout shapes, it is hypothesized that the mouse will be the better device for this task. No age by device interaction is hypothesized.

Tracing task

Description. The tracing task requires the participant to trace a line-art image as best they can in as short a time as they can (see Figure 12). This task addresses how effective a tool is for freeform, drawing movement. The shape for tracing included straight lines, curves that come to sharp points, and smooth curves.

Figure 12. Tracing task.

Practice task. The practice task was the same as the real task, except that a simpler shape was provided for tracing (see Figure 13).

Figure 13. Tracing practice task.

Analysis. Although there is a single goal for this task, there are several methods the participant could use. They could attempt a continuous, uninterrupted trace, or trace the shape in segments. Continuous tracing only requires a single start point, a single click and a long, controlled drag, but the long drag requires much concentration. The second method involves numerous starting points, and repeated executions of the click, drag, and release, but less concentration to maintain the long dragging motion. No precision movement is required *per se*, although maintaining continuous control over a sustained period of time makes this task quite different from the other tasks. It is also a direct analog for drawing with pencil and paper when using the pen device, but not when using the mouse.

Because of the strong analog between this task and drawing with a pen, as well as the continuous nature of the task, it is hypothesized that the pen will be the better device for this task. An interaction is hypothesized between age and device, such that age related decrements in performance are less with the pen than with the mouse.

Cross-referencing “form filling” task

Description. This is the first of the two cross-referencing tasks – complex, realistic tasks that require the participant to refer to a printed document to make changes to an electronic document. The participant refers to a printed document that has multiple choice data items on it, such as might be collected in a typical survey. The task is to accurately and quickly enter the data from the printed document (see Figure 14) into an electronic, web-based, equivalent form (see Figure 15). Only standard HTML form widgets were used, e.g., radio buttons, checkboxes, and drop menus. Although there are two pages on the paper survey, the online form is a single webpage that the participant must scroll down.

Practice task. The practice task was a simple web form that included all of the widgets the participant would be using on the actual task: radio buttons, check boxes, and drop menus (see Figure 16). Participants were encouraged to change settings multiple times until they felt comfortable using all the widgets with the input device.

Mutual Fund Survey

Demographic Information

1. What is your age?

☐ 18 to 24 ☐ 30 to 39 ☒ 50 to 59 ☐ 65 to 69
☐ 25 to 29 ☐ 40 to 49 ☐ 60 to 64 ☐ 70 and over

2. Are you:

☐ Female ☒ Male

3. What is the highest level of education you have completed?

☐ High school ☐ Some college ☒ College graduate ☐ Advanced degree

4. What is your employment status?

☐ Self-employed ☐ Employed part time ☐ Retired
☒ Employed full-time ☐ Not employed ☐ Student

5. Which broad category best estimates your household's annual income before taxes?

☐ Under \$35,000 ☐ \$50,000 to \$74,999 ☐ \$100,000 to \$149,999
☐ \$35,000 to \$49,999 ☒ \$75,000 to \$99,999 ☐ \$150,000 or more

Investment Profile

6. For what purpose(s) are you investing in mutual funds ? Please fill in all that apply.

☒ Build estate to provide for family ☒ Tax shelter
☐ Current income ☒ Emergency funds
☒ Children's education ☐ Home purchase
☐ Part of general savings plan ☐ Parental/elder care
☐ Travel ☐ Maintain current lifestyle
☒ Retirement planning ☐ Long-term healthcare

Figure 14. Cross-referencing “form filling” task - first page of paper survey that the participant references for the task.

Mutual Fund Survey Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Mutual Fund Survey

Demographic Information

1. What is your age?

☐ 18 to 24 ☐ 30 to 39 ☐ 50 to 59 ☐ 65 to 69
☐ 25 to 29 ☐ 40 to 49 ☐ 60 to 64 ☐ 70 and over

2. Are you:

☐ Female ☐ Male

3. What is the highest level of education you have completed?

☐ High School ☐ Some college ☐ College graduate ☐ Advanced degree

4. What is your employment status?

☐ Self-employed ☐ Employed part-time ☐ Retired
☐ Employed full-time ☐ Not employed ☐ Student

5. Which broad category best estimates your household's annual income before taxes?

☐ Under \$35,000 ☐ \$50,000 to \$74,999 ☐ \$100,000 to \$149,999
☐ \$35,000 to \$49,999 ☐ \$75,000 to \$99,999 ☐ \$150,000 or more

Investment Profile

6. For what purpose(s) are you investing in mutual funds? Check all that apply.

☐ Build estate to provide for family	☐ Tax shelter
☐ Current income	☐ Emergency funds
☐ Children's education	☐ Home purchase
☐ Part of general savings plan	☐ Parental/elder care
☐ Travel	☐ Maintain current lifestyle
☐ Retirement planning	☐ Long-term healthcare

Done

Figure 15. Cross-referencing “form filling” task - Electronic equivalent of the same survey.

T5a - Cross Referencing Form Filling - Practice - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://localhost/T5a_practice.html

Practice Survey

Radio Buttons

Select one choice

☐ A
 ☐ C
 ☐ E
 ☐ G
☐ B
 ☐ D
 ☐ F
 ☐ H

Checkboxes

Check multiple options.

☐ Apple
 ☐ Kiwi
☐ Banana
 ☐ Orange
☐ Pear
 ☐ Peach
☐ Cherry
 ☐ Nectarine
☐ Plum
 ☐ Grape
☐ Papaya
 ☐ Mango

Option Menus

Rate an item

Item A.

Item B.

Figure 16. Cross-referencing “form filling” practice task.

Analysis. Although there was a single basic sub-goal of matching a question on the printed document to the electronic one and updating the electronic form, the form filling task could be divided into three sub-goals, based on the properties of the tasks. First, complete the radio button questions (5 instances); second, complete the checkbox questions (11 instances, grouped as 5 in one question and 6 in the other); finally, completing the drop-menu questions (16 instances, grouped as 6, 3, 3 and 4 items). Both the radio button and checkbox tasks were simple ballistic tasks requiring pointing and

clicking, although the targets were small. Mental components of these tasks involved comparing the printed document to the electronic document and finding the correct target. The drop-menu component had the added mental challenge of requiring the user to translate information that was represented visually on the printed document to the same information represented more abstractly on the electronic form (see Figure 17).

A.

	Excellent	Very Good	Average	Fair	Poor	Does Not Apply
Our Telephone Representatives						
A. Promptness in answering call	☐	☐	☒	☐	☐	☐
B. Courtesy and helpfulness	☐	☒	☐	☐	☐	☐
C. Clarity and completeness	☐	☐	☒	☐	☐	☐
D. Timeliness of follow-up call	☐	☐	☐	☒	☐	☐
E. Accuracy of response	☐	☒	☐	☐	☐	☐
F. Ease of understanding	☐	☒	☐	☐	☐	☐

B.

Our Telephone Representatives

A. Promptness in answering calls

B. Courtesy and helpfulness

C. Clarity and completeness

D. Timeliness and follow-up call

E. Accuracy of response

F. Ease of understanding

C.

Set value... → →

Set value... → →

Set value... → →

Set value... → →

Set value... → →

Set value... → →

Figure 17. Drop-menu items in the cross-referencing “form filling” task. (A) as shown on the printed document; (B) as they appear on the electronic form; (C) the actions required to set the equivalent value.

It is difficult to determine a definitive hypothesis favoring one device or the other for this task because while the primarily ballistic nature of the task suggests the pen would be the better device, the amount of precision required to hit the small radio buttons and checkboxes may make the mouse the better tool. Because younger adults with their better motor control are less likely to have problems with the precision component of this task (small targets) it is hypothesized that they will benefit by using the pen. However, older adults, who may have more difficulty with the precision component of the task, are expected to perform equally well with either device.

Cross-referencing “document editing” task

Description. In the final task, the “document editing” task, the participant refers to a printed document that has been marked up (or annotated) for layout changes (See Figure 18 and Figure 19). The “editor” has requested changes to the formatting of text (e.g., making some passages bold or italic), changes to the positioning of graphic elements, and manual correction of typographical errors (i.e., using the input device to select the misplaced letter, and manually drag it to the correct place). Although with today’s word processing applications it is much more likely that a user would use spell checking or select a different method for correcting the typo, participants were constrained to a single method for the purpose of testing the input device on a precision oriented component of the document editing task.

Real Estate News You Can Use *beld*
 Issue Number 1: June 20, 2006
 John Smith, A1 Real Estate Brokers, Los Angeles, California
 seller1@a1realestate.com or (310) 555-1234

Dear Clients, Friends, Neighbors, Associates:

I sincerely appreciate your trust, your business, and your referrals. You supply not only my livelihood but many wonderful relationships. I welcome all questions, comments, or opportunities to help you and your acquaintances. In order to provide better service and the latest relevant information regarding your real estate interests, I will be sending this newsletter to you periodically. Below are some of your most frequently asked questions or issues.

Halie Please rest assured that your personal information is secure with me, and that this is not part of any outside vendor connection. Please feel free to discontinue these mailings at any time by sending a quick "opt out" email to me at seller1@a1realestate.com, or by leaving me a voice mail at my direct phone number above.

Q: Where is the local real estate market headed? *beld*

A: Despite all the media attention paid to the potential for a residential real estate "bubble" bursting, the February 2006 sales figures are impressive. According to DataQuick Information Systems, 6,405 new and existing homes were sold in Los Angeles County. This figure is down 9.2% from the number sold last year in February 2005, but that was nearly a record year of frenzied market activity. As for prices, the median sale price (half above and half below) for LA County was UP another 15.6% from the 2005 median, another double digit gain. The predictions are that those yearly gains will soften, possibly into the single digits by this fall, but this hardly seems like a collapse. Interest rates are creeping upward, but still remain historically low in a forty year perspective. As buyers resist these higher prices and mortgages, and as more inventory competes in the marketplace, there will probably be a GRADUAL adjustment to a more "normal market" environment.

Q: What are the most valuable home improvements you can make? *beld*

A: Would you believe that a bathroom remodel returns 102.2% of the cost when you sell?... that a minor kitchen remodel returns 98.5% of its cost? According to a joint survey by Realtor Magazine and Remodeling Magazine, those are the national average returns for those two projects. Many more projects are detailed in their study which appeared in December, 2005. They have even qualified their figures based on regional locations and local contractor estimates. For further information, please send me a request.

Figure 18. Cross-referencing "document editing" task - first page of the paper annotated document.

Q: Will you get hit with "capital gains" taxes if you sell? *bold*

A: With the incredible gains in prices over just the last several years, many owners have now found that their home sale profits will exceed the \$250,000 single/\$500,000 couple exemptions. In fact, many owners have taken advantage of that equity by trading up BEFORE they incur a tax liability. By consulting your tax advisor, you may learn that after as little as two years of owning and occupying your principal residence, you can cash out that equity, placing it directly into a new home purchase or in the bank. If you do buy another home, it can be either more or less expensive than the home you sell, and you start the two year cycle all over again. That tax free trade is an incredible gift from both the IRS and the State, and no one knows how much longer it will last. Please consult your tax advisor for details, and remember that profit is calculated not based on your loans but upon your sale price minus the price you paid. If you need a Comparative Market Price Analysis or a Broker Price Opinion, please contact me.

Q: How can you find out about crime in your neighborhood? *bold*

A: There is a great new program initiated by the Los Angeles Police Department which actually maps the recent crimes in your area. You simply enter your street address and it returns reported incidents up to several months ago. The web address is: www.lapderimemaps.org/. Check it out.

Q: What is the best way for married couples or domestic partners to hold title to their homes? *bold*

A: Only you, in consultation with your legal and financial advisors can decide, but a recent change to California law has added a new designation called "Community Property with the Right of Survivorship". It offers married couples the benefits of holding title as "Community Property" while adding the financial and inheritance benefits of a "Joint Tenancy". I have a terrific article written by the legal staff of the California Association of Realtors which outlines those benefits. There is also a recent change in California law regarding exemption from property tax increases when domestic partners transfer title to one another. Either of these changes could provide significant financial benefits.

I hope you will be able to use some of this information. Thanks again for your friendship and your real estate business.

John Smith (310) 555-1234
seller1@arealestate.com

Figure 19. Cross-referencing "document editing" task - second page of the paper annotated document

Practice task. The practice task was a simple Word document with nonsense text and blank images (see Figure 20). The participant was able to practice all of the sub-goals – selecting and formatting text, moving objects, and manually correcting typographical errors.

Figure 20. Cross-referencing "document editing" practice task.

Analysis. There are three distinct sub-goals of the document editing task: move the pictures (6 instances), make changes to text formatting (7 instances), and correcting typographical errors (4 instances). Each sub-goal also involved the mental component of referring to the printed document for guidance. Participants were free to complete the various sub-goals in whatever order they pleased, for example by moving down the page, or by doing a group of the same sub-goals in sequence. The sub-goal of moving the

image objects is essentially the same as the PowerPoint dragging task – it involves a simple ballistic movement to select the object, a quick dragging movement, and a release. Some control is required to release the object in the correct place. The text formatting changes were somewhat more complex, requiring the participant to use some precision in selecting the correct range of text (accomplished usually by clicking, dragging, and releasing over the correct range). After selecting text successfully the participant must hit a formatting button – a straightforward ballistic task, although the buttons appear rather small on the screen. Correcting the typos is the most difficult sub-goal in the task, and perhaps the most difficult task participants were asked to complete in the entire study. After first searching and finding the correct word in the electronic document, the participant must select a very small target – a letter within a word. This task is more difficult because Microsoft Word tends to snap to select an entire word, so the participant must exert both precision and control to select the letter. Next, the participant must select and move the letter, which requires high precision to activate the move tool, and to drag the letter the short distance to its correct destination.

Because of the amount of precision required to accomplish the typo sub-goal, along with the variety of task demands, it is hypothesized that the mouse will be the better device for this task. No age by device interaction is hypothesized.

Results

Data preparation – overview

This section provides an overview of the approach used for examining the data associated with each dependent variable (time, errors, RPE for physical effort, RPE for mental effort) captured for each task. The results of data examination (e.g., specific removal of data points, data transformations applied) is described in the results section for each task.

Error analysis

Each participant's session, as recorded onto digital media files and video, was reviewed for an error analysis. Errors were classified as slips, lapses, or mistakes (Reason, as cited in Stanton, 2003). A slip was an instance where the participant had the correct intent, but failed to accomplish their goal due to an error in execution. Slips were the error type of interest in this study, as they reflect a failure of the input device to support the user in achieving their goal. A lapse was an instance where the participant forgot to complete some portion of the task. Although infrequent, in some cases lapses could result in the disqualification of a record from the analysis, as too many lapses would adversely affect timing, and the participant would not have the same basis as others to judge perceived physical and mental effort. A final type of error – a mistake – was an instance where the participant did not understand how to use the software or was confused about the goal of the task. Although attempts were made to eliminate mistakes by allowing participants to practice tasks first, in some cases they still occurred. Too many mistakes could result in disqualification of a record from the analysis, as they

affected time on task and had unknown impact on ratings of perceived effort. The details and results of each error analysis is described in the task-by-task sections that follow.

Testing for order effects

Each age group was counterbalanced so that half of the participants were run through the protocol with the pen first, and half with the mouse first. Although not the focus of this study, the methodological factor of order (pen first, mouse first) was included in the omnibus ANOVA for each task to determine if any order (i.e., carryover) effects were present. Main effects of order were considered of interest but not problematic because of the counterbalanced design. In cases where an interaction between order and another factor was present (i.e., differential carryover was present) a secondary analysis was conducted using only the data from the first device used. Only significant order effects or interactions are reported in the sections that follow.

Effect size

The effect size statistic reported in the present study is partial η^2 , which is a sample-based measure consisting of the sums of squares associated with the factor of interest divided by the sums of squares of the effect plus the error term, as shown in Equation 1.

Equation 1. Calculation of partial eta-squared effect size statistic.

$$\text{partial } \eta^2 = \frac{SS_{\text{effect}}}{SS_{\text{effect}} + SS_{\text{error}}} \quad (1)$$

It was necessary to use a non-additive measure of effect size because of the mixed factorial design; there are different error terms for the between-subjects factors and the within-subjects factor. However, the result is that the measure of effect size is artificially inflated because it is both sample-based (as opposed to population-based) and does not represent the effect in the presence of effects of other factors in the study (Tabachnick & Fidell, 2001). This should be kept in mind when interpreting effect size statistics.

Excel dragging task

Error analysis and data preparation

There were three types of slips that occurred in this task which were included in the error tally. The first was when a range of cells was selected rather than the target cell being dragged, i.e., the participant selected the wrong tool. The second was when a participant activated a cell by double clicking on it. The third was when a cell was dropped enroute to its intended destination. See Figure 21 for examples of the error types.

Figure 21. Slip errors for Excel dragging task.

Participant 17 (P17) had little experience with Excel and did not seem to understand the task mechanics, despite the practice time. P17's RPE ratings were also off the scale (ratings of 11). Thus P17 was removed from the analysis.

Descriptive statistics

Refer to Table 6 for descriptive statistics of the Excel dragging task.

Timing. Older adults appeared to perform slower overall – 57.50 seconds compared to 34.46 seconds for younger adults – and participants were also generally slower using the pen than when using the mouse, with overall time of 53.04 seconds for the pen and 37.91 seconds for the mouse. In most cases, participants finished the task in under 1 minute, with younger participants averaging 39.67 seconds with the pen and 29.25 seconds with the mouse, and older participants averaging 1 minute 7.64 seconds with the pen and 47.36 seconds with the mouse. Older participants also had greater variability in their time. Standard deviations for older participants was 24.48 seconds and 13.10 seconds for pen and mouse, respectively, compared to younger adults' 9.28 seconds and 5.14 seconds for pen and mouse, respectively.

Errors. Older adults made more errors overall – around 3 ($M = 3.05$) compared to between 1 and 2 ($M = 1.38$) for younger adults – and participants generally made more errors using the pen than when using the mouse, with average overall error count of 3.48 for the pen and between 0 and 1 ($M = .87$) for the mouse. On a device by device basis, younger adults made fewer errors on average than older adults, 2.42 with the pen and 0.33 with the mouse, compared to 4.64 with the pen and 1.45 with the mouse, respectively. As with timing, there was greater variability among older adults. Standard deviations for older adults ranged from 4.30 with the pen to 1.70 with the mouse, compared to younger adults 1.98 with the pen and .65 with the mouse, respectively.

RPE – physical. Younger adults provided ratings of perceived physical effort about the same as older adults, 1.57 compared to 1.37, corresponding to between “very weak” and “weak” on the Borg CR-10 scale. Overall ratings by device fell into a similar range, with an average rating of 1.75 for the pen, and 1.20 for the mouse. However, younger adults rated the pen higher than the mouse, 2.19 (“weak”) compared to 0.95 (“very weak”), whereas older adults rated both devices about the same, 1.26 and 1.48 (both between “very weak” and “weak”) for the pen and mouse, respectively. Younger adults showed greater variability in rating the pen ($SD = 1.51$) compared to the mouse ($SD = 0.75$), where older adults had less variability with then pen ($SD = 0.72$) than the mouse ($SD = 1.09$).

RPE – mental. Younger adults provided ratings of perceived mental effort about the same as older adults, 1.44 compared to 1.97, corresponding to just less than “weak” on the Borg CR-10 scale. Overall ratings by device fell into a similar range, with an average rating of 1.90 for the pen, and 1.49 for the mouse. However, older adults rated the pen higher than the mouse, 2.23 (just above “weak”) compared to 1.71 (just below “weak”), whereas younger adults rated both devices about the same, 1.59 and 1.29 (both between “very weak” and “weak”) for the pen and mouse, respectively. Both younger and older adults had consistent variability in their ratings, with younger adults’ standard deviation of 1.09 for pen and 1.01 for mouse, and older adults’ standard deviation of .85 for pen and .79 for mouse. While ratings of perceived physical and mental effort all fell in relatively low ranges, typically between “very weak” and “weak”, younger adults

seemed to be more sensitive to physical differences between the devices, whereas older adults seemed to be more sensitive to mental differences.

Table 6. Descriptive statistics for Excel dragging task

Time			Pen	Mouse	Device
	Younger	M	39.67	29.25	34.46
		SD	9.28	5.14	6.50
		N	12	12	12
	Older	M	67.64	47.36	57.50
		SD	27.83	12.66	17.50
		N	11	11	11
	Age	M	53.04	37.91	
		SD	24.48	13.10	
		N	23	23	
Errors			Pen	Mouse	Device
	Younger	M	2.42	0.33	1.38
		SD	1.98	0.65	1.00
		N	12	12	12
	Older	M	4.64	1.45	3.05
		SD	4.30	1.70	2.36
		N	11	11	11
	Age	M	3.48	0.87	
		SD	3.41	1.36	
		N	23	23	
RPE Physical			Pen	Mouse	Device
	Younger	M	2.19	0.95	1.57
		SD	1.51	0.75	1.04
		N	12	12	12
	Older	M	1.26	1.48	1.37
		SD	0.72	1.09	0.53
		N	11	11	11
	Age	M	1.75	1.20	
		SD	1.27	0.95	
		N	23	23	
RPE Mental			Pen	Mouse	Device
	Younger	M	1.59	1.29	1.44
		SD	1.09	1.01	0.92
		N	12	12	12
	Older	M	2.23	1.71	1.97
		SD	0.85	0.79	0.63
		N	11	11	11
	Age	M	1.90	1.49	
		SD	1.01	0.91	
		N	23	23	

Inferential statistics

Refer to Table 7 for a complete summary of effects statistics for the Excel dragging task.

Table 7. Effects summary for Excel dragging task

		<i>F</i> (1, 19)	<i>p</i>	η^2
Time on Task	Age	17.18**	< .01	.47
	Device	15.34**	< .01	.45
	Age x Device	1.67	.21	.08
Errors	Age	4.50*	< .05	.19
	Device	11.84**	< .01	.38
	Age x Device	.49	.49	.02
RPE Physical	Age	.32	.58	.02
	Device	3.28^	.09	.14
	Age x Device	6.02*	< .05	.25
	Younger x Device	9.55**	< .01	.33
	Older x Device	.27	.61	.01
RPE Mental	Age	2.62	.12	.12
	Device	4.15^	.06	.18
	Age x Device	.38	.54	.02
<hr/>				
^ marginal significance, $p < .10$		**	significant, $p < .01$	
* significant, $p < .05$		***	significant, $p < .001$

Figure 22. Excel dragging task measures: (A) Time on task; (B) Number of errors; (C) Ratings of perceived physical effort; (D) Ratings of perceived mental effort. Error bars show +/- 1 standard error of the mean.

Time on task

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and time on task as the dependent variable. A main effect of age was found to be significant, $F(1, 19) = 17.18$, $p < .01$, $\eta^2 = .47$, indicating that across both devices, younger adults ($M = 34.46$,

$SD = 6.50$) were faster than older adults ($M = 57.50, SD = 17.50$) at performing the task. A main effect of device was also found to be significant, $F(1, 19) = 15.34, p < .01, \eta^2 = .45$, indicating that across all participants, performance time was slower when using the pen ($M = 53.04, SD = 24.48$) than when using the mouse ($M = 37.91, SD = 13.10$). The interaction between age and device was not significant, $F(1, 19) = 1.67, p = .21, \eta^2 = .08$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 22 (A).

Errors

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and number of errors as the dependent variable. A main effect of age was found to be significant, $F(1, 19) = 4.50, p < .05, \eta^2 = .19$, indicating that across both devices, younger adults ($M = 1.38, SD = 1.00$) made fewer errors than older adults ($M = 3.05, SD = 2.36$) when performing the task. A main effect of device was also found to be significant, $F(1, 19) = 11.84, p < .01, \eta^2 = .38$, indicating that across all participants, more errors were made when using the pen ($M = 3.48, SD = 3.41$) than when using the mouse ($M = .87, SD = 1.36$). The interaction between age and device was not significant, $F(1, 19) = .49, p = .49, \eta^2 = .02$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 22 (B).

RPE – physical

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and rating of perceived physical effort as the dependent variable. The main effect of age was not significant, $F(1, 19) = .32, p = .58, \eta^2 = .02$, indicating that across both devices, younger adults ($M = 1.57, SD = 1.04$) perceived the same amount of physical effort as older adults ($M = 1.37, SD = .53$) when performing the task. A main effect of device was marginal but not significant, $F(1, 19) = 3.28, p = .09, \eta^2 = .14$, indicating that across all participants, there was no difference in perceived physical effort when using the pen ($M = 1.75, SD = 1.27$) compared to using the mouse ($M = 1.20, SD = .95$). The interaction between age and device was significant, $F(1, 19) = 6.02, p < .05, \eta^2 = .25$, indicating that the direction and magnitude of device effects on ratings of perceived physical effort varied across age groups.

To determine the nature of the interaction between age and device found in the omnibus ANOVA, an analysis of the simple effects of the within-subjects factor (device) at each age group (younger, older) was conducted. Younger adults were found to rate higher perceived physical effort with the pen ($M = 2.19, SD = 1.51$) than the mouse ($M = .95, SD = .75$), $F(1, 19) = 9.25, p < .01, \eta^2 = .33$. Older adults did not rate the pen ($M = 1.26, SD = .72$) significantly different from the mouse ($M = 1.48, SD = 1.09$), $F(1, 19) = .27, p = .61, \eta^2 = .01$. Refer to Figure 22 (C).

RPE – mental

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and rating of perceived mental effort as the dependent variable. The main effect of age was not significant, $F(1, 19) = 2.62, p = .12, \eta^2 = .12$, indicating that across both devices, younger adults ($M = 1.44, SD = .92$) perceived the same amount of mental effort as older adults ($M = 1.97, SD = .63$) when performing the task. A main effect of device was marginal but not significant, $F(1, 19) = 4.15, p = .06, \eta^2 = .18$, indicating that across all participants, there was no difference in perceived mental effort when using the pen ($M = 1.90, SD = 1.01$) compared to using the mouse ($M = 1.49, SD = .91$). The interaction between age and device was not significant, $F(1, 19) = .38, p = .54, \eta^2 = .02$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 22 (D).

PowerPoint dragging task

Error analysis and data preparation

There were two types of slips that occurred in this task which were included in the error tally. The first was when a participant missed a target object. The second was when a participant dropped an object enroute to its intended drop point. See Figure 23 for examples of the error types.

Figure 23. Slip errors for PowerPoint dragging task

Participants were instructed to move the shapes into the gray box area so that none of the objects touched each other and none touched the outside border. If, at the end of the task, participants had objects touching either another object or the border of the box, it was not considered an important threat to the data, as long as the participant made an effort to position the shapes correctly.

Descriptive statistics

Refer to Table 8 for descriptive statistics of the PowerPoint dragging task.

Timing. Older adults appeared to perform slower overall – 128.00 seconds compared to 82.17 seconds for younger adults – but it took participants about the same time using the pen or the mouse, with overall time of 102.96 seconds for the pen and 107.21 seconds for the mouse. This similarity was apparent when looking at individual age groups as well - younger participants averaged 83.50 seconds with the pen and 80.83 seconds with the mouse, and older participants averaged 122.42 seconds with the pen and 133.58 seconds with the mouse. Variability in timing was similar for all conditions:

younger participants' time had standard deviations of 24.27 and 20.24 for pen and mouse, respectively, and for older participants, 32.18 and 45.18 for pen and mouse, respectively.

Errors. Older adults made more errors overall – between 1 and 2 ($M = 1.58$) compared to between 0 and 1 ($M = .71$) for younger adults – and participants generally made more errors using the pen than when using the mouse, with average overall error count of 1.83 for the pen and between 0 and 1 ($M = .46$) for the mouse. The average amount of errors was relatively low for both age groups; younger adults made on average 1.25 errors with the pen and 0.17 with the mouse, while older adults made 2.42 with the pen and .75 with the mouse, respectively. There was greater variability among older adults – standard deviations for older adults ranged from 2.50 with the pen to 1.14 with the mouse, compared to younger adults 1.22 with the pen and .39 with the mouse, respectively.

RPE – physical. Younger adults provided ratings of perceived physical effort slightly higher than older adults, 1.83 compared to 1.55, both however corresponding to just below “weak” on the Borg CR-10 scale. Overall ratings by device fell into a similar range, with an average rating of 1.85 for the pen, and 1.53 for the mouse. However, younger adults rated the pen higher than the mouse, 2.22 (“weak”) compared to 1.44 (between “very weak” and “weak”), whereas older adults rated both devices about the same, 1.48 and 1.63 (both between “very weak” and “weak”) for the pen and mouse, respectively. Younger adults showed greater variability in rating the pen ($SD = 1.65$) compared to the mouse ($SD = .94$), where older adults had about the same variability with then pen ($SD = 0.88$) and the mouse ($SD = .83$).

RPE – mental. Younger adults provided ratings of perceived mental effort about the same as older adults, 2.13 compared to 2.46, corresponding to slightly more than “weak” on the Borg CR-10 scale. Overall ratings by device fell into a similar range, with an average rating of 2.13 for the pen, and 2.47 for the mouse. Both younger and older adults rated performing the task with the pen as requiring less mental effort than with the mouse – 1.88 and 2.39 for younger adults, and 2.38 and 2.54 for older adults, for pen and mouse, respectively. Variability was consistent for younger adults, with standard deviation of 1.18 for the pen and 1.17 for the mouse, whereas older adults had more variability in rating perceived mental effort with the mouse ($SD = 1.76$) than with the pen ($SD = 1.17$).

Table 8. Descriptive statistics for PowerPoint dragging task

Time			Pen	Mouse	Device
	Younger	<i>M</i>	83.50	80.83	82.17
		<i>SD</i>	24.27	20.24	20.39
		<i>N</i>	12	12	12
	Older	<i>M</i>	122.42	133.58	128.00
		<i>SD</i>	32.18	45.18	32.04
		<i>N</i>	12	12	12
	Age	<i>M</i>	102.96	107.21	
		<i>SD</i>	34.23	43.56	
		<i>N</i>	24	24	
Errors			Pen	Mouse	Device
	Younger	<i>M</i>	1.25	0.17	0.71
		<i>SD</i>	1.22	0.39	0.62
		<i>N</i>	12	12	12
	Older	<i>M</i>	2.42	0.75	1.58
		<i>SD</i>	2.50	1.14	1.51
		<i>N</i>	12	12	12
	Age	<i>M</i>	1.83	0.46	
		<i>SD</i>	2.01	0.88	
		<i>N</i>	24	24	
RPE Physical			Pen	Mouse	Device
	Younger	<i>M</i>	2.22	1.44	1.83
		<i>SD</i>	1.65	0.94	1.06
		<i>N</i>	12	12	12
	Older	<i>M</i>	1.48	1.63	1.55
		<i>SD</i>	0.88	0.83	0.76
		<i>N</i>	12	12	12
	Age	<i>M</i>	1.85	1.53	
		<i>SD</i>	1.34	0.87	
		<i>N</i>	24	24	
RPE Mental			Pen	Mouse	Device
	Younger	<i>M</i>	1.88	2.39	2.13
		<i>SD</i>	1.18	1.17	1.09
		<i>N</i>	12	12	12
	Older	<i>M</i>	2.38	2.54	2.46
		<i>SD</i>	1.17	1.76	1.25
		<i>N</i>	12	12	12
	Age	<i>M</i>	2.13	2.47	
		<i>SD</i>	1.18	1.47	
		<i>N</i>	24	24	

Inferential statistics

Refer to Table 9 for a complete summary of effects statistics for the PowerPoint dragging task.

Table 9. Effects summary for PowerPoint dragging task

		<i>F</i> (1, 20)	<i>p</i>	η^2
Time on Task	Age	16.06**	< .01	.45
	Device	.59	.45	.03
	Age x Device	1.57	.22	.07
	Order x Device	13.70**	< .01	.41
	Pen first x Device	4.29*	< .05	.18
	Mouse first x Device	10.00**	< .01	.33
Time on Task (first device only)	Age	12.13**	< .01	.38
	Device	.05	.82	.00
	Age x Device	.63	.44	.03
Errors	Age	3.25^	.09	.14
	Device	12.07**	< .01	.38
	Age x Device	.54	.47	.03
RPE Physical	Age	.52	.48	.03
	Device	1.37	.26	.06
	Age x Device	2.87	.11	.13
RPE Mental	Age	.44	.51	.02
	Device	1.66	.21	.08
	Age x Device	.47	.50	.02
<hr/>				
^ marginal significance, $p < .10$		**	significant, $p < .01$	
* significant, $p < .05$		***	significant, $p < .001$

Figure 24. PowerPoint dragging task measures: (A) Time on task; (B) Number of errors; (C) Ratings of perceived physical effort; (D) Ratings of perceived mental effort. Error bars show +/- 1 standard error of the mean.

Time on task

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and time on task as the dependent variable. A main effect of age was found to be significant, $F(1, 20) = 16.06$,

$p < .01$, $\eta^2 = .45$, indicating that across both devices, younger adults ($M = 82.17$, $SD = 20.39$) were faster than older adults ($M = 128.00$, $SD = 32.04$) at performing the task. However, the main effect of device was not significant, $F(1, 20) = .59$, $p = .45$, $\eta^2 = .03$, indicating that across all participants, performance time was the same when using the pen ($M = 102.96$, $SD = 34.23$) as when using the mouse ($M = 107.21$, $SD = 43.56$). The interaction between age and device was not significant, $F(1, 20) = 1.57$, $p = .22$, $\eta^2 = .07$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 24 (A).

Looking at the factor of order, an interaction between order and device was found to be significant, $F(1, 20) = 13.70$, $p < .01$, $\eta^2 = .41$, indicating a differential carryover effect was present. To determine the nature of the interaction between order and device, an analysis of the simple effects of the within-subjects factor (device) at each order (pen first, mouse first) was conducted. Participants who used the pen first were found to complete the task slower with the pen ($M = 113.50$, $SD = 40.22$) than the mouse ($M = 97.33$, $SD = 33.02$), $F(1, 20) = 4.29$, $p < .05$, $\eta^2 = .18$. Similarly, participants who used the mouse first were slower with the mouse ($M = 117.08$, $SD = 51.62$) than with the pen ($M = 92.42$, $SD = 24.29$), $F(1, 20) = 10.00$, $p < .01$, $\eta^2 = .33$. Refer to Figure 25.

Figure 25. PowerPoint dragging task: Time on task for each device as a function of order. Error bars show +/- 1 standard error of the mean.

To test the research hypothesis and to compensate for differential carryover effects, a 2x2 between-subjects ANOVA was conducted with age group (younger, older) and device (pen, mouse) as the independent factors and time on task as the dependent factor. Only data from the first device used was included. A main effect of age was found to be significant, $F(1, 20) = 12.13, p < .01, \eta^2 = .38$, indicating that across both devices, younger adults ($M = 88.33, SD = 23.50$) performed faster than older adults ($M = 142.25, SD = 46.39$) when performing the task. The main effect of device was not significant, $F(1, 20) = .05, p = .82, \eta^2 = .00$, indicating that across all participants, time on task did not differ when using the pen ($M = 113.50, SD = 40.22$) or the mouse ($M = 117.08, SD = 51.62$). The interaction between age and device was not significant, $F(1, 20) = .63, p = .44, \eta^2 = .03$, indicating that the direction and magnitude of device effects

did not vary across age groups. The results concur with the results from the full data set with regard to significant effects. Refer to Figure 26.

Figure 26. PowerPoint dragging task: Time on task for each device as a function of age, limited to data from first device used. Error bars show +/- 1 standard error of the mean.

Errors

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and number of errors as the dependent variable. A main effect of age was marginal but not significant, $F(1, 20) = 3.25, p = .09, \eta^2 = .14$, indicating that across both devices, younger adults ($M = .71, SD = .62$) made the same amount of errors as older adults ($M = 1.58, SD = 1.51$) when performing the task. However, the main effect of device was significant, $F(1, 20) = 12.07, p < .01, \eta^2 = .38$, indicating that across all participants, more errors were made

when using the pen ($M = 1.83$, $SD = 2.01$) than when using the mouse ($M = .46$, $SD = .88$). The interaction between age and device was not significant, $F(1, 20) = .54$, $p = .47$, $\eta^2 = .03$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 24 (B).

RPE – physical

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and rating of perceived physical effort as the dependent variable. The main effect of age was not significant, $F(1, 20) = .52$, $p = .48$, $\eta^2 = .03$, indicating that across both devices, younger adults' perceived physical effort ($M = 1.83$, $SD = 1.06$) was not different from that of older adults ($M = 1.55$, $SD = .76$) when performing the task. The main effect of device was not significant, $F(1, 20) = 1.37$, $p = .26$, $\eta^2 = .06$, indicating that across all participants, perceived physical effort was no different using the pen ($M = 1.85$, $SD = 1.34$) from using the mouse ($M = 1.53$, $SD = .87$). The interaction between age and device was not significant, $F(1, 20) = 2.87$, $p = .11$, $\eta^2 = .13$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 24 (C).

RPE – mental

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and perceived mental effort as the dependent variable. The main effect of age was not significant, $F(1, 20) = .44$,

$p = .51$, $\eta^2 = .02$, indicating that across both devices, younger adults' perceived mental effort ($M = 2.13$, $SD = 1.09$) was not different from that of older adults ($M = 2.46$, $SD = 1.25$) when performing the task. The main effect of device was not significant, $F(1, 20) = 1.66$, $p = .21$, $\eta^2 = .08$, indicating that across all participants, perceived mental effort was no different using the pen ($M = 2.13$, $SD = 1.18$) than when using the mouse ($M = 2.47$, $SD = 1.47$). The interaction between age and device was not significant, $F(1, 20) = .47$, $p = .50$, $\eta^2 = .02$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 24 (D).

PowerPoint manipulation task

Error analysis and data preparation

Three types of slip errors occurred in this task: (1) selecting and using the wrong tool when attempting to manipulate the callout shape, (2) attempting to select but missing the tool, and (3) inadvertent double clicking. The numbers of these types of slips were summed to create the error score for the task. See Figure 27 for examples of the error types.

Figure 27. Slip errors for PowerPoint manipulation task

Another error type that occurred frequently was when a participant attempted to move the callout shape and selected text instead of the perimeter of the box. While this could be considered a slip, based on observation it was generally more likely that it was a mistake, i.e., the participant had trouble understanding how to manipulate the box, rather than had trouble doing it with the input device. Even though participants were shown this issue in the practice session it still proved vexing for most. Thus this error was not considered part of the error tally. Unfortunately it did contribute to timing differences on the task, as some participants made repeated attempts to perform this component of the task unsuccessfully. See Figure 28 for an example.

Figure 28. Text selection error (mistake) in PowerPoint manipulation task

Three participants failed the task and were omitted from the analysis. All three who failed (P4, P13, P17) were not able to complete the task in the pen condition in under 5 minutes. Also these participants had very high incidents of text selection errors, making their trials more fraught with mistakes than useful data on slips with the input devices.

Descriptive statistics

Refer to Table 10 for descriptive statistics of the PowerPoint manipulation task.

Timing. Older adults appeared to perform slower overall – 162.65 seconds compared to 87.82 seconds for younger adults – and participants were overall slower with the pen than with the mouse, with overall time of 141.10 seconds for the pen and 105.81 seconds for the mouse. When looking at individual age groups participants were slower with the pen than the mouse as well – younger participants averaged 98.36 seconds with the pen and 77.27 seconds with the mouse, and older participants averaged 188.10 seconds with the pen and 137.20 seconds with the mouse. Variability in timing was similar for younger participants, with standard deviations of time of 27.69 and 22.44 for pen and mouse, respectively. Older adults, however, exhibited much greater variability in the pen condition ($SD = 73.90$) than the mouse condition ($SD = 43.12$).

Errors. Older adults made more errors overall – about 3 ($M = 2.90$) compared to about 2 ($M = 1.95$) for younger adults – and participants generally made more errors using the pen than when using the mouse, with average overall error count of 4.14 for the pen and between 0 and 1 ($M = .67$) for the mouse. Younger adults made on average more errors with the pen ($M = 3.45$) than with the mouse ($M = .45$), as did older adults, who made 4.90 with the pen and .90 with the mouse, respectively. Patterns of variability were consistent between groups, with much higher standard deviations for the pen condition ($SD = 6.17$ for younger adults and 3.87 for older adults) than the mouse condition ($SD = .52$ for younger adults and .74 for older adults).

RPE – physical. Younger adults provided ratings of perceived physical effort slightly lower than older adults, 2.01 (corresponding to “weak” on the Borg CR-10 scale) compared to 2.80 (closer to “moderate”). Overall ratings by device indicated more

“moderate” physical effort for the pen ($M = 2.94$) and more “weak” physical effort for the pen ($M = 1.83$). Both younger and older adults provided higher perceived physical effort ratings for the pen than the mouse, and older adults rated higher physical effort for both. Younger adults on average rated 2.47 (between “weak” and “moderate”) for the pen and 1.55 (between “very weak” and “weak”) for the mouse. Older adults rated 3.45 (between “moderate” and “strong”) for the pen and 2.15 (just above “weak”) for the mouse. Variability was generally higher for the pen with standard deviation of ratings for younger adults of 1.48 for the pen and .69 for the mouse, and older adults’ standard deviation of ratings of 2.23 for the pen and 1.13 for the mouse.

RPE – mental. Younger adults provided lower ratings of perceived mental effort than older adults, 2.35 compared to 3.24, corresponding to slightly more than “weak” on the Borg CR-10 scale for the younger adults and slightly more than “moderate” for the older adults. Perceived mental effort ratings by device were similar, with 3.14 for the pen, and 2.40 for the mouse. Younger adults rated performing the task as having a similar range of mental effort for both devices, 2.52 for the pen, and 2.17 for the mouse, both between “weak” and “moderate” on the Borg CR-10 scale. Older adults indicated higher mental effort for the pen ($M = 3.82$, close to “strong”) than for the mouse ($M = 2.65$, between “weak” and “moderate”). With younger adults, there was more variability for the pen than the mouse, with standard deviation of 1.35 for the pen and .95 for the mouse. Similarly, older adults had more variability in rating perceived mental effort with the pen ($SD = 2.44$) than with the mouse ($SD = 1.75$).

Table 10. Descriptive statistics for PowerPoint manipulation task

Time			Pen	Mouse	Device
	Younger	<i>M</i>	98.36	77.27	87.82
		<i>SD</i>	27.69	22.44	18.47
		<i>N</i>	11	11	11
	Older	<i>M</i>	188.10	137.20	162.65
		<i>SD</i>	73.90	43.12	49.80
		<i>N</i>	10	10	10
	Age	<i>M</i>	141.10	105.81	
		<i>SD</i>	70.36	45.04	
		<i>N</i>	21	21	
Errors			Pen	Mouse	Device
	Younger	<i>M</i>	3.45	0.45	1.95
		<i>SD</i>	6.17	0.52	2.99
		<i>N</i>	11	11	11
	Older	<i>M</i>	4.90	0.90	2.90
		<i>SD</i>	3.87	0.74	2.02
		<i>N</i>	10	10	10
	Age	<i>M</i>	4.14	0.67	
		<i>SD</i>	5.13	0.66	
		<i>N</i>	21	21	
RPE Physical			Pen	Mouse	Device
	Younger	<i>M</i>	2.47	1.55	2.01
		<i>SD</i>	1.48	0.69	1.04
		<i>N</i>	11	11	11
	Older	<i>M</i>	3.45	2.15	2.80
		<i>SD</i>	2.23	1.13	1.61
		<i>N</i>	10	10	10
	Age	<i>M</i>	2.94	1.83	
		<i>SD</i>	1.89	0.95	
		<i>N</i>	21	21	
RPE Mental			Pen	Mouse	Device
	Younger	<i>M</i>	2.52	2.17	2.35
		<i>SD</i>	1.35	0.95	1.06
		<i>N</i>	11	11	11
	Older	<i>M</i>	3.82	2.65	3.24
		<i>SD</i>	2.44	1.75	1.94
		<i>N</i>	10	10	10
	Age	<i>M</i>	3.14	2.40	
		<i>SD</i>	2.01	1.37	
		<i>N</i>	21	21	

Inferential statistics

Refer to Table 11 for a complete summary of effects statistics for the PowerPoint manipulation task.

Table 11. Effects summary for PowerPoint manipulation task

		<i>F</i> (1, 17)	<i>p</i>	η^2
Time on Task	Age	18.97***	< .001	.53
	Device	12.05**	< .01	.41
	Age x Device	2.03	.17	.11
Errors	Age	.45	.51	.03
	Device	9.08**	< .01	.35
	Age x Device	.06	.81	.00
RPE Physical	Age	1.49	.24	.08
	Device	18.42***	< .001	.52
	Age x Device	.78	.39	.04
RPE Mental	Age	1.45	.25	.08
	Device	13.40**	< .01	.44
	Age x Device	4.24^	.06	.20
	Order x Device	11.48**	< .01	.40
	Pen first x Device	19.63***	< .001	.54
	Mouse first x Device	.04	.84	.00
RPE Mental (first device only)	Age	2.72	.12	.14
	Device	1.63	.22	.09
	Age x Device	.22	.65	.01
<hr/>				
^ marginal significance, $p < .10$		**	significant, $p < .01$	
* significant, $p < .05$		***	significant, $p < .001$

Figure 29. PowerPoint manipulation task measures: (A) Time on task; (B) Number of errors; (C) Ratings of perceived physical effort; (D) Ratings of perceived mental effort. Error bars show ± 1 standard error of the mean.

Time on task

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and time on task as the dependent variable. A main effect of age was found to be significant, $F(1, 17) = 18.97$, $p < .001$, $\eta^2 = .53$, indicating that across both devices, younger adults ($M = 87.82$,

$SD = 18.47$) were faster than older adults ($M = 162.65$, $SD = 49.80$) at performing the task. The main effect of device was also significant, $F(1, 17) = 12.05$, $p < .01$, $\eta^2 = .41$, indicating that across all participants, performance time was slower when using the pen ($M = 141.10$, $SD = 70.36$) than when using the mouse ($M = 105.81$, $SD = 45.04$). The interaction between age and device was not significant, $F(1, 17) = 2.03$, $p = .17$, $\eta^2 = .11$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 29 (A).

Errors

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and number of errors as the dependent variable. The main effect of age was not significant, $F(1, 17) = .45$, $p = .51$, $\eta^2 = .03$, indicating that across both devices, there was no difference in number of errors made by younger adults ($M = 1.95$, $SD = 2.99$) compared to older adults ($M = 2.90$, $SD = 2.02$). The main effect of device was significant, $F(1, 17) = 9.08$, $p < .01$, $\eta^2 = .35$, indicating that across all participants, number of errors was higher when using the pen ($M = 4.14$, $SD = 5.13$) than when using the mouse ($M = .67$, $SD = .66$). The interaction between age and device was not significant, $F(1, 17) = .06$, $p = .81$, $\eta^2 = .00$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 29 (B).

RPE – physical

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and ratings of perceived physical effort as the dependent variable. The main effect of age was not significant, $F(1, 17) = 1.49, p = .24, \eta^2 = .08$, indicating that across both devices, there was no difference in perceived physical effort by younger adults ($M = 2.01, SD = 1.04$) compared to older adults ($M = 2.80, SD = 1.61$). The main effect of device was significant, $F(1, 17) = 18.42, p < .001, \eta^2 = .52$, indicating that across all participants, perceived physical effort was higher when using the pen ($M = 2.94, SD = 1.89$) than when using the mouse ($M = 1.83, SD = .95$). The interaction between age and device was not significant, $F(1, 17) = .78, p = .39, \eta^2 = .04$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 29 (C).

RPE – mental

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and ratings of perceived mental effort as the dependent variable. The main effect of age was not significant, $F(1, 17) = 1.45, p = .25, \eta^2 = .08$, indicating that across both devices, there was no difference in perceived mental effort by younger adults ($M = 2.35, SD = 1.06$) when compared to older adults ($M = 3.24, SD = 1.94$). The main effect of device was significant, $F(1, 17) = 13.40, p < .01, \eta^2 = .44$, indicating that across all participants, perceived mental effort

was higher when using the pen ($M = 3.14$, $SD = 2.01$) than when using the mouse ($M = 2.40$, $SD = 1.37$). The interaction between age and device was marginal but not significant, $F(1, 17) = 4.24$, $p = .06$, $\eta^2 = .20$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 29 (D).

Looking at the factor of order, an interaction between order and device was found to be significant, $F(1, 17) = 11.48$, $p < .01$, $\eta^2 = .40$, indicating a differential carryover effect was present. To determine the nature of the interaction between order and device, an analysis of the simple effects of the within-subjects factor (device) at each order (pen first, mouse first) was conducted. Participants who used the pen first perceived higher mental effort with the pen ($M = 3.58$, $SD = 1.82$) than with the mouse ($M = 1.94$, $SD = 1.04$), $F(1, 17) = 19.63$, $p < .001$, $\eta^2 = .54$. However, participants who used the mouse first did not perceive significantly different mental effort with the pen ($M = 2.81$, $SD = 2.15$) or the mouse ($M = 2.74$, $SD = 1.53$), $F(1, 17) = .04$, $p = .84$, $\eta^2 = .00$. Refer to Figure 30.

Figure 30. Ratings of perceived mental effort for PowerPoint manipulation task for each device as a function of device order. Error bars show +/- 1 standard error of the mean.

To test the research hypothesis and to compensate for differential carryover effects, a 2x2 between-subjects ANOVA was conducted with age group (younger, older) and device (pen, mouse) as the independent factors, and perceived mental effort as the dependent factor. Only data from the first device used was included. The main effect of age was not significant, $F(1, 17) = 2.72, p = .12, \eta^2 = .14$, indicating that across both devices, perceived mental effort was the same for younger adults ($M = 2.58, SD = 1.22$) as for older adults ($M = 3.67, SD = 1.97$). The main effect of device was not significant, $F(1, 17) = 1.63, p = .22, \eta^2 = .09$, indicating that across all participants, perceived mental effort did not differ when using the pen ($M = 3.58, SD = 1.82$) or the mouse ($M = 2.74, SD = 1.53$). The interaction between age and device was not significant, $F(1, 17) = .22, p = .65, \eta^2 = .01$, indicating that the direction and magnitude of device effects did not

vary across age groups. The results of this analysis differ from the previous one in that the main effect of device is no longer significant. Refer to Figure 31.

Figure 31. Ratings of perceived mental effort for PowerPoint manipulation task for each device as a function of age, limited to first device used. Error bars show +/- 1 standard error of the mean.

Tracing task

Error analysis and data preparation

To evaluate errors on the tracing task, three elements were recorded for each resulting image: the number of gaps in the tracing (i.e., every time white space was visible between the tracing image and the participant's line), size of the gaps, rated from 1 to 3, with 1 being small (alongside the tracing line), 2 being medium (a bump or more noticeable gap) and 3 being large (way off). The final measure was a subjective rating of the line and tracing quality from 1 to 5 with 1 being smooth and well traced, and 5 being rough and poorly traced (see Figure 32). An error/quality metric was derived from these

three values by dividing the number of gaps by the inverse of the gap size (so a large number of small sized gaps would not be penalized as much as a large number of larger gaps), and adding the subjective rating. Thus, the best attainable score would be 1, with higher numbers representing worse scores.

Figure 32. Examples of tracing images

Descriptive statistics

Refer to Table 12 for descriptive statistics of the tracing task.

Timing. Overall both age groups performed the task in similar time – 44.50 seconds for younger adults compared to 50.21 seconds for older adults. However, participants were overall faster with the pen than with the mouse, with overall time of 37.71 seconds for the pen and 57 seconds for the mouse. When looking at individual age groups participants were faster with the pen than the mouse as well – younger participants averaged 36.58 seconds with the pen and 52.42 seconds with the mouse, and older participants averaged 38.83 seconds with the pen and 61.58 seconds with the

mouse. Variability in timing was similar across both conditions for older adults, with standard deviations of time of 17.06 and 19.52 for pen and mouse, respectively. Younger adults, however, exhibited greater variability in the mouse condition ($SD = 25.55$) than the pen condition ($SD = 12.71$).

Errors. The error measure for the tracing task was somewhat higher (i.e., more and larger gaps and poorer tracing quality) for older adults ($M = 6.58$) compared to younger adults ($M = 4.93$) – and participants had a lower error measure using the pen than when using the mouse, with average overall score of 3.56 for the pen and 7.95 for the mouse. Both younger and older adults had higher error scores with the mouse than with the pen – younger adults scored 2.96 and 6.90 with the pen and mouse, respectively, and older adults scored 4.15 and 9 with the pen and mouse, respectively. Variability was lowest for young adults with the pen ($SD = 1.39$) and highest for older adults with the mouse ($SD = 3.25$), and about the same for young adults with the mouse ($SD = 2.37$), and older adults with the pen ($SD = 2.30$).

RPE – physical. Ratings of perceived physical effort were about the same for both younger and older adults overall, 2.56 and 2.73, respectively, corresponding to between “weak” and “moderate” on the Borg CR-10 scale. However, overall ratings by device were higher for the mouse – 3.66, corresponding to between “moderate” and “strong” – than for the pen – 1.63, corresponding to between “very weak” and “weak”. Both younger and older adults rated lower perceived physical effort working with the pen than the mouse. Younger adults’ average rating was 1.68 for the pen and 3.44 for the mouse, and older adults were similarly 1.58 for the pen and 3.88 for the mouse.

Variability of ratings was fairly consistent, although older adults' standard deviation was higher for the mouse at 2.13 than the pen at 1.04. Younger adults' standard deviations of ratings were 1.03 for the pen and 1.51 for the mouse.

RPE – mental. Younger adults provided somewhat lower ratings of perceived mental effort than older adults, 2.15 compared to 2.72, corresponding to slightly more than “weak” on the Borg CR-10 scale for the younger adults and slightly less than “moderate” for the older adults. A greater difference was reported across all participants between devices, with 1.52 for the pen (between “very weak” and “weak”) and 3.35 for the mouse (slightly above “moderate”). Both younger and older adults rated lower perceived mental effort working with the pen than the mouse. Younger adults' average rating was 1.39 for the pen and 2.92 for the mouse, and older adults were similarly 1.65 for the pen and 3.79 for the mouse. Variability of ratings was higher for the mouse, with younger adults' standard deviation of rating for the mouse of 2.01 and older adults, 1.85. For the pen, younger adults' ratings had a standard deviation of .87, and older adults, 1.07.

Table 12. Descriptive statistics for tracing task

Time			Pen	Mouse	Device
	Younger	<i>M</i>	36.58	52.42	44.50
		<i>SD</i>	12.71	25.55	18.32
		<i>N</i>	12	12	12
	Older	<i>M</i>	38.83	61.58	50.21
		<i>SD</i>	17.06	19.52	15.62
		<i>N</i>	12	12	12
	Age	<i>M</i>	37.71	57.00	
		<i>SD</i>	14.75	22.73	
		<i>N</i>	24	24	
Errors			Pen	Mouse	Device
	Younger	<i>M</i>	2.96	6.90	4.93
		<i>SD</i>	1.39	2.37	1.69
		<i>N</i>	12	12	12
	Older	<i>M</i>	4.15	9.00	6.58
		<i>SD</i>	2.30	3.25	2.33
		<i>N</i>	12	12	12
	Age	<i>M</i>	3.56	7.95	
		<i>SD</i>	1.95	2.98	
		<i>N</i>	24	24	
RPE Physical			Pen	Mouse	Device
	Younger	<i>M</i>	1.68	3.44	2.56
		<i>SD</i>	1.03	1.51	1.06
		<i>N</i>	12	12	12
	Older	<i>M</i>	1.58	3.88	2.73
		<i>SD</i>	1.04	2.13	1.19
		<i>N</i>	12	12	12
	Age	<i>M</i>	1.63	3.66	
		<i>SD</i>	1.01	1.82	
		<i>N</i>	24	24	
RPE Mental			Pen	Mouse	Device
	Younger	<i>M</i>	1.39	2.92	2.15
		<i>SD</i>	0.87	2.01	1.37
		<i>N</i>	12	12	12
	Older	<i>M</i>	1.65	3.79	2.72
		<i>SD</i>	1.07	1.85	1.16
		<i>N</i>	12	12	12
	Age	<i>M</i>	1.52	3.35	
		<i>SD</i>	0.96	1.94	
		<i>N</i>	24	24	

Inferential statistics

Refer to Table 13 for a complete summary of effects statistics for the tracing task.

Table 13. Effects summary for tracing task

		<i>F</i> (1, 20)	<i>p</i>	η^2
Time on Task	Age	.94	.35	.04
	Device	27.28***	< .001	.58
	Age x Device	.88	.36	.04
	Order	10.33**	< .01	.34
Errors	Age	4.68*	< .05	.19
	Device	63.02***	< .001	.76
	Age x Device	.67	.42	.03
	Order	6.22*	< .05	.24
RPE Physical	Age	.13	.72	.01
	Device	28.18***	< .001	.58
	Age x Device	.49	.49	.02
RPE Mental	Age	1.25	.28	.06
	Device	30.32***	< .001	.60
	Age x Device	.86	.37	.04
	Order x Age x Device	4.42*	< .05	.18
		<i>F</i> (1, 10)	<i>p</i>	η^2
Younger adults	Pen first x Device	16.52**	< .01	.62
	Mouse first x Device	2.47	.15	.20
Older adults	Pen first x Device	3.38	.10	.25
	Mouse first x Device	13.83**	< .01	.58
		<i>F</i> (1, 20)	<i>p</i>	η^2
RPE Mental (first device only)	Age	4.03^	.06	.17
	Device	7.12*	< .05	.26
	Age x Device	3.32^	.08	.14
^		marginal significance, $p < .10$	**	significant, $p < .01$
*		significant, $p < .05$	***	significant, $p < .001$

Figure 33. Tracing task measures: (A) Time on task; (B) Number of errors; (C) Ratings of perceived physical effort; (D) Ratings of perceived mental effort. Error bars show ± 1 standard error of the mean.

Time on task

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and time on task as the dependent variable. A main effect of age was not significant, $F(1, 20) = .94, p = .35$, $\eta^2 = .04$, indicating that across both devices, time on task for younger adults ($M = 44.50$,

$SD = 18.32$) did not differ from older adults ($M = 50.21$, $SD = 15.62$). However, the main effect of device was significant, $F(1, 20) = 27.28$, $p < .001$, $\eta^2 = .58$, indicating that across all participants, performance time was faster when using the pen ($M = 37.71$, $SD = 14.75$) than when using the mouse ($M = 57.00$, $SD = 22.73$). The interaction between age and device was not significant, $F(1, 20) = .88$, $p = .36$, $\eta^2 = .04$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 33 (A).

Looking at the order factor, the main effect of order was found to be significant, $F(1, 20) = 10.33$, $p < .01$, $\eta^2 = .34$, indicating that across both devices, performance time was faster when the pen was first ($M = 37.87$, $SD = 9.71$) than when the mouse was first ($M = 56.83$, $SD = 17.52$). Because the experimental design was counterbalanced, this effect was not considered a threat to the interpretation of the results.

Errors

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and error score as the dependent variable. The main effect of age was significant, $F(1, 20) = 4.68$, $p < .05$, $\eta^2 = .19$, indicating that across both devices, younger adults ($M = 4.93$, $SD = 1.69$) had lower error scores than older adults ($M = 6.58$, $SD = 2.33$). The main effect of device was also significant, $F(1, 20) = 63.02$, $p < .001$, $\eta^2 = .76$, indicating that across all participants, error score was lower when using the pen ($M = 3.56$, $SD = 1.95$) than when using the mouse ($M = 7.95$, $SD = 2.98$). The interaction between age and device was not

significant, $F(1, 20) = .67, p = .42, \eta^2 = .03$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 33 (B).

Looking at the order factor, the main effect of order was found to be significant, $F(1, 20) = 6.22, p < .05, \eta^2 = .24$, indicating that across both devices, error score was higher when the pen was first ($M = 6.70, SD = 2.11$) than when the mouse was first ($M = 4.80, SD = 1.84$). Because the experimental design was counterbalanced, this effect was not considered a threat to the interpretation of the results.

RPE – physical

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and ratings of perceived physical effort as the dependent variable. The main effect of age was not significant, $F(1, 20) = .13, p = .72, \eta^2 = .01$, indicating that across both devices, there was no difference in perceived physical effort by younger adults ($M = 2.56, SD = 1.06$) compared to older adults ($M = 2.73, SD = 1.19$). The main effect of device was significant, $F(1, 20) = 28.18, p < .001, \eta^2 = .58$, indicating that across all participants, perceived physical effort was lower when using the pen ($M = 1.63, SD = 1.01$) than when using the mouse ($M = 3.66, SD = 1.82$). The interaction between age and device was not significant, $F(1, 20) = .49, p = .49, \eta^2 = .02$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 33 (C).

RPE – mental

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and ratings of perceived mental effort as the dependent variable. The main effect of age was not significant, $F(1, 20) = 1.25, p = .28, \eta^2 = .06$, indicating that across both devices, there was no difference in perceived mental effort by younger adults ($M = 2.15, SD = 1.37$) compared to older adults ($M = 2.72, SD = 1.16$). The main effect of device was significant, $F(1, 20) = 30.32, p < .001, \eta^2 = .60$, indicating that across all participants, perceived mental effort was lower when using the pen ($M = 1.52, SD = .96$) than when using the mouse ($M = 3.35, SD = 1.94$). The interaction between age and device was not significant, $F(1, 20) = .86, p = .37, \eta^2 = .04$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 33 (D).

Looking at the factor of order, a three-way interaction between order, age, and device was significant, $F(1, 20) = 4.42, p < .05, \eta^2 = .18$, indicating that differential carryover effects varied by age group. When exposed to the pen first, younger adults provided significantly lower ratings of perceived mental effort for the pen ($M = 1.55, SD = 1.02$) than for the mouse ($M = 3.75, SD = 2.21$), $F(1, 10) = 16.52, p < .01, \eta^2 = .62$. However, when exposed to the mouse first, younger adults' ratings of perceived mental effort was not significantly different for the pen ($M = 1.23, SD = .75$) or the mouse ($M = 2.08, SD = 1.53$), $F(1, 10) = 2.47, p = .15, \eta^2 = .20$. Conversely, when older adults were exposed to the pen first, they did not provide significantly different ratings of perceived

mental effort for the pen ($M = 1.67$, $SD = 1.03$) or the mouse ($M = 3.08$, $SD = 1.11$), $F(1, 10) = 3.38$, $p = .10$, $\eta^2 = .25$. However, when older adults were exposed to the mouse first, they rated the pen ($M = 1.63$, $SD = 1.21$) significantly lower than the mouse ($M = 4.50$, $SD = 2.26$), $F(1, 10) = 13.83$, $p < .01$, $\eta^2 = .58$. Refer to Figure 34.

Figure 34. Ratings of perceived mental effort for tracing task - 3-way interaction of device, age, and order. Error bars show ± 1 standard error of the mean.

To test the research hypothesis and to compensate for differential carryover effects, a 2x2 between-subjects ANOVA was conducted with age group (younger, older) and device (pen, mouse) as the independent factors and ratings of perceived mental effort as the dependent factor. Only data from the first device used was included. A main effect of age was marginal but not significant, $F(1, 20) = 4.03$, $p = .06$, $\eta^2 = .17$, indicating that across both devices, there was no difference in perceived mental effort between younger adults ($M = 1.82$, $SD = 1.27$) and older adults ($M = 3.08$, $SD = 2.23$).

The main effect of device was significant, $F(1, 20) = 7.12, p < .05, \eta^2 = .26$, indicating that across all participants, perceived mental effort was lower when using the pen ($M = 1.61, SD = .98$) than when using the mouse ($M = 3.29, SD = 2.23$). The interaction between age and device was not marginal but not significant, $F(1, 20) = 3.32, p = .08, \eta^2 = .14$, indicating that the direction and magnitude of device effects did not vary across age groups. The results concur with the results from the full data set with regard to significant effects. Refer to Figure 35.

Figure 35. Ratings of perceived mental effort for tracing task for both devices as a function of age, limited to first device used. Error bars show +/- 1 standard error of the mean.

Cross-referencing “form filling” task

Error analysis and data preparation

Slip errors were noted when a participant attempted to hit a checkbox or radio button and missed. Also slips were noted with the drop menus, by counting when a participant selected a wrong value and immediately changed it. In some cases a participant selected a wrong value and immediately changed it. In some cases a participant made a mistake, in that they set a value wrong (i.e., they made a mistake) but they were able to set the value they initially intended. These instances were not counted as slips.

Errors were also inferred by evaluating the accuracy of the data entered. However, for participants with more than 3 errors, a closer examination of their data was carried out, to make sure the problem was not one of mistaking the form data. In several cases the participants’ inaccuracy counts were reduced based on review of the DVD tape and analysis of the pattern of error. The observed slips in using the device were added to the inaccuracies (inferred slips) to create the final error count for the task. Examples of the observed slips are shown in Figure 36.

Missing a form widget	Resetting a drop-menu

Figure 36. Examples of slip errors in the cross-referencing “form filling” task

Descriptive statistics

Refer to Table 14 for descriptive statistics of the cross-referencing “form filling” task.

Timing. Overall younger adults were faster than older adults, taking on average 110.54 to complete the task compared to 186.17 for older adults. However, participants took about the same amount of time with the pen as with the mouse, with overall time of 155.04 seconds for the pen and 141.67 seconds for the mouse. Within age groups, both younger and older adults were only slightly slower with the pen than the mouse. Younger adults took 120.83 seconds with the pen and 100.25 seconds with the mouse, while older adults took 189.25 seconds with the pen and 183.08 with the mouse. Timing was less variable with younger adults, whose standard deviation of time was 34.10 for the pen and 22.46 for the mouse; older adults’ standard deviation of time was 49.20 with the pen and 43.15 with the mouse.

Errors. Younger adults made on average between 1 and 2 ($M = 1.58$) errors, while older adults made between 2 and 3 ($M = 2.63$). Across both age groups, participants made close to 3 errors with the pen ($M = 2.92$) compared to closer to 1 with the mouse ($M = 1.29$). Both younger and older adults made more errors with the pen than with the mouse, but older adults appeared to have a wider range of errors – younger adults made on average 1.83 and 1.33 errors with the pen and mouse, respectively, whereas older adults made on average 4.00 and 1.25 with the pen and mouse, respectively. Older adults using the pen had a high amount of variability in their error tally, with a standard deviation of 4.13, compared to 1.14 for the mouse. Younger adults’

error tallies were more consistently variable, with standard deviation of 1.75 and 1.44 for the pen and mouse, respectively.

RPE – physical. Ratings of perceived physical effort were about the same for both younger and older adults overall, 1.82 and 1.83, respectively, corresponding to just below “weak” on the Borg CR-10 scale. Overall ratings by device were somewhat higher for the pen – 1.97, corresponding to “weak” – than for the mouse – 1.68, corresponding to a bit less than “weak”. Both younger and older adults rated slightly higher perceived physical effort working with the pen than the mouse. Younger adults’ average rating was 1.98 for the pen and 1.65 for the mouse, and older adults were similarly 1.95 for the pen and 1.71 for the mouse. Variability of ratings was fairly consistent, with standard deviations for younger adults of 1.08 and .97 for the pen and mouse, respectively. Similarly, older adults’ standard deviations of ratings were 1.39 for the pen and .94 for the mouse.

RPE – mental. Younger adults provided slightly lower ratings of perceived mental effort than older adults, 2.70 compared to 3.03, corresponding to a bit below “moderate” on the Borg CR-10 scale for the younger adults and “moderate” for older adults. Across all participants, performing the task with the pen was rated slightly lower than with the mouse – 2.75 for the pen (below “moderate”) and 2.99 for the mouse (“moderate”). Younger adults rated pen and mouse equally for perceived mental effort, whereas older adults rated the pen lower than the mouse. Younger adults’ average ratings were 2.68 for the pen and 2.72 for the mouse, and older adults’ ratings were 2.82 for the pen and 3.25 for the mouse. Variability of ratings for younger adults was

consistent, with standard deviations of 1.40 and 1.34 for pen and mouse, respectively.

Older adults had slightly less variability in their ratings of the pen, with standard deviations of 1.30 for the pen, and 1.57 for the mouse.

Table 14. Descriptive statistics for cross-referencing “form filling” task

Time			Pen	Mouse	Device
	Younger	<i>M</i>	120.83	100.25	110.54
		<i>SD</i>	34.10	22.46	26.70
		<i>N</i>	12	12	12
	Older	<i>M</i>	189.25	183.08	186.17
		<i>SD</i>	49.20	43.15	38.67
		<i>N</i>	12	12	12
	Age	<i>M</i>	155.04	141.67	
		<i>SD</i>	54.17	54.05	
		<i>N</i>	24	24	
Errors			Pen	Mouse	Device
	Younger	<i>M</i>	1.83	1.33	1.58
		<i>SD</i>	1.75	1.44	0.76
		<i>N</i>	12	12	12
	Older	<i>M</i>	4.00	1.25	2.63
		<i>SD</i>	4.13	1.14	2.07
		<i>N</i>	12	12	12
	Age	<i>M</i>	2.92	1.29	
		<i>SD</i>	3.30	1.27	
		<i>N</i>	24	24	
RPE Physical			Pen	Mouse	Device
	Younger	<i>M</i>	1.98	1.65	1.82
		<i>SD</i>	1.08	0.97	0.77
		<i>N</i>	12	12	12
	Older	<i>M</i>	1.95	1.71	1.83
		<i>SD</i>	1.39	0.94	1.04
		<i>N</i>	12	12	12
	Age	<i>M</i>	1.97	1.68	
		<i>SD</i>	1.22	0.93	
		<i>N</i>	24	24	
RPE Mental			Pen	Mouse	Device
	Younger	<i>M</i>	2.68	2.72	2.70
		<i>SD</i>	1.40	1.34	1.28
		<i>N</i>	12	12	12
	Older	<i>M</i>	2.82	3.25	3.03
		<i>SD</i>	1.30	1.57	1.32
		<i>N</i>	12	12	12
	Age	<i>M</i>	2.75	2.99	
		<i>SD</i>	1.33	1.45	
		<i>N</i>	24	24	

Inferential statistics

Refer to Table 15 for a complete summary of effects statistics for the cross-referencing “form filling” task.

Table 15. Effects summary for cross-referencing “form filling” task

		<i>F</i> (1, 20)	<i>p</i>	η^2
Time on Task	Age	29.22***	< .001	.59
	Device	5.20*	< .05	.21
	Age x Device	1.51	.23	.07
	Order x Device	17.73***	< .001	.47
	Pen first x Device	21.06***	< .001	.51
	Mouse first x Device	1.87	.19	.09
Time on Task (first device only)	Age	28.97***	< .001	.59
	Device	.67	.42	.03
	Age x Device	1.38	.25	.06
Errors	Age	2.49	.13	.11
	Device	4.30^	.05	.17
	Age x Device	2.06	.17	.09
RPE Physical	Age	.00	.98	.00
	Device	1.48	.24	.07
	Age x Device	.04	.85	.00
	Order x Device	6.01*	< .05	.23
	Pen first x Device	6.73*	< .05	.25
	Mouse first x Device	.76	.39	.04
RPE Physical (first device only)	Age	.02	.88	.00
	Device	1.15	.30	.05
	Age x Device	.05	.83	.00
RPE Mental	Age	.36	.56	.02
	Device	1.31	.27	.06
	Age x Device	.89	.36	.04

^ marginal significance, $p < .10$ ** significant, $p < .01$

* significant, $p < .05$ *** significant, $p < .001$

Figure 37. Cross-referencing "form filling" task measures: (A) Time on task; (B) Number of errors; (C) Ratings of perceived physical effort; (D) Ratings of perceived mental effort. Error bars show ± 1 standard error of the mean.

Time on task

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and time on task as the dependent variable. The main effect of age was significant, $F(1, 20) = 29.22, p < .001$, $\eta^2 = .59$, indicating that across both devices, younger adults ($M = 110.54, SD = 26.70$)

took less time than older adults ($M = 186.17$, $SD = 38.67$) to complete the task.

Additionally, the main effect of device was significant, $F(1, 20) = 5.20$, $p < .05$, $\eta^2 = .21$, indicating that across all participants, performance time was slower when using the pen ($M = 155.04$, $SD = 54.17$) than when using the mouse ($M = 141.67$, $SD = 54.05$). The interaction between age and device was not significant, $F(1, 20) = 1.51$, $p = .23$, $\eta^2 = .07$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 37 (A).

Looking at the factor of order, an interaction between order and device was found to be significant, $F(1, 20) = 17.73$, $p < .001$, $\eta^2 = .47$, indicating a differential carryover effect was present. To determine the nature of the interaction between order and device, an analysis of the simple effects of the within-subjects factor (device) at each order (pen first, mouse first) was conducted. Participants who used the pen first were slower with the pen ($M = 167.25$, $SD = 59.12$) than with the mouse ($M = 129.17$, $SD = 47.59$), $F(1, 20) = 21.06$, $p < .001$, $\eta^2 = .51$. However, participants who used the mouse first did not differ in time on task with the pen ($M = 142.83$, $SD = 48.12$) or the mouse ($M = 154.17$, $SD = 59.19$), $F(1, 20) = 1.87$, $p = .19$, $\eta^2 = .09$. Refer to Figure 38.

Figure 38. Time on task for form filling task for each device as a function of order. Error bars show +/- 1 standard error of the mean.

To test the research hypothesis and to compensate for differential carryover effects, a 2x2 between-subjects ANOVA was conducted with age group (younger, older) and device (pen, mouse) as the independent factors and time on task as the dependent factor. Only data from the first device used was included. The main effect of age was significant, $F(1, 20) = 28.97, p < .001, \eta^2 = .59$, indicating that across both devices, time on task was shorter for younger adults ($M = 117.75, SD = 34.56$) than for older adults ($M = 203.67, SD = 43.25$). The main effect of device was not significant, $F(1, 20) = .67, p = .42, \eta^2 = .03$, indicating that across all participants, time on task did not differ when using the pen ($M = 167.25, SD = 59.12$) or the mouse ($M = 154.17, SD = 59.19$). The interaction between age and device was not significant, $F(1, 20) = 1.38, p = .25, \eta^2 = .06$, indicating that the direction and magnitude of device effects did not vary across age

groups. The results of this analysis differ from the previous one in that the main effect of device is no longer significant. Refer to Figure 39.

Cross-ref. "form filling" - Time on task, limited to first device

Figure 39. Time on task for form filling task for each device as a function of age, limited to first device used. Error bars show +/- 1 standard error of the mean.

Errors

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and number of errors as the dependent variable. The main effect of age was not significant, $F(1, 20) = 2.49, p = .13, \eta^2 = .11$, indicating that across both devices, there was no difference in number of errors made by younger adults ($M = 1.58, SD = .76$) compared to older adults ($M = 2.63, SD = 2.07$). The main effect of device was marginal but not significant, $F(1, 20) = 4.30, p = .05, \eta^2 = .17$, indicating that across all participants, number of errors did not differ

when using the pen ($M = 2.92$, $SD = 3.30$) or the mouse ($M = 1.29$, $SD = 1.27$). The interaction between age and device was not significant, $F(1, 20) = 2.06$, $p = .17$, $\eta^2 = .09$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 37 (B).

RPE – physical

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and ratings of perceived physical effort as the dependent variable. The main effect of age was not significant, $F(1, 20) = .00$, $p = .98$, $\eta^2 = .00$, indicating that across both devices, there was no difference in perceived physical effort by younger adults ($M = 1.82$, $SD = .77$) compared to older adults ($M = 1.83$, $SD = 1.04$). The main effect of device also was not significant, $F(1, 20) = 1.48$, $p = .24$, $\eta^2 = .07$, indicating that across all participants, perceived physical effort did not differ when using the pen ($M = 1.97$, $SD = 1.22$) or the mouse ($M = 1.68$, $SD = .93$). The interaction between age and device was not significant, $F(1, 20) = .04$, $p = .85$, $\eta^2 = .00$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 37 (C).

Looking at the factor of order, an interaction between order and device was found to be significant, $F(1, 20) = 6.01$, $p < .05$, $\eta^2 = .23$, indicating a differential carryover effect was present. To determine the nature of the interaction between order and device, an analysis of the simple effects of the within-subjects factor (device) at each order (pen first, mouse first) was conducted. Participants who used the pen first had higher

perceived physical effort when using the pen ($M = 2.38$, $SD = 1.33$) than when using the mouse ($M = 1.51$, $SD = .95$), $F(1, 20) = 6.73$, $p < .05$, $\eta^2 = .25$. However, participants who used the mouse first did not differ in perceived physical effort with the pen ($M = 1.56$, $SD = .99$) or the mouse ($M = 1.85$, $SD = .93$), $F(1, 20) = .76$, $p = .39$, $\eta^2 = .04$. Refer to Figure 40.

Figure 40. Ratings of perceived physical effort for form filling task for each device as a function of device order. Error bars show +/- 1 standard error of the mean.

To test the research hypothesis and to compensate for differential carryover effects, a 2x2 between-subjects ANOVA was conducted with age group (younger, older) and device (pen, mouse) as the independent factors and perceived physical effort as the dependent factor. Only data from the first device used was included. The main effect of age was not significant, $F(1, 20) = .02$, $p = .88$, $\eta^2 = .00$, indicating that across both devices, perceived physical effort was the same for younger adults ($M = 2.08$, $SD = 1.05$)

as for older adults ($M = 2.15$, $SD = 1.29$). The main effect of device was not significant, $F(1, 20) = 1.15$, $p = .30$, $\eta^2 = .05$, indicating that across all participants, perceived physical effort did not differ when using the pen ($M = 2.38$, $SD = 1.33$) or the mouse ($M = 1.85$, $SD = .93$). The interaction between age and device was not significant, $F(1, 20) = .05$, $p = .83$, $\eta^2 = .00$, indicating that the direction and magnitude of device effects did not vary across age groups. The results of this analysis do not differ from the analysis with the full data set with regard to significance of effects. Refer to Figure 41.

Cross-ref. "form filling" - RPE (physical), limited to first device

Figure 41. Ratings of perceived physical effort for form filling task for each device as a function of age, limited to first device used. Error bars show +/- 1 standard error of the mean.

RPE – mental

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and ratings of perceived mental effort

as the dependent variable. The main effect of age was not significant, $F(1, 20) = .36, p = .56, \eta^2 = .02$, indicating that across both devices, there was no difference in perceived mental effort experienced by younger adults ($M = 2.70, SD = 1.28$) compared to that experienced by older adults ($M = 3.03, SD = 1.32$). The main effect of device also was not significant, $F(1, 20) = 1.31, p = .27, \eta^2 = .06$, indicating that across all participants, perceived mental effort did not differ when using the pen ($M = 2.75, SD = 1.33$) or the mouse ($M = 2.99, SD = 1.45$). The interaction between age and device was not significant, $F(1, 20) = .89, p = .36, \eta^2 = .04$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 37 (D).

Cross-referencing “document editing” task

Error analysis and data preparation

There were three types of slips that occurred in the task which were included in the error tally. The first was when a range of text was selected incorrectly so that the participant had to go back and reselect it to make correct formatting changes. The second was if a participant encountered difficulty in the precision task of editing the typographical errors (“typos”); usually this took the form of either inaccurately selecting text or moving a letter to the wrong place. The third type of error was if a participant missed a target, such as a button to change font formatting. See Figure 42 for samples.

<p>Real Estate News You Can Use Issue Number 1 June 2006 John Smith. A1 Real Estate Brokers. I</p>
Over-selecting a range of text	Precision errors

Figure 42. Examples of slip errors in the cross-referencing “document editing” task

Participants 13 and 17 failed the cross-referencing document edit task in the pen condition by omitting or failing the typo component of the task. Thus their data was not included in the analysis.

Descriptive statistics

Refer to Table 16 for descriptive statistics of the cross-referencing “document editing” task.

Timing. Younger adults were faster than older adults overall across both devices, averaging 133.92 seconds to complete the task, compared to 227.05 seconds for older adults. Participants of both age groups were overall slower with the pen than the mouse, averaging 197.05 seconds to complete the task with the pen, and 155.45 seconds with the mouse. The difference in time between devices was less pronounced for younger adults, who took 145.92 seconds with the pen and 121.92 seconds with the mouse. Older adults, by contrast, took 258.40 seconds with the pen and 195.70 seconds with the mouse. Variability was fairly consistent within age groups, although older adults had higher standard deviations than younger adults. Younger adults’ standard deviation of time measures were 35.74 for the pen and 29.56 for the mouse, while older adults were 56.30 for the pen and 50.11 for the mouse.

Errors. Younger adults made fewer errors than older adults on average, between 2 and 3 ($M = 2.29$) for younger adults and between 6 and 7 ($M = 6.40$) for older adults. Overall, participants made more errors with the pen, between 6 and 7 ($M = 6.55$), than with the mouse, between 1 and 2 ($M = 1.77$). The difference in number of errors between devices was less pronounced for younger adults, who made on average 3.50 errors with

the pen and 1.08 with the mouse. Older adults, by contrast, made on average 10.77 errors with the pen and 2.60 with the mouse. There was also a much higher amount of variability in older adults' error performance, particularly with the pen. Older adults' standard deviation of error tallies with the pen was 10.77 and with the mouse was 3.03. Younger adults, by contrast, had a standard deviation of error tallies with the pen of 2.71 and the mouse only .67.

RPE – physical. Younger and older adults had about the same perception of physical effort for the task overall – across both devices younger adults' rating was 2.52 and older adults' was 2.49, both corresponding to between “weak” and “moderate” on the Borg CR-10 scale. When looking at overall ratings by device, ratings for both devices still averaged between “weak” and “moderate”, although the pen, at 2.85 was closer to “moderate” and the mouse, at 2.15, was closer to “weak”. This pattern was apparent within age groups as well, although more pronounced for older adults. Younger adults rated the pen on average 2.78, and the mouse 2.26, while older adults rated the pen on average 2.95, and the mouse, 2.15. Variability of ratings was fairly consistent, though somewhat higher for ratings of the pen. Younger adults had a standard deviation of rating scores of 1.59 for the pen and .78 for the mouse, while older adults had 1.34 for the pen and 1.02 for the mouse.

RPE – mental. Overall, across both devices, younger adults perceived less mental effort on the task than older adults, providing ratings of 2.85, corresponding to just below “moderate” on the Borg CR-10 scale, compared to older adults' average rating of 3.51, corresponding to between “moderate” and “strong”. Participants also reported higher

perceived mental effort when using the pen – 3.41, corresponding to between “moderate” and “strong” – compared to using the mouse – 2.89, corresponding to just below “moderate”. Younger adults were fairly consistent between devices, reporting 2.72 for the pen and 2.99 for the mouse, whereas older adults reported more effort with the pen ($M = 4.25$) than for the mouse ($M = 2.77$). Variability of ratings was relatively low and consistent, although standard deviation for older adults’ ratings for the pen was highest at 1.83 compared to 1.16 for the mouse. Younger adults’ standard deviations of ratings was .94 for the pen and 1.16 for the mouse.

Table 16. Descriptive statistics for cross-referencing “document editing” task

Time			Pen	Mouse	Device
	Younger	M	145.92	121.92	133.92
		SD	35.74	29.56	27.75
		N	12	12	12
	Older	M	258.40	195.70	227.05
		SD	56.30	50.11	44.17
		N	10	10	10
	Age	M	197.05	155.45	
		SD	72.90	54.29	
		N	22	22	
Errors			Pen	Mouse	Device
	Younger	M	3.50	1.08	2.29
		SD	2.71	0.67	1.41
		N	12	12	12
	Older	M	10.20	2.60	6.40
		SD	10.77	3.03	6.43
		N	10	10	10
	Age	M	6.55	1.77	
		SD	8.08	2.18	
		N	22	22	
RPE Physical			Pen	Mouse	Device
	Younger	M	2.78	2.26	2.52
		SD	1.59	0.78	0.85
		N	12	12	12
	Older	M	2.95	2.03	2.49
		SD	1.34	1.02	1.00
		N	10	10	10
	Age	M	2.85	2.15	
		SD	1.45	0.88	
		N	22	22	
RPE Mental			Pen	Mouse	Device
	Younger	M	2.72	2.99	2.85
		SD	0.94	1.16	0.81
		N	12	12	12
	Older	M	4.25	2.77	3.51
		SD	1.83	1.16	1.23
		N	10	10	10
	Age	M	3.41	2.89	
		SD	1.58	1.14	
		N	22	22	

Inferential statistics

Refer to Table 17 for a complete summary of effects statistics for the cross-referencing “document editing” task.

Table 17. Effects summary for cross-referencing “document editing” task

		<i>F</i> (1, 18)	<i>p</i>	η^2
Time on Task	Age	32.21***	< .001	.64
	Device	28.02***	< .001	.61
	Age x Device	6.54*	< .05	.27
Simple effects of device by age group	<i>Younger x Device</i>	4.21^	.06	.19
	<i>Older x Device</i>	23.97***	< .001	.57
	Order x Device	9.72**	< .01	.35
Simple effects of device by order	<i>Pen first x Device</i>	28.69***	< .001	.61
	<i>Mouse first x Device</i>	2.66	.12	.13
Time on Task (first device only)	Age	31.00***	< .001	.63
	Device	3.52^	.08	.16
	Age x Device	.55	.47	.03
Errors	Age	3.93^	.06	.18
	Device	12.30**	< .01	.41
	Age x Device	3.43^	.08	.16
RPE Physical	Age	.00	.99	.00
	Device	4.77*	< .05	.21
	Age x Device	.51	.49	.03
RPE Mental	Age	2.50	.13	.12
	Device	4.22^	.06	.19
	Age x Device	8.27*	< .05	.31
Simple effects of device by age group	<i>Younger x Device</i>	.38	.55	.02
	<i>Older x Device</i>	9.16**	< .01	.34

Figure 43. Cross-referencing "document editing" task measures: (A) Time on task; (B) Number of errors; (C) Ratings of perceived physical effort; (D) Ratings of perceived mental effort. Error bars show +/- 1 standard error of the mean.

Time on task

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and time on task as the dependent variable. The main effect of age was significant, $F(1, 18) = 32.21, p < .001$, $\eta^2 = .64$, indicating that across both devices, younger adults ($M = 133.92, SD = 27.75$)

took less time than older adults ($M = 227.05$, $SD = 44.17$) to complete the task.

Additionally, the main effect of device was significant, $F(1, 18) = 28.02$, $p < .001$, $\eta^2 = .61$, indicating that across all participants, performance time was slower when using the pen ($M = 197.05$, $SD = 72.90$) than when using the mouse ($M = 155.45$, $SD = 54.29$).

The interaction between age and device was also significant, $F(1, 18) = 6.54$, $p < .05$, $\eta^2 = .27$, indicating that the direction and magnitude of device effects varied across age groups.

To determine the nature of the interaction between age and device found in the omnibus ANOVA, an analysis of the simple effects of the within-subjects factor (device) at each age group (younger, older) was conducted with time as the dependent variable. Younger adults were marginally but not significantly slower with the pen ($M = 145.92$, $SD = 35.74$) than the mouse ($M = 121.92$, $SD = 29.56$), $F(1, 18) = 4.21$, $p = .06$, $\eta^2 = .19$. Older adults took significantly more time with the pen ($M = 258.40$, $SD = 56.30$) than with the mouse ($M = 195.70$, $SD = 50.11$), $F(1, 18) = 23.97$, $p < .001$, $\eta^2 = .57$. Refer to Figure 43 (A).

Looking at the factor of order, an interaction between order and device was found to be significant, $F(1, 18) = 9.72$, $p < .01$, $\eta^2 = .35$, indicating a differential carryover effect was present. To determine the nature of the interaction between order and device, an analysis of the simple effects of the within-subjects factor (device) at each order (pen first, mouse first) was conducted. Participants who used the pen first were found to complete the task slower with the pen ($M = 198.10$, $SD = 66.42$) than the mouse ($M = 129.50$, $SD = 40.48$), $F(1, 18) = 28.69$, $p < .001$, $\eta^2 = .61$. However, participants

who used the mouse first had no difference between time using the pen ($M = 196.17$, $SD = 80.83$) or the mouse ($M = 177.08$, $SD = 56.25$), $F(1, 18) = 2.66$, $p = .12$, $\eta^2 = .13$.

Refer to Figure 44.

Figure 44. Time on task for document editing task for each device as a function of device order. Error bars show +/- 1 standard error of the mean.

To test the research hypothesis and to compensate for differential carryover effects, a 2x2 between-subjects ANOVA was conducted with age group (younger, older) and device (pen, mouse) as the independent factors and time on task as the dependent factor. Only data from the first device used was included. A main effect of age was found to be significant, $F(1, 18) = 31.00$, $p < .001$, $\eta^2 = .63$, indicating that across both devices, younger adults ($M = 145.92$, $SD = 32.43$) performed faster than older adults ($M = 235.50$, $SD = 48.97$) when performing the task. The main effect of device was marginal but not significant, $F(1, 18) = 3.52$, $p = .08$, $\eta^2 = .16$, indicating that across all

participants, time on task did not differ when using the pen ($M = 198.10$, $SD = 66.42$) or the mouse ($M = 177.08$, $SD = 56.25$). The interaction between age and device was not significant, $F(1, 18) = .55$, $p = .47$, $\eta^2 = .03$, indicating that the direction and magnitude of device effects did not vary across age groups. Unlike the analysis with the full data set, only the effect of age group was significant in this analysis. Refer to Figure 45.

Figure 45. Time on task for document editing task for each device as a function of age, limited to first device used. Error bars show +/- 1 standard error of the mean.

Errors

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and number of errors as the dependent variable. A main effect of age was marginal but not significant, $F(1, 18) = 3.93$, $p = .06$, $\eta^2 = .18$, indicating that across both devices, younger adults ($M = 2.29$,

$SD = 1.41$) did not make more errors than older adults ($M = 6.40$, $SD = 6.43$) when performing the task. However, the main effect of device was significant, $F(1, 18) = 12.30$, $p < .01$, $\eta^2 = .41$, indicating that across all participants, more errors were made when using the pen ($M = 6.55$, $SD = 8.08$) than when using the mouse ($M = 1.77$, $SD = 2.18$). The interaction between age and device was marginal but not significant, $F(1, 18) = 3.43$, $p = .08$, $\eta^2 = .16$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 43 (B).

RPE – physical

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and rating of perceived physical effort as the dependent variable. The main effect of age was not significant, $F(1, 18) = .00$, $p = .99$, $\eta^2 = .00$, indicating that across both devices, younger adults' perceived physical effort ($M = 2.52$, $SD = .85$) was not different from that of older adults ($M = 2.49$, $SD = 1.00$) when performing the task. The main effect of device was significant, $F(1, 18) = 4.77$, $p < .05$, $\eta^2 = .21$, indicating that across all participants, perceived physical effort was greater when using the pen ($M = 2.85$, $SD = 1.45$) than when using the mouse ($M = 2.15$, $SD = .88$). The interaction between age and device was not significant, $F(1, 18) = .51$, $p = .49$, $\eta^2 = .03$, indicating that the direction and magnitude of device effects did not vary across age groups. Refer to Figure 43 (C).

RPE – mental

To test the research hypothesis, a 2x2x2 mixed-factor ANOVA was conducted with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, device (pen, mouse) as the within-subjects factor, and perceived mental effort as the dependent variable. The main effect of age was not significant, $F(1, 18) = 2.50, p = .13, \eta^2 = .12$, indicating that across both devices, younger adults' perceived mental effort ($M = 2.85, SD = .81$) was not different from that of older adults ($M = 3.51, SD = 1.23$) when performing the task. The main effect of device was marginal but not significant, $F(1, 18) = 4.22, p = .06, \eta^2 = .19$, indicating that across all participants, perceived mental effort was no different when using the pen ($M = 3.41, SD = 1.58$) from when using the mouse ($M = 2.89, SD = 1.14$). The interaction between age and device, however, was significant, $F(1, 18) = 9.16, p < .01, \eta^2 = .34$, indicating that the direction and magnitude of device effects varied across age groups.

To determine the nature of the interaction between age and device found in the omnibus ANOVA, an analysis of the simple effects of the within-subjects factor (device) at each age group (younger, older) was conducted with ratings of perceived mental effort as the dependent variable. Younger adults did not rate the pen ($M = 2.72, SD = .94$) significantly different from the mouse ($M = 2.99, SD = 1.16$), $F(1, 18) = .38, p = .55, \eta^2 = .02$. Older adults, however, rated significantly higher perceived mental effort with the pen ($M = 4.25, SD = 1.83$) than with the mouse ($M = 2.77, SD = 1.16$), $F(1, 18) = 9.16, p < .01, \eta^2 = .34$. Refer to Figure 43 (D).

Usability ratings of input devices

Participants answered a usability dimensions survey after using each device for the complete set of tasks. The survey was designed to capture 5 dimensions of usability: comfort, efficiency, usefulness, satisfaction, and appropriateness. Each dimension had at least two questions associated with it, as shown in Table 18.

Table 18. Dimensions of usability survey breakdown

Dimension	#	Question	Direction
Comfort	A2	This device fits my hand well	Positive
	A9	It felt comfortable to use this device	Positive
	B1	Natural - - - Awkward	Negative
	B2	Uncomfortable - - - Comfortable	Positive
Efficiency	A4	This device slowed me down	Negative
	A8	I was able to work efficiently with this device	Positive
Usefulness	A5	I can imagine many cases where I'd want to use this device	Positive
	A6	I would never use this device if I didn't have to	Negative
Satisfaction	A1	I think that I would like to use this device frequently	Positive
	A3	I would recommend this device to a friend	Positive
Appropriateness	A7	I would prefer to use this device for all my computer tasks	Negative
	B1	Universal - - - Task Specific	Positive

A correlation matrix was generated for responses to the surveys across both devices to check the general validity of the usability dimension questions. Question 10 (“Some tasks seemed easier with this device”) was removed from the analysis, as the responses did not correlate as anticipated. Items that had negative direction responses were recoded in the opposite direction (e.g., a response of 2 was recoded to 4). For each of the 5 usability dimensions, the associated response scores were combined and averaged to create a single measure of the usability dimension on a 5 point scale.

A 2x2x2 mixed-factor ANOVA was conducted for each usability dimension measure to determine if order effects were present, with age group (younger, older) and order (pen first, mouse first) as the between-subjects factors, and device (pen, mouse) as the within subjects factor. For the “satisfaction” measure, an interaction between order and device was found to be significant, $F(1, 20) = 6.75, p < .05, \eta^2 = .25$, indicating a differential carryover effect was present. A follow-up analysis was conducted to determine the nature and direction of the effect, and found that when the pen was first, there was a significant difference between “satisfaction” ratings, with a lower rating for the pen ($M = 2.92, SD = 1.43$) than the mouse ($M = 4.71, SD = .45$), $F(1, 20) = 14.16, p < .01, \eta^2 = .41$. However, when the mouse was first, there was not a significant difference between “satisfaction” ratings for the pen ($M = 3.96, SD = .94$) or the mouse ($M = 4.00, SD = 1.07$), $F(1, 20) = .01, p = .93, \eta^2 = .00$. Thus, for the “satisfaction” measure a second analysis was conducted with just the data from the first device used. The results are presented in the following sections.

Descriptive statistics

There was very little variation in results of usability dimensions between age groups overall (i.e., for both devices) with the exception of “satisfaction,” where younger adults provided overall higher ratings of satisfaction ($M = 4.00$, $SD = .67$) than older adults ($M = 2.92$, $SD = 1.65$). Between devices, however, the mouse received higher ratings than the pen for comfort ($M = 4.31$, $SD = .84$ for the mouse, $M = 3.55$, $SD = .90$ for the pen), efficiency ($M = 3.95$, $SD = .93$ for the mouse, $M = 2.73$, $SD = .86$ for the pen), usefulness ($M = 4.27$, $SD = 1.03$ for the mouse, $M = 3.52$, $SD = 1.26$ for the pen), and satisfaction ($M = 4.00$, $SD = 1.07$ for the mouse, $M = 2.92$, $SD = 1.43$ for the pen). The pen was identified as more task-specific, by gaining higher ratings for appropriateness ($M = 3.33$, $SD = 1.05$ for the pen, $M = 2.03$, $SD = .96$ for the mouse). Refer to Table 19.

Table 19. Descriptive statistics of usability dimensions survey results

		Pen	Mouse	Device
Comfort				
Younger ($N = 12$)	<i>M</i>	3.94	4.31	4.13
	<i>SD</i>	0.72	0.73	0.50
Older ($N = 12$)	<i>M</i>	3.17	4.31	3.74
	<i>SD</i>	0.93	0.97	0.51
Age ($N = 24$)	<i>M</i>	3.55	4.31	
	<i>SD</i>	0.90	0.84	
Efficiency				
Younger ($N = 12$)	<i>M</i>	2.83	3.96	3.40
	<i>SD</i>	0.49	0.78	0.45
Older ($N = 12$)	<i>M</i>	2.63	3.94	3.28
	<i>SD</i>	1.13	1.10	0.57
Age ($N = 24$)	<i>M</i>	2.73	3.95	
	<i>SD</i>	0.86	0.93	
Usefulness				
Younger ($N = 12$)	<i>M</i>	3.79	4.42	4.10
	<i>SD</i>	0.78	0.67	0.62
Older ($N = 12$)	<i>M</i>	3.25	4.13	3.69
	<i>SD</i>	1.59	1.32	0.60
Age ($N = 24$)	<i>M</i>	3.52	4.27	
	<i>SD</i>	1.26	1.03	
Satisfaction (1st Device Only)				
Younger ($N = 12$)	<i>M</i>	3.67	4.33	4.00
	<i>SD</i>	0.68	0.52	0.67
Older ($N = 12$)	<i>M</i>	2.17	3.67	2.92
	<i>SD</i>	1.63	1.40	1.65
Age ($N = 24$)	<i>M</i>	2.92	4.00	
	<i>SD</i>	1.43	1.07	
Appropriateness				
Younger ($N = 12$)	<i>M</i>	3.17	1.87	2.52
	<i>SD</i>	0.81	0.71	0.29
Older ($N = 12$)	<i>M</i>	3.50	2.19	2.85
	<i>SD</i>	1.26	1.16	0.46
Age ($N = 24$)	<i>M</i>	3.33	2.03	
	<i>SD</i>	1.05	0.96	

Inferential statistics

To determine if differences between user ratings of usability dimensions were significantly different, for each measure the results of the aforementioned ANOVAs (used to test for order effects) were consulted, and the factors of age group, device, and the interaction between age group and device were examined. The main effect of age group was not significant for comfort, $F(1, 20) = 3.62, p = .07, \eta^2 = .15$, efficiency, $F(1, 20) = .28, p = .60, \eta^2 = .01$, or usefulness, $F(1, 20) = 2.60, p = .12, \eta^2 = .11$, indicating that overall ratings did not differ between age groups for these measures. For the measure of appropriateness, the main effect of age was marginal but not significant, $F(1, 20) = 4.27, p = .05, \eta^2 = .18$. For the measure of satisfaction, the main effect of age was significant, $F(1, 20) = 5.25, p < .05, \eta^2 = .21$, indicating that across both devices younger adults rated higher overall satisfaction ($M = 4.00, SD = .67$) than older adults ($M = 2.92, SD = 1.65$).

The main effect of device was significant in all but one of the measures, and in that one it was marginal. For comfort, participants rated the mouse ($M = 4.31, SD = .84$) significantly higher than the pen ($M = 3.55, SD = .90$), $F(1, 20) = 8.66, p < .01, \eta^2 = .30$. For efficiency, participants rated the mouse ($M = 3.95, SD = .93$) significantly higher than the pen ($M = 2.73, SD = .86$), $F(1, 20) = 18.57, p < .001, \eta^2 = .48$. The measure of usefulness was marginally significant, $F(1, 20) = 4.22, p = .05, \eta^2 = .17$, thus with this sample there was no difference between ratings for the mouse ($M = 4.27, SD = 1.03$) or the pen ($M = 3.52, SD = 1.26$). For satisfaction, participants rated the mouse ($M = 4.00, SD = 1.07$) significantly higher than the pen ($M = 2.92, SD = 1.43$), $F(1, 20) = 5.25,$

$p < .05$, $\eta^2 = .21$. Finally, for appropriateness, participants rated the pen ($M = 3.33$, $SD = 1.05$) significantly higher than the mouse ($M = 2.03$, $SD = .96$), $F(1, 20) = 12.77$, $p < .01$, $\eta^2 = .39$. Refer to Figure 46.

Figure 46. Mean scores of usability dimensions by device – both age groups combined. Data for Satisfaction reflects only half of the participants to compensate for differential carryover effects. Error bars show +/- 1 standard error of the mean.

None of the age group by device interactions were significant, indicating that the magnitude and direction of device effects did not vary between age groups. Refer to Table 20 for complete statistics associated with the analysis.

Table 20. Summary of effects for usability dimensions survey

Dimension	Factor(s)	<i>F</i> (1, 20)	<i>P</i>	η^2
Comfort	Age	3.62 [^]	.07	.15
	Device	8.66**	< .01	.30
	Age x Device	2.22	.15	.10
Efficiency	Age	0.28	.60	.01
	Device	18.57***	< .001	.48
	Age x Device	0.11	.74	.01
Usefulness	Age	2.60	.12	.11
	Device	4.22 [^]	.05	.17
	Age x Device	0.12	.74	.01
Satisfaction	Age	5.25*	< .05	.21
	Device	5.25*	< .05	.21
	Age x Device	0.78	.39	.04
Appropriateness	Age	4.27 [^]	.05	.18
	Device	12.77*	< .01	.39
	Age x Device	0.00	.98	.00
[^] marginal significance, $p < .10$		**	significant, $p < .01$	
* significant, $p < .05$		***	significant, $p < .001$	

Device preference

Participants were asked to indicate their device preference on a task-by-task basis with a forced-choice questionnaire. The mouse was the preferred device for the Excel dragging task, the PowerPoint manipulation task, and both cross-referencing tasks for both younger and older adults. However, participants were fairly evenly split on the PowerPoint dragging task. Exactly half of younger adults preferred the pen on the task, while 58.33% of older adults preferred the pen to the mouse. While participants preferred the mouse to the pen on the cross-referencing form filling task, it was by a lesser margin than other tasks, with a third of participants in both age groups preferring

the pen. The pen was the unanimous choice for the tracing task in both groups. Refer to Table 21, Figure 47, and Figure 48.

Table 21. Device preference by task

Task	Younger		Older	
	Mouse	Pen	Mouse	Pen
Excel dragging	91.67%	8.33%	75.00%	25.00%
PowerPoint dragging	50.00%	50.00%	41.67%	58.33%
PowerPoint manipulation	83.33%	16.67%	91.67%	8.33%
Tracing	0.00%	100.00%	0.00%	100.00%
Cross-ref form filling	66.67%	33.33%	66.67%	33.33%
Cross-ref document edit	83.33%	16.67%	83.33%	16.67%

Figure 47. Device preference by task, younger adults only

Figure 48. Device preference by task, older adults only

Finally, participants were asked to rate the devices overall on a scale from 1 to 5 with 1 being the worst and 5 being the best, with the caveat that they could not provide the same rating to both devices (i.e., they had to rank the devices). The mouse received a higher rating overall across both groups ($M = 4.58$, $SD = .78$) compared to the pen ($M = 3.08$, $SD = 1.18$) although greater variability was observed in the ratings of the pen, as evidenced by the higher standard deviation. When looked at between age groups, the mouse still fared better, garnering a higher rating from both younger ($M = 4.83$, $SD = .39$) and older ($M = 4.33$, $SD = .98$) groups compared to their ratings for the pen ($M = 3.58$, $SD = .51$ and $M = 2.58$, $SD = 1.44$, respectively). Overall, younger participants gave higher ratings to both devices combined ($M = 4.21$, $SD = .40$) than did the older participants ($M = 3.46$, $SD = .66$). Thus it appears that the younger participants generally gave higher ratings to both input devices and rated them closer together. The older adults

appeared to have stronger feelings about the difference between the devices, and rated the pen lower.

However, interactions between order and device, as well as a three-way interaction between order, age, and device were present in the user preference overall rating, $F(1, 20) = 10.25, p < .01, \eta^2 = .34$, and $F(1, 20) = 8.10, p < .05, \eta^2 = .29$, respectively, indicating the presence of differential carryover effects. Therefore, to test for significant differences between factors and to compensate for differential carryover effects, a second analysis was conducted using only the data from the first device used. The results from the 2x2 between-subjects ANOVA with age (younger, older) and device (pen, mouse) as the independent variables, and overall rating as the dependent variable, yielded a similar outcome. Younger adults ($M = 4.17, SD = .83$) rated devices significantly higher overall than older adults ($M = 2.75, SD = 1.60$), $F(1, 20) = 14.60, p < .01, \eta^2 = .42$, and the pen received lower marks overall ($M = 2.58, SD = 1.31$) than the mouse ($M = 4.33, SD = .98$), $F(1, 20) = 22.27, p < .001, \eta^2 = .53$. The interaction between age and device was not significant, $F(1, 20) = 1.26, p = .27, \eta^2 = .06$, indicating that the magnitude and direction of device effects did not vary between age groups. Refer to Figure 49.

Figure 49. Overall ratings of device by age group, limited to first device used. Error bars show +/- 1 standard error of the mean.

Post-trial interviews

Participants were interviewed in a semi-structured manner after completing the protocol with both devices. The responses tended to focus on use of the pen, as the pen was the novel device for participants. Participants were asked for overall impressions of using both devices, which tasks they preferred devices for, any problems they noticed, and what suggestions they might have for the design of future pen computing systems. In the following description of the results of the interviews, multiple comments that addressed the same issue have been paraphrased into representative comments.

All participants had a positive experience with the pen on the tracing task, and some expressed surprise that the mouse was such a poor device to use for the task – it had not occurred to some that the mouse was not suitable for all tasks. Eight participants also

commented that they “liked the pen for pointing, it seemed more natural for the PowerPoint dragging,” and one added that it was “less stressful mentally” for that task. Several participants also cited the survey (cross-referencing form filling) task, stating “[the] survey was easier with the pen because clicking on it was more natural.” Half of the older participants specifically commented that they appreciated the direct interaction, i.e., they “liked working on the screen.” A number of participants commented positively on the design of the pen itself, one stating “I liked the pen, it’s small, comfortable, fit my hand, feels natural – it’s fun.”

Not all participants were as positive about the pen, however. One participant with big hands felt he “didn’t have good control, it wasn’t comfortable to hold.” Another, with minor arthritic nodes on her fingers commented that the “pen is small and has no grip especially for someone older with nodes on their fingers.” Several participants, speaking to the issue of direct interaction, “liked the mouse better. I could move around easier; the pen wasn’t fast enough.” With regard to the tasks, the overwhelming majority of participants commented on the difficulty of using the pen for certain tasks – “[it was] harder to do tasks with fine detail or small tools – I didn’t have enough precise control,” and some found it frustrating to use for all tasks, stating “all tasks were pretty difficult with the pen.”

Participants cited several specific interaction difficulties they encountered with the pen. The problem of parallax was noted by 7 participants, saying “it felt like I was clicking in the right spot but I wasn’t, it got out of calibration too easily.” Others noted that “it was hard to grab the vertical scrollbar with the pen,” referring the scrollbar on the

right side of the screen, for which the pen often seemed out of calibration. The problem of occlusion was noted by 11 participants, who complained that while working on a task “[my] hand and the pen obstructs the view for detailed tasks.” There was 1 left-handed participant in the study, who also commented that form widgets (drop-menus) appeared under her hand. Five participants commented on the right-click interaction (press and hold) as being “annoying.” Finally, a number of participants commented on the overall use of the tablet. Many commented “it was awkward, difficult to cradle the tablet – the tablet was too heavy.” Several also complained of glare on the screen limiting their ability to find a comfortable position in which to use the tablet.

Comments on the mouse were fewer but generally positive. Several participants commented that it was easier working in Excel with the mouse, and several noted that there was “less physical effort using the mouse.” Two participants also commented that they “like the sound feedback of the mouse [when clicking].” One participant who did not like using the mouse explained that he “didn’t like to take up desk real estate and prefer a trackball.”

When asked what design changes or improvements they would make to the tablet PC, participants provided a wide range of ideas. Several commented on the footprint of the pen itself, noting that it “needs to be more substantial – not so plasticky.” Others noted that the pen should be “bigger, fatter for an easier grip, more ergonomically designed.” Ideas for making the pen more “ergonomic” included adding a rubber grip to the pen, trying a two-fingered pen, or a mount-on-finger pen for people with arthritis. There were also comments about the tip of the pen, that it needed to be “sharper and

thinner, it felt too fat.” Several participants commented on the button on the pen, suggesting to add more buttons, and to label the button so it is more obvious to prevent accidental activation. One participant commented on the overall tablet PC itself, suggesting to “add rubber feet so the user could position it how they want.”

Suggestions for software improvements included making tools larger, as “the PowerPoint diamond tool [in the manipulation task] is too small for the pen.” Although participants liked the direct interaction associated with the pen, several participants suggested having a “mirror” capability, or offsetting the calibration of the pen to the pointer, to mitigate the occlusion problem. Finally, one participant pointed out that when using a real pen, “you never double-tap, it’s not natural, so need to come up with a different interaction for that.”

When asked in what contexts they could envision using a tablet PC, participants thought it would be “beneficial to use the pen in a mobile context, on the bus or commuting on the train” as well as “on a plane, in a lecture, cramped spaces.” For application contexts, many participants identified graphic applications, drawing, or photo editing. Two of the younger participants thought the tablet would be useful for note-taking, and one older participant thought the pen could be used for games or online jigsaw puzzles, as she enjoyed the physical interaction of the pen and found it fun to use.

Discussion

Review of hypotheses

The research hypotheses for the study were based on the known effects of age on human performance, the task properties, and the properties of the devices. Specifically, those tasks that required precision work (e.g., activating small tools, precise selection and manipulation) were hypothesized to be better performed with the mouse because the mouse uses relative positioning and indirect manipulation, affording a better view of the task activity. Those tasks that had a compelling analog to an actual physical activity one might do with a stylus or finger were hypothesized to be better performed (or equally well performed) with the pen because the pen is a direct manipulation device using absolute positioning. Finally, the amount of continuous action (e.g., dragging, drawing) was considered as a factor generally favoring use of the pen. Interactions were hypothesized to be present for tasks favoring the pen indicating a benefit for older adults to use the pen instead of the mouse. Refer to Table 22 for a summary of task properties and hypothesized results.

Table 22. Time and error hypotheses

Task	Task components			Hypotheses		
	Precision	Analog	Continuous	A: Age	B: Device	AxB
Excel dragging	✓	✓	✓	$Y < O$	$P > M$	✗
PowerPoint dragging	✗	✓	✓	$Y < O$	$P \leq M$	✓
PowerPoint manipulation	✓✓	✗	✓	$Y < O$	$P > M$	✗
Tracing	✗	✓✓	✓✓	$Y < O$	$P \leq M$	✓
Cross-ref form filling	✓	✓	✗	$Y < O$	$P = M$	✓
Cross-ref doc editing	✓✓	✗	✗	$Y < O$	$P > M$	✗

Hypotheses regarding perceived physical and mental effort were similar to performance hypotheses, although age differences were generally hypothesized to be present only in the more difficult tasks. Because direct manipulation is more physically involved than indirect manipulation, in general the pen was hypothesized to require greater physical effort except in cases where strong analogs to real world tasks are present. Mental effort with the pen was hypothesized to be generally the same as the mouse, except in cases where a strong analog is present, in which case the presence of a motor program should make mental effort lower for the pen than for the mouse.

Interactions were not hypothesized for perceived effort ratings. Refer to Table 23 for a summary of task properties and hypothesized results.

Table 23. Ratings of perceived physical and mental effort hypotheses

Task	Task components			Hypotheses		
	Precision	Analog	Continuous	A: Age	B: Device	AxB
Excel dragging	✓	✓	✓	$Y = O$	$P > M$	✗
PowerPoint dragging	✗	✓	✓	$Y = O$	$P = M$	✗
PowerPoint manipulation	✓✓	✗	✓	$Y < O$	$P > M$	✗
Tracing	✗	✓✓	✓✓	$Y = O$	$P \leq M$	✗
Cross-ref form filling	✓	✓	✗	$Y = O$	$P = M$	✗
Cross-ref doc editing	✓✓	✗	✗	$Y < O$	$P > M$	✗

Discussion of quantitative results

Participant expertise with devices. Participants had different levels of expertise with the two input devices, and thus interpretation of the results with regard to the device should take this into consideration. Since participants had 10 years (median) experience with a mouse, and 0 years (median) experience with an alternative device such as a stylus, participants could be considered experts at using the mouse and novices at using the pen. Two of the older participants had considerable experience using a Wacom tablet, and preferred it to a mouse, but all other participants had virtually no experience with a stylus. As a result, in cases where the pen might be hypothesized to be better than the mouse, a result of no difference between the two (i.e., accepting the null hypothesis) might reflect a positive result for the pen, particularly for performance measures (i.e., time and errors).

Although participant expertise with the input devices was considered important for the interpretation of performance measures, its impact on the subjective perceived effort measures is less clear. Although participants would likely become more skilled with the pen after having more experience with it, it is not clear that their perceived physical effort would be greatly changed because the physical requirements of moving the device and working with it would not be any different whether the participant is experienced or not. If anything, it is possible that perceived mental effort would be affected by lack of expertise, because the participant may have to concentrate harder using a new device in the absence of a motor program. However, the degree to which expertise may have impacted these measures is unknown. Thus, the issue of expertise was not considered in the interpretation of perceived effort measures as it was for performance measures.

Age hypotheses. As hypothesized, older adults were generally slower than younger adults on all tasks, with the exception of the tracing task, where time was the same for both groups. Age effects for time consistently had large effect sizes (around .50), which is to be expected, given the well-established, robust effect of age on human performance. The lack of an age effect in the tracing task is likely due to the unusual nature of the tracing task, i.e., because it has such a strong analog to its pencil-and-paper equivalent, the motor program for using a stylus for drawing was likely well enough established with older adults to transcend any differences from younger adults.

Age effects for errors were less pronounced, likely reflecting a speed-accuracy tradeoff. Older adults made more errors than younger adults on the Excel dragging task,

and the tracing task, although marginal age effects were also noted in the PowerPoint dragging task and cross-referencing document editing task. Only on the first task, the Excel dragging task, did older adults make more errors as well as take more time than younger adults. As the Excel dragging task was early in the protocol it is possible that participants were still getting used to using the pen. On the tracing task, it appears that older adults engaged in a speed-accuracy tradeoff favoring speed, as their timing was the same as younger adults, but they made more errors.

For ratings of perceived effort, it is interesting to note that age did not come into play, largely as hypothesized. While it was hypothesized that age effects would be present for the most difficult of the precision tasks, i.e., the PowerPoint manipulation task and the cross-referencing document editing task (with its difficult typo editing component), even in these cases the null hypothesis was not rejected for the main effect of age. The only place age came into play with these measures was when it interacted with device effects on the Excel dragging task and the cross-referencing document editing task, which is examined further in the following discussion of device effects.

Device hypotheses. The hypotheses regarding device effects were that tasks requiring precision actions would generally be performed faster and with fewer errors with the mouse than the pen, yet tasks not requiring precision that have a compelling analog to an actual physical activity one might do with a stylus or finger would be performed better (or equally well) with the pen. Taking into account the relative expertise of participants with the devices, the hypotheses were for the most part supported by the study.

For the Excel dragging, PowerPoint manipulation, and cross-referencing document editing task, all of which require precision activity to select the correct tools or perform actions, participants were clearly more successful with the mouse. Although only on the tracing task did participants perform faster and make fewer errors with the pen, on the PowerPoint dragging task and cross-referencing form filling task, there was no difference in time performance between using the pen or the mouse. Considering that participants were novices with the pen, these results suggest that the pen supported these tasks at least as well as the mouse, and with practice, may be a better device to use than the mouse. Alternately, it may be that regardless of expertise, both devices are equally suited for the tasks. This interpretation seems likely for the cross-referencing form filling task because its properties favor both the mouse (precision work) and the pen (ballistic actions, analog to real activity). With the exception of the tracing task, however, more errors were made with the pen. A high amount of variability in error scores in the pen condition for most of the tasks suggests that individual differences in ability to learn and use the novel device is an important factor.

Ratings of perceived physical effort are a better subjective indicator of participants' experience with the input device than perceived mental effort, which more reflects the task complexity. These ratings were collected separately specifically for the purpose of distinguishing use of the input device from task demands. Thus, while perceived mental effort may in some cases provide some insights into using different devices with different age groups, for the most part perceived physical effort is the measure of interest when looking at device effects.

As hypothesized, ratings of perceived physical effort were mostly higher for the pen on tasks requiring precision, especially as the tasks became more challenging. Higher ratings of perceived physical effort on the PowerPoint manipulation task and the cross-referencing editing task, with its challenging “typo correction” component, confirm that the pen is not suitable for the tasks not only by performance metrics, but also from the subjective perspective of the user. The tracing task was the only task for which perceived physical effort was greater for the mouse than for the pen, as hypothesized. The tracing task, with its analog to real-world drawing, lack of precise manipulation but requirement for continuous, controlled activity, was not surprisingly better served by the pen.

Age and device interactions. One of the goals of the present study was to investigate interactions of age and device in more realistic tasks. Previous work in lab studies found a benefit for older adults using a pen, i.e., the age-related decline in performance was mitigated by use of the pen instead of the mouse, and it was hypothesized that these effects would be seen in the tasks that did not require precision actions. Such a finding would have real-world implications for potential uses of computer devices for aging populations. However, in all tasks, the null hypothesis for interactions between age and device was retained for the performance measures of time and errors. Furthermore, the effect size for age by device interactions were small, ranging from .00 to .16, with a median of .04 (these effect sizes are considered small because partial η^2 is an inflated statistic, as described at the beginning of the results

section). Thus the effects that were observed in elemental task studies do not appear to hold up in more complex, realistic tasks.

There were two interactions between age and device with the subjective ratings of perceived effort, and both occurred in tasks that involved precision components. In the Excel dragging task, which had less intense precision work than the other tasks, younger adults rated the pen higher in physical effort, while older adults rated both devices the same. As it is known that older adults have some problems using the mouse because of declines in motor control associated with age, this result seems to support the idea that the pen may be beneficial for older adults. However, this conclusion must be qualified, as the effect was not seen in other tasks, where older adults did not differ from younger adults in their perceived physical effort. The other interaction between device and age was for perceived mental effort for the cross-referencing “document editing” task. In this case, older adults rated higher mental effort with the pen, while younger adults did not rate any difference between devices. This likely reflects the difficulty of the typo correction component of the task, which older adults clearly had more difficulty with and had to concentrate more on to accomplish. Thus it is worth noting that degree of difficulty of performing a task with different devices may vary for different ages. Again, because this was only observed in one of the tasks, any conclusions must be made with caution.

Differential carryover effects. In some cases, the order in which participants were exposed to the devices affected their performance results or ratings of the devices, and although this was not the focus of the study, because of its occurrence some discussion is

warranted. In most cases, the pattern was such that when participants used the pen first, their performance or ratings were worse for the pen, but when the mouse was first, there was no difference. The disparity in expertise on the two devices is likely responsible for these effects, as the participant's ability to learn and perform the task appears to be influenced by the lack of experience with the device. In one case (timing for the PowerPoint dragging task), a more straightforward effect was observed: when the pen was first, time was worse with the pen, but when the mouse was first, time was worse with the mouse. In this case it seems that after completing the task the first time, the participant had a better strategy for completing the task the second time, and which device they used did not make a difference. As the remedy for resolving these effects was to only use the first device data for the measure in question, a larger sample size would have been beneficial for the analysis. Also, providing more practice for learning the task mechanics, while not feasible because of time limitation in this protocol, could also have reduced or eliminated the effects.

Summary. The robust effect of age on performance, while not surprising, nevertheless confirms that in complex, realistic computing tasks, and even despite more years of experience with computers, older adults are still generally slower and make more errors than younger adults. With regard to device choice, clearly the task demands affect which device is most suitable for the task. While this again is not surprising, it is worth noting that in the tasks studied, even the presence of a single task component (e.g., the typo component of the document editing task) can determine the overall effect of device superiority. It does not appear that device effects on performance vary between age

groups on the more complex tasks used in the present study. Thus, while older adults may benefit from using a pen on tasks that it is suited for, the results from the present study do not support the adoption of a pen device by older adults as a way to mitigate the effects of age on human performance.

User preferences and qualitative indicators

While not surprising that participants unanimously preferred the pen over the mouse for the tracing task, it bodes well for the pen that a respectable portion of participants in both age groups also preferred the pen for the other two tasks that did not require highly precise actions and which had a compelling analog to a real activity (the PowerPoint dragging task and the cross-referencing form filling task). These preferences are mirrored by the comments user made in the post-interviews, where they identified having difficulty with the pen on tasks requiring precision activity and small tools. Participants also identified the pen as more task specific than the mouse, rating it higher on the “appropriateness” dimension of the usability survey.

However, in all of the usability dimensions except appropriateness, the mouse surpassed the pen, and the overall ratings for the devices were much higher for the mouse than the pen, particularly among older adults. While the pattern of device preference for tasks was generally the same for younger adults as it was for older adults, the younger adults gave higher overall ratings to the pen than the mouse, suggesting a greater willingness on the part of the younger adults to try using the pen for computing. This was also reflected in observations during the study, in which younger adults seemed less troubled by the difficulties of using the pen, whereas older adults often had more extreme

reactions to the problems encountered. The willingness to adopt new technology seems to be greater amongst the younger users.

The participant feedback about difficulties with using the pen provide some valuable insight and data into future development of pen computing. Overwhelmingly, participants complained about occlusion of the display while trying to perform precision tasks, and many also were frustrated by parallax-related problems in actuating targets. Additionally, participants noted the difficulty of working with small tools, and cited the size and shape of the pen as being a problem in accessing these tools. The neutral to low ratings of efficiency and satisfaction with the pen are consistent with these comments. A combination of software and hardware solutions could begin to address these problems. Software solutions could focus on making tools and widgets easier to actuate using the pen, as well as attempting to compensate for occlusion problems. Hardware solutions could continue to work on reducing parallax problems as technology improves, but also provide a wider variety of pen shapes and designs that are available to use with the tablet. Finally, the tablet itself could be improved to better accommodate its most likely use – mobile computing contexts – by moving away from the boxy, heavy form and providing users with options to customize the position of the tablet in order to minimize glare and afford flexibility in the position of the tablet.

Limitations

The tablet PC used in the study, while representative of current consumer technology, had specifications that may have caused problems for participants. Specifically, with its 14” screen, it is one of the largest convertible tablet PCs on the

market, thus making it heavier and more awkward to use in a table configuration.

Additionally, the pen provided with the tablet is small and lightweight and, as noted by some participants, less substantive than a typical pen or pencil. Other tablet PCs in the same product class use styluses with different parameters that might have been more favorable for the users.

The tasks used in the study, while more complex than a typical lab study with abstract pointing and clicking gestures, were still significantly constrained in that they did not incorporate text entry, and only represented a limited interaction with the various software applications. Also in some cases participants were prohibited from using application aids that would have made the task much easier (e.g., using spell checking in Word for the typo tasks). Thus the conclusions should be interpreted recognizing that the tasks were still simpler than real-world use of a personal computer.

The presence of differential carryover effects in many of the tasks suggests that for these types of studies, more practice and training is required with a novel device before it can be reliably compared to a device with which participants have experience. Alternately, tasks could be learned with the more familiar device first in all cases, then tested on the novel device.

The present study was conducted in a laboratory set up as a typical office work space. While a reasonable place to begin with a comparison study of this type, aspects of the setup may have biased results in favor of the mouse and against the pen. Having a desk as the primary workspace was well suited for the laptop and mouse configuration, but participants often struggled trying to find a comfortable position for the tablet PC,

and the keyboard tray, which most participants used to balance the tablet, was slippery. A more neutral environmental context or a greater range of contexts could address this issue.

The participant sample was too small to detect subtle effects and interactions. Also there was too much variation in experience with software applications; too many of the older participants had trouble understanding tasks. Future studies should use larger samples and control more for participant experience. In the absence of age-related interactions, it may be reasonable to eliminate the age factor, use more homogenous participants, and focus more on task variables.

The disparity in expertise with the input devices among participants is an enduring problem in this type of study. Given the prevalence of the mouse it may not be possible to find an adequately unbiased sample. Some studies have had participants use devices with their non-preferred hand, or use an unfamiliar relative device instead of a mouse to mitigate this confound, but in these cases the participants are either functioning at a severe handicap or using devices they would not normally use. Adequate training over a period of time is probably the best solution to this problem, so that participants are at a normal level of proficiency with the new device.

Suggestions for future research

Now that a broader range of more realistic tasks has been conducted, future studies could use more focused, controlled comparisons of task elements. For example, varying the size of application widgets or text and manipulating both with a mouse and a pen might yield significant data to bolster design requirements for pen computing. The

data from the present study suggest that, given participants' lack of difficulty with small widgets when using the mouse, if a comparison study were done varying widget size and device, an interaction between device and widget size should show marked improvement for performance with the pen while performance with the mouse would likely stay much the same.

The contexts of use that participants in this study envisioned for the tablet PC (along with others) could be explored in future work. Computing in a mobile or cramped environment, using the computer for photographic, graphic, or gaming applications, and other types of uses should be explored. The benefits of pen computing and using a tablet PC may be better realized in a study that includes varied contexts of use.

Age as a factor in cross-sectional studies like the present one is confounded by the fact that the current younger adults have grown up with computers, and been exposed to them in their formative years. Longitudinal studies of aging and HCI could now be embarked upon with the intent of studying aging with populations that have grown up with computers. Although technology will no doubt advance significantly in the next generation, it would be instructive to see if age differences in HCI become less important when individuals have been using computers since childhood.

Design recommendations

Note that these recommendations do not address hardware performance and technical issues as they may have been specific to the tablet PC used in the present study. Also, it is a given for example that parallax problems should be reduced, but that is a technological limitation.

In designing software for use with the pen, occlusion of the display must be taken into consideration; the user needs to be able to see what they are doing. Various approaches could be used, from simple repositioning of some elements, to co-locating controls or even creating a micro-relative-control application of the pen (i.e., the user would manipulate interface elements offset from the actual tools).

Tools used in applications intended for use with the pen should be sized appropriately, so that users are not hampered by small widgets that they cannot easily manipulate. For example, Microsoft One Note, which is intended for use exclusively with the pen, still has small buttons and tabs with which the pen-wielding user may have difficulty.

Applications that might be used dually with a pen or a mouse should have alternate views so that pen users can effectively use the applications (e.g., larger icons and buttons, repositioning of elements so they do not appear under the user's hand). Alternately, such applications should be designed so that elements that are difficult with the pen are made easier; larger buttons and tools will not make it any more difficult to use the application with the mouse.

Tablet PCs should accommodate a range of stylus input devices. While other pens may work with the tablet used in the study, the design of the tablet is such that it encourages use of only the pen supplied with the product, as there is a storage slot for the device that will only accept the one pen. Thus providing a "drawer" or more variable pen storage component would allow for a broader range of input devices and users with different needs could select the device that works best for them. While the market may

not be large enough to support a wide range of stylus products, even allowing simple modifiers to the stylus, such as the addition of a rubber ring for better grip, would be beneficial.

References

- Adler, A., Gujar, A., Harrison, B. L., O'Hara, K., & Sellen, A. (1998). A diary study of work-related reading: Design implications for digital reading devices. *Proceedings of CHI '98: ACM Conference on Human Factors in Computing Systems*, 241-248.
- Albert, A. E. (1982). The effect of graphic input devices on performance in a cursor positioning task. In *Proceedings of the Human Factors Society 26th Annual Meeting*, 54-58. Santa Monica, CA: Human Factors and Ergonomics Society.
- Alvarado, M. M. (2004). Effects of computer input device on movement time and error among older adults. *Masters Abstracts International*, 42 (06), 2311, (AAT 1420418).
- Anderson, R., Anderson, R., Simon, B., Wolfman, S. A., VanDeGrift, T., & Yasuhara, K. (2004). Experiences with a tablet PC based lecture presentation system in computer science courses. *Proceedings of the 35th SIGCSE Technical Symposium on Computer Science Education*, 56-60.
- Andre, A. D., & Wickens, C. D. (1995). When users want what's NOT best for them. *Ergonomics in Design*, October 1995, 10-14.
- Apte, A., & Kimura, T. D. (1993). A comparison study of the pen and the mouse in editing graphic diagrams. *Proceedings 1993 IEEE Symposium on Visual Languages*, 1993, 352-357.
- Borg, G. (1998). *Borg's Perceived Exertion and Pain Scales*. US: Human Kinetics.
- Bullinger, H., Kern, P., & Braun, M. (1997). Controls. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (pp. 697-728). Mahwah, NJ: Lawrence Erlbaum.
- Buxton, W. (1990). A three-state model of graphical input. In D. Diaper et. al. (Eds.) *Human-Computer Interaction – INTERACT '90*, 449-456. Amsterdam: Elsevier.
- Chaparro, A., Bohan, M., Fernandez, J., Choi, S. D., & Kattel, B. (1999). The impact of age on computer input device use: Psychophysical and physiological measures. *International Journal of Industrial Ergonomics*, 24, 503-513.
- Charness, N., Holley, P., Feddon, J., & Jastrzembski, T. (2004). Light pen use and practice minimize age and hand performance differences in pointing tasks. *Human Factors*, 46(3), 373-384.

- Cohen, O., Meyer, S., and Nilsen, E. (1993) Studying the movement of high tech rodents: Pointing and dragging. *Adjunct Proceedings of INTERCHI '93: ACM New York*, 135-136.
- Czaja, S. J. (1997). Computer technology and the older adult. In M Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction. Second, completely revised edition* (pp. 797-812). New York, NY: Elsevier Science.
- Elliott, A., & Hearst, M. A. (2002). A comparison of the affordances of a digital desk and tablet for architectural image tasks. *International Journal of Human-Computer Studies*, 56(2), 173-197.
- Fitzmaurice, G. W., & Buxton, W. (1997). An empirical evaluation of graspable user interfaces: towards specialized, space-multiplexed input. *Proceedings of CHI '97: ACM Conference on Human Factors in Computing Systems, PAPERS: Handy User Interfaces, Vol. 1*, 43-50.
- Greenstein, J. S. (1997). Pointing devices. In M Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of Human-Computer Interaction. Second, completely revised edition* (pp. 1317-1348). New York, NY: Elsevier Science.
- Haller, R., Mutschler, H., & Voss, M. (1984). Comparison of input devices for correction of typing errors in office systems. In *Proceedings of the Interact '84 Conference, First IFIP Conference on 'Human-Computer Interaction', Volume 2*, 218-223.
- Hinckley, K. (2003). Input technologies and techniques. In J. A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook* (pp. 151-168). Mahwah, NJ: Lawrence Erlbaum.
- John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: comparison and contrast. *ACM Transactions on Computer-Human Interaction*, 3(4), 320-351.
- Karlqvist, L., Bernmark, E., Ekenvall, L., Hagberg, M., Isaksson, A., Rostö, T. (1999). Computer mouse and track-ball operation: similarities and differences in posture, muscular load and perceived exertion. *International Journal of Industrial Ergonomics*, 23, 157-169.
- Keely, B. (2005). Living without paper: A challenge for the man-machine interface. *Keynote presentation at ADEAC Conference*. Retrieved February 16, 2006 from <http://www.edn.com/media/SIDADEAC/Microsoft/LivingWithoutPaper.ppt>

- Kimura, T. D., Citrin, W., Halbert, D., Hewitt, C. Meyrowitz, N., & Shneiderman, B. (1993). Potentials and limitations of pen-based computers. *Proceedings of the 1993 ACM Conference on Computer Science*, 536-539. New York: ACM Press.
- Machado, R. L. (2004). An introduction to tablet PCs. *Family Practice Management*, October 2004, 36-38.
- Mack, R., & Montaniz, F. (1991). A comparison of touch and mouse interaction techniques for a graphical windowing software environment. In *Proceedings of the Human Factors Society 35th Annual Meeting*, 286-289. Santa Monica, CA: Human Factors and Ergonomics Society.
- MacKenzie, S., Sellen, A., & Buxton, W. (1991). A comparison of input devices in elemental pointing and dragging tasks. *Proceedings of ACM CHI '91 Conference on Human Factors in Computing Systems and Graphics Interface*, 161-166. New York: ACM Press.
- Meyer, A. (1995). Pen computing: a technology overview and a vision. *ACM SigCHI Bulletin*, 27(3), 46-90.
- Meyer, S., Cohen, O., & Nilsen, E. (1994). Device comparisons for goal-directed drawing tasks. *Proceedings of CHI '94 Conference Companion on Human Factors in Computing Systems*, 251-252.
- Murata, A., & Iwase, H. (2006). Usability of touch-panel interfaces for older adults. *Human Factors*, 47(4), 767-776.
- Phillips, L. M. (2003). Effect of computer input device on speed and accuracy of older adults. *Masters Abstracts International*, 42 (05), 1842, (AAT 1418714).
- Proctor, R. W., Vu, K. L., Pick, D. F. (2005). Aging and response selection in spatial choice tasks. *Human Factors*, 47(2), 250-269.
- Rogers, W. A., Fisk, A. D., McLaughlin, A. C., & Pak, R. (2005). Touch a screen or turn a knob: Choosing the best device for the job. *Human Factors*, 47(2), 271-288.
- Sears, A., & Schneiderman, B. (1991). High precision touchscreens: Design strategies and comparisons with a mouse. *International Journal of Man-Machine Studies*, 34, 593-613.
- Smith, M. W., Sharit, J., & Czaja, S. J. (1999). Aging, motor control, and the performance of computer mouse tasks. *Human Factors*, 41(3), 388-396.

- Soukoreff, R.W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *Int. J. Human-Computer Studies*, 61, 751-789.
- Stanton, N. A. (2003). Human error identification in human-computer interaction. In J. A. Jacko & A. Sears (Eds.), *The Human-Computer Interaction Handbook* (pp. 371-383). Mahwah, NJ: Lawrence Erlbaum.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.
- Vercruyssen, M. (1997). Movement control and speed of behavior. In A. D. Fisk & W. A. Rogers (Eds.), *Handbook of Human Factors and the Older Adult*. San Diego, CA: Academic Press.
- Vicente, K. J. (1997). Heeding the legacy of Meister, Brunswik, & Gibson: Toward a broader view of human factors research. *Human Factors*, 39(2), 323-328.
- Walker, N., Millians, J., & Worden, A. (1996). Mouse accelerations and performance of older computer users. *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting - 1996*, 151-154.
- Worden, A., Walker, N., Bharat, K., & Hudson, S. (1997). Making computers easier for older adults to use: area cursors and sticky icons. *Proceedings of the CHI '97 Conference on Human Factors in Computing Systems*. New York: ACM.

Appendix A

Participant Recruitment Materials

Recruitment Script

Participant Screening Script
Thanks for your interest in our study. I need to ask you some questions first to make sure you qualify for the study.
I. Demographic information
What is your age?
You are a male/female?
Which is your preferred hand?
II. Computing experience
How many years have you been using a personal computer? Which personal computers do you use?
I'm going to ask about your experience with several software applications. For each one, please tell me if you use them on a daily basis (1), a weekly basis (2), a monthly basis (3), less than monthly (4), or never (5)?
<ul style="list-style-type: none"> • Microsoft Word
<ul style="list-style-type: none"> • Microsoft Excel
<ul style="list-style-type: none"> • Microsoft PowerPoint
<ul style="list-style-type: none"> • Any freehand drawing graphics creation program
For how many years have you been using a mouse for computer tasks?
Have you ever used a tablet PC, graphic tablet, or other type of pen computing device? Which?
<ul style="list-style-type: none"> • If yes, for how many years?
III. Physical capabilities
Do you have any hand tremors?
Do you have severe arthritis that affects your ability to use your hands?
Do you have normal vision with or without corrective lenses?
Do you have any physical limitations that would prevent you from doing an hour and a half of typical computer office tasks?

Recruitment website link from flyer

Web page hosted at <http://www.digivitz.com/research>.

Participants needed for new technology study at SJSU

Participants needed for new technology study at SJSU

Are you 18 to 30? Do you use a personal computer?

SJSU Graduate Student is seeking participants for a study on improving the experience of personal computing.

To qualify:

- you must be 18-30 years old;
- you must have some familiarity with typical office software; and
- you must be free of physical limitations like hand tremors or severe arthritis that would interfere with computing task

This 2 hour study will expose you to exciting new technology on the market, and your participation will help in the understanding of how people will use and respond to these technologies.

Study dates and times: Tue/Thu afternoons and eves; Wed/Fri *all day*; Sat/Sun *all day* - NOW through SEPTEMBER 27.

Study location: the study will be conducted at a facility near Cupertino by Highway 85 and De Anza Blvd. Details will be provided upon your acceptance in the study. You will need your own transportation to get to the facility.

Compensation: No monetary compensation is provided for this study. SJSU Psychology 1 students receive 2 credit hours for participating in research.

Contact Information - please leave me your phone number so I can call you back!

- Jamie Kravitz
- 408-297-0554
- research@digivitz.com

Recruitment advertisement used on craigslist.org for older adults

\$50 for 2 hour research study AGES 55-75

Reply to: gigs-208191176@craigslist.org

Date: 2006-09-16, 12:18PM PDT

PARTICIPANTS NEEDED - MEN AND WOMEN - for new technology study

Are you aged 55 to 75? Do you use a personal computer?

SJSU Graduate Student is seeking participants for a study on improving the experience of personal computing.

To qualify, you must have some familiarity with typical office software, and be free of physical limitations like hand tremors or severe arthritis that would interfere with computing tasks.

This 2 hour study will expose you to exciting new technology on the market, and your participation will help in the understanding of how people will use and respond to these technologies.

Study dates and times: September 17-24, 2006, morning, afternoon, and evening slots available! Now booking for Sunday 17th as well!

Study location: the study is being conducted at a facility in San Jose near Cupertino by 85 and De Anza Blvd. Details will be provided upon your acceptance in the study.

Compensation: \$50

Contact Information:

Jamie Kravitz

408-297-0554

or email response to this ad - be sure to provide phone number for callback

this is in or around Cupertino

no -- it's NOT ok to contact this poster with services or other commercial interests

Compensation: \$50 fee for 2 hours

Appendix B

Experimental Materials

Agreement to participate in research (consent) form

San José State
UNIVERSITY

Department of Psychology

DMH 157

One Washington Square
San Jose, CA 95192-1020
Voice: 408-924-5600
Fax: 408-924-5605
www.psych.sjsu.edu

**AGREEMENT TO PARTICIPATE IN RESEARCH AT
SAN JOSE STATE UNIVERSITY**

Responsible Investigators: James Kravitz and Kevin Jordan, Professor of Psychology

I have been invited to participate in research on differences between age groups when performing tasks with different computer input devices. The possible benefits I might gain from my participation include exposure to new, publicly available, consumer computer technology. There are no possible risks from participating other than minimal physical discomfort equivalent to what might occur from 90 minutes normal office work on a computer. I understand that my participation in the experiment is voluntary.

If I decide to participate, I will be asked to perform several common computer tasks using two different input devices: a mouse and a stylus (or "pen"). The tasks will be conducted on a convertible notebook/tablet personal computer. I will receive preliminary training and orientation to the computer, the devices, and the tasks to be performed. After each task I will be asked to feedback on my experience. I will be asked to fill out several surveys on my experience with the devices. The study will take approximately an hour and a half.

Data gathered from this study will be stored on a computer drive and videotapes which no one but the experimenter will be able to access. In case the results of this study are published, any information that is obtained from me in connection with this study and that can be identified with me will remain confidential.

My decision to participate or not participate will not in any way prejudice my future relations with San José State University. If I decide to participate, I am free to withdraw my consent and to discontinue my participation at any time without penalty.

If I have any questions, I may ask them prior to the start of the experiment. If I have any questions after the experiment, I may contact Dr. Jordan at 924-5626 or drop by DMH 342. If I have any complaints about the procedure, I may contact Dr. Sheila Bienenfeld, Psychology Department Chair at 924-5600 (DMH 157). For questions about research participants' rights, or in the event of research-related injury, I may contact Dr. Pamela Stacks (Associate Academic Vice President for Graduate Studies) at 924-2480.

I am making a decision whether or not to participate. My signature indicates that I have decided to participate having read the information provided above. My signature indicates that I agree to be videotaped for the purposes of this study. I have received a copy of this consent form for my records.

X _____ Date: _____
SIGNATURE

PRINTED NAME

SIGNATURE OF INVESTIGATOR (James F. Kravitz)

The California State University
University Center
Bakersfield, Chico, Clarks, Chico
Domestic Hills, East Bay, Fresno,
Fullerton, Hayward, Long Beach,
Los Angeles, Maritime Academy,
Monterey Bay, Northridge, Pomona,
Sacramento, San Bernardino, San
Diego, San Francisco, San Jose, San
Luis Obispo, San Marcos, Sonoma,
Stanislaus

Digit Span Test

I need to do a brief assessment of your short term memory. I'm going to say a short sequence of numbers to you and ask you to repeat them back to me. We'll do four each of 4, 5, and 6 digit number sequences. Any questions?

I'll begin with 4 digit sequences. Ready?

Repeat the 4-digit sequences back to me:

4 9 2 5

2 7 9 6

3 1 8 4

5 2 0 7

Next, Repeat the 5-digit sequences back to me:

8 5 1 9 2

4 7 3 0 6

7 5 2 9 4

1 8 5 2 9

Last set. Repeat the 6-digit sequences back to me:

7 0 2 6 9 5

2 7 9 1 3 8

3 8 6 9 1 4

8 2 5 3 0 6

Borg RPE CR-10 Scale

0	Nothing at all	"No P"
0.3		
0.5	Extremely weak	Just noticeable
1	Very weak	
1.5		
2	Weak	Light
2.5		
3	Moderate	
4		
5	Strong	Heavy
6		
7	Very strong	
8		
9		
10	Extremely strong "Max P"	
11		
↖		
●	Absolute maximum	Highest possible

Borg CR10 scale

© Gunnar Borg, 1981, 1982, 1998

Usability Dimensions Survey**Participant #** _____**Device** (circle one): MOUSE | PEN

A. For each of the following **statements**, mark an “X” in the box that best represents your **level of agreement**. Mark only one “X” per row.

		Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree
1	I think that I would like to use this device frequently					
2	This device fits my hand well					
3	I would recommend this device to a friend					
4	This device slowed me down					
5	I can imagine many cases where I'd want to use this device					
6	I would never use this device if I didn't have to					
7	I would prefer to use this device for all my computer tasks					
8	I was able to work efficiently with this device					
9	It felt comfortable to use this device					
10	Some tasks seemed easier with this device					

B. For each of the following **opposing word pairs**, mark an “X” in the box that best represents **your assessment** of the device. Mark only one “X” per row.

Natural						Awkward
Uncomfortable						Comfortable
Universal						Task Specific

User Preference Survey (page 1 of 2)

Participant # _____

- A. **Overall preference.** On a **scale of 1 to 5**, with 1 being the worst, and 5 being the best, **rate each of the two devices** you used today. Mark an "M" for the mouse and a "P" for the pen. Do not mark more than one letter in a box.

1 = worst	2	3	4	5 = best

- B. **Task preference.** Recall each of the tasks you performed today. For **each task**, indicate **which device you preferred** using. Mark only one device for each task.

<p>Task 1. Adjusting Windows</p>	<p>MOUSE ☐</p>	<p>PEN ☐</p>																																																															
<p>Task 2a. Excel Dragging Task</p>	<p>MOUSE ☐</p>	<p>PEN ☐</p>																																																															
<p>Task 2b. Excel Scrolling Task</p> <table border="1" data-bbox="397 1513 852 1753"> <tr> <td>A</td><td>A</td><td>8</td><td>11</td><td>8</td><td>8</td><td>11</td></tr> <tr> <td>A</td><td>A</td><td>12</td><td>17</td><td>12</td><td>12</td><td>17</td></tr> <tr> <td>8</td><td>11</td><td>8</td><td>25</td><td>18</td><td>18</td><td>25</td></tr> <tr> <td>12</td><td>17</td><td>12</td><td>35</td><td>24</td><td>24</td><td>35</td></tr> <tr> <td>18</td><td>25</td><td>18</td><td>0</td><td>-13</td><td>-13</td><td>0</td></tr> <tr> <td>24</td><td>35</td><td>24</td><td>53</td><td>36</td><td>36</td><td>53</td></tr> <tr> <td>30</td><td>0</td><td>-13</td><td>11</td><td>8</td><td>12</td><td>61</td></tr> <tr> <td>36</td><td>53</td><td>36</td><td>17</td><td>12</td><td>18</td><td>66</td></tr> <tr> <td>8</td><td>11</td><td>8</td><td>25</td><td>18</td><td>24</td><td>70</td></tr> </table>	A	A	8	11	8	8	11	A	A	12	17	12	12	17	8	11	8	25	18	18	25	12	17	12	35	24	24	35	18	25	18	0	-13	-13	0	24	35	24	53	36	36	53	30	0	-13	11	8	12	61	36	53	36	17	12	18	66	8	11	8	25	18	24	70	<p>MOUSE ☐</p>	<p>PEN ☐</p>
A	A	8	11	8	8	11																																																											
A	A	12	17	12	12	17																																																											
8	11	8	25	18	18	25																																																											
12	17	12	35	24	24	35																																																											
18	25	18	0	-13	-13	0																																																											
24	35	24	53	36	36	53																																																											
30	0	-13	11	8	12	61																																																											
36	53	36	17	12	18	66																																																											
8	11	8	25	18	24	70																																																											

User Preference Survey (page 2 of 2)

Participant # _____

<p>Task 3a. PowerPoint Dragging Task</p>	<p>MOUSE ☐</p>	<p>PEN ☐</p>
<p>Task 3b. PowerPoint Manipulation Task</p>	<p>MOUSE ☐</p>	<p>PEN ☐</p>
<p>Task 4. Tracing</p>	<p>MOUSE ☐</p>	<p>PEN ☐</p>
<p>Task 5a. Cross-Referencing: Form Entry</p>	<p>MOUSE ☐</p>	<p>PEN ☐</p>
<p>Task 5b. Cross-Referencing: Document Editing</p>	<p>MOUSE ☐</p>	<p>PEN ☐</p>

Experimenter data collection instrument

Participant #	Digit Span Errors:
---------------	--------------------

Device	☐ Mouse	☐ Pen
Participant OK with Device?	☐ Yes	☐ No

• **Start Camtasia and DVD!**

Task	Set:	OK?	RPE Phys	RPE Mentl	Time	Err	Notes
1 windows os							
2a excel dragging							
2b excel scrolling							
3a ppt moving							
3b ppt manip							
4 tracing							
5a form entry							
5b newsletter							

Device	☐ Mouse	☐ Pen
Participant OK with Device?	☐ Yes	☐ No

• **Start Camtasia and DVD!**

Task	Set:	OK?	RPE Phys	RPE Mentl	Time	Err	Notes
1 windows os							
2a excel dragging							
2b excel scrolling							
3a ppt moving							
3b ppt manip							
4 tracing							
5a form entry							
5b newsletter							

Appendix C

Procedural Scripts

General Instruction Script

Let me explain to you what we are going to do. Using this convertible Tablet PC, I'm going to ask you to perform a series of tasks first using one input device (the mouse/pen), then using another input device (the mouse/pen). For using the pen, the PC converts into a tablet and you use the pen directly on the surface of the screen. The tasks will be similar but not identical between each set. Before each task, you'll have a practice task to get used to using the device for the kinds of interactions required.

After you are finished with each task, I'll be asking you to provide a rating of perceived effort, which I'll explain more in a moment. After you are finished with all the tasks for one of the input devices, I'll be asking you to fill out a brief survey about that input device. Finally, after completing all tasks with both devices I'll be asking you to fill out a final survey about which device you preferred overall, and on a task by task basis.

The purpose of the study is NOT about your ability to understand or figure out the tasks themselves. Rather, it is about how well the input device supports you in your execution of the task. That's why there is a practice task first, and I'll ask to make sure you understand what you are being asked to do. If you don't understand a task, please ask me to clarify it before continuing.

We will also spend some time with each device before beginning any of the tasks to get you comfortable with using the device, and comfortable with the placement and position of the computer. Any questions?

Borg CR-10 Participant Instructions

You will use this scale to tell how strong your perception is of exertion to accomplish the task with this device. Ten, “Extremely strong – Max P,” is the main anchor. It is the strongest perception (P) you have ever experienced. It may be possible, however, to experience or to imagine something even stronger. Therefore, “Absolute maximum” is placed somewhat further down the scale without a fixed number and marked with a dot “●”. If you perceive an intensity stronger than 10, you may use a higher number. “Extremely weak,” corresponding to 0.5 on the scale, is something just noticeable, i.e., something that is on the boundary of what is possible to perceive.

Always start with a *verbal expression* and then choose a *number*. If your perception is “Very weak,” say 1; if “Moderate,” say 3; and so on. You are welcome to use half values (such as 1.5, or 3.5) or decimals (for example, 0.3, 0.8 or 2.3). It is very important that you answer what *you* perceive and not what you believe you ought to answer. Be as honest as possible and try not to overestimate or underestimate the intensities.

Any questions?

Borg CR-10 Test items

To see that you have understood the instruction and how to use the scale, please answer the following questions:

1. How black do you perceive a piece of pure black charcoal to be? (9) How white?
(0.5)
2. How loud do you perceive an ordinary conversation between two people to be?
(2.5)
3. How white do you perceive a piece of pure white sugar to be? (9) How black?
(0.5)
4. How sour do you perceive a lemon to be? (7)
5. How sweet do you perceive a ripe banana to be? (3.5)

Appendix D

Task Analysis

1. Excel dragging task

- Goal: Move all cells from row into column so they are in sequence
 - Sub-goal: Move individual cell from row to correct position in column
 - Drag cell (x10)
 - Choose cell to work with
 - Method 1. Numeric sequence low first
 - Method 2. Numeric sequence high first
 - Method 3. Random sequence/closest cell
 - Select cell
 - Discriminate move tool from select tool
 - Click and hold with move tool
 - Drag object to destination
 - Release cell to destination
 - Recenter to row of cells

2. PowerPoint dragging task

- Goal: Move all shapes inside the box so they are not touching each other or the box edges
 - Sub-goal: Move individual shape to a destination in the box (x29)
 - Choose shape to move
 - Choose by proximity to pen
 - Choose by size
 - Choose by shape/pattern (e.g., clockwise)
 - Select object click
 - Move object drag
 - Drop object release-click
 - Recenter/prepare for next

3. PowerPoint manipulation task

- Goal: Move all callout shapes to appropriate position and manipulate shape so it lines up callout with red dots
 - Sub-goal: move shapes away from center (x8)
 - Select object click
 - Move object drag
 - Drop object release-click
 - Recenter/prepare for next
 - Sub-goal: move shape to correct compass point (x8)
 - Identify correct location for callout
 - Select object click
 - Move object drag
 - Drop object release-click
 - Recenter/prepare for next
 - Sub-goal: manipulate shape so callout lines up with center box (x8)
 - Select callout object
 - Discriminate manipulate tool from move tool
 - Click and hold with manipulate tool
 - Identify correct destination
 - Drag callout element to destination
 - Release callout to destination
 - Recenter to next object

4. Tracing task

- Goal: Trace shape
 - Method 1 – continuous trace
 - Identify a starting point
 - Click and hold on start point
 - Drag along shape line
 - Control device so that continuous tracing remains accurate
 - Method 2 – trace in segments
 - Trace segment (x ??)
 - Identify a starting point *for segment*
 - Click and hold on start point
 - Drag along shape line
 - Control device so that continuous tracing remains accurate
 - Release device
 - Reposition device
 - Continue from where stopped
 - Start at another point

5. Form filling task

- Goal: enter form data from printed document to electronic document
 - Sub-goal: complete radio button questions (x5)
 - Identify which field to check from printed document
 - Method 1. One item at a time
 - Method 2. Memorize multiple items
 - Find same field on electronic form
 - Click on radio button
 - Reset to next item
 - Sub-goal: complete checkbox questions (x11, grouped as 5, 6)
 - Identify which checkboxes to activate from printed document
 - Method 1. One checkbox at a time per question
 - Method 2. Multiple checkboxes (column)
 - Method 3. Whole item
 - Find same field on electronic form
 - Click on check box
 - Reset to next item
 - Sub-goal: complete drop-menu questions (x16, grouped as 6, 3, 3, 4)
 - Identify which choice was marked on printed document
 - *First time only*: realize/translate boxes into drop-menu
 - Retain choice in STM while activating drop-menu
 - Click on drop-menu to open
 - Move pointer to correct choice
 - Click on correct choice
 - Reset to next item

6. Document editing task

- Goal: make changes to electronic document based on annotated printed document
 - Sub-goal: move pictures (x6)
 - Refer to printed document for item instructions
 - Identify corresponding object on electronic form
 - Click and hold on object
 - Drag object
 - Release object
 - Reset to next item
 - Sub-goal: text formatting changes (x7)
 - Refer to printed document for item instructions
 - Identify corresponding object on electronic form
 - Highlight text to reformat
 - Method 1. Click and drag across text
 - Method 2. Position cursor in left margin and click to select entire line
 - Identify correct button to hit
 - Click formatting button
 - Reset to next
 - Sub-goal: correct typos (x4)
 - Refer to printed document for item instructions
 - Identify corresponding text on electronic form
 - Highlight text to move
 - Identify which letter to select
 - Click and hold on one side of letter
 - Drag across letter
 - Release button
 - Drag letter
 - Click and hold on selected text
 - Drag to destination position
 - Release button
 - Reset to next

Appendix E

Omitted Tasks

Windows task

Note: this task was omitted from the final results because a disproportionate number of participants were not familiar enough with the Windows manipulations. Many had to be coached through the task, or referred back to the instruction sheet frequently. It was felt that too much cognitive effort was being exerted on the task, and as it was a short task to begin with, the data would not be reliable. It did serve, however, as a useful additional warm-up for participants.

Description. This task is a common manipulation of windows elements using Microsoft Windows XP Tablet Edition. Starting by opening a window with various files in it, the task requires the participant to change the view of the window from “icons” to “details”, adjust the width of the “Name” column so that longer file names are readable, move the “Date modified” column adjacent to the “Name” column, and sort the window by “Date modified” to show most recent files at the top of the list. This task involves menu selection, operating interface widgets, and dragging both with both a high level of accuracy (column width) and a lower level of accuracy (moving the column).

Windows Task. Starting screen (A) and finished screen (B).

Excel scrolling task

Note: this task was omitted from the final results because a disproportionate number of participants omitted steps of the task. Specifically, participants did not find all of the target cells because they did not search effectively. While these results may have been useful to look at search strategies, as the focus of the present study was on devices, it was determined that too many records were invalid.

The Excel scrolling task asks the participant to scroll through a large range of random numbers, find colored grouped boxes of letters, and change their formatting. This task utilizes scrolling (both vertically and horizontally), selecting, and accessing menu items, which are common goals in many applications. Refer to the following figure.

Search through all the data and change all the yellow boxed A's to bold, red type

A	A	8	11	8	8	11
A	A	12	17	12	12	17
8	11	8	25	18	18	25
12	17	12	35	24	24	35
18	25	18	0	-13	-13	0
24	35	24	53	36	36	53
30	0	-13	11	8	12	61
36	53	36	17	12	18	66
8	11	8	25	18	24	70

Appendix F

Letter of Approval from Human Subjects Institutional Review Board

San José State
UNIVERSITY

Office of the Provost
Associate Vice President
Graduate Studies & Research

One Washington Square
San Jose, CA 95192-0025
Voice: 408-924-2427
Fax: 408-924-2477

E-mail: gradstudies@sjsu.edu
<http://www.sjsu.edu>

To: James Kravitz
225 N. Morrison Ave #4
San Jose, CA 95126

From: Pamela Stacks, Ph.D.
Associate Vice President
Graduate Studies and Research

Date: August 30, 2006

The Human Subjects-Institutional Review Board has approved your request to use human subjects in the study entitled:

"Effects of input device on different age groups performing goal-oriented tasks"

This approval is contingent upon the subjects participating in your research project being appropriately protected from risk. This includes the protection of the anonymity of the subjects' identity when they participate in your research project, and with regard to all data that may be collected from the subjects. The approval includes continued monitoring of your research by the Board to assure that the subjects are being adequately and properly protected from such risks. If at any time a subject becomes injured or complains of injury, you must notify Dr. Pamela Stacks, Ph.D. immediately. Injury includes but is not limited to bodily harm, psychological trauma, and release of potentially damaging personal information. This approval for the human subject's portion of your project is in effect for one year, and data collection beyond August 30, 2007 requires an extension request.

Please also be advised that all subjects need to be fully informed and aware that their participation in your research project is voluntary, and that he or she may withdraw from the project at any time. Further, a subject's participation, refusal to participate, or withdrawal will not affect any services that the subject is receiving or will receive at the institution in which the research is being conducted.

If you have any questions, please contact me at (408) 924-2480.

cc. Kevin Jordan, DMH 342-0120

The California State University:
Chancellor's Office
Bakersfield, Channel Islands, Chico
Dominguez Hills, East Bay, Fresno
Fullerton, Hayward, Long Beach
Los Angeles, Maritime Academy
Monterey Bay, Northridge, Pomona
Sacramento, San Bernardino, San Diego
San Francisco, San Jose, San Luis Obispo
San Marcos, Stanislaus