Productivity and Multi-Screen Displays

A Comparison of Single Monitor, Multiple Monitor, and Multiple Monitor with Hydravision® Computer Displays over Simulated Office Tasks across Performance and Usability

James A. Anderson, Ph.D., F.I.C.A., Principal Investigator

Janet Colvin

Nancy Tobler

University of Utah

Initiated by:

Don Lindsay, Chief Investigator

Lindsay Research & Consulting, Inc.

Sponsored by:

Neil Rennert

Richard Mulcahy

ATI Technologies Inc.

Christopher Connery

NEC/MItsubishi

CIC Report 200311

Productivity and Multi-Screen Displays

Executive Summary

One hundred eight university and non university personnel participated in a comparison of single monitor, multi-monitor, and multi-monitor with Hydravision display configurations. Respondents edited slide shows, spreadsheets, and text documents in a simulation of office work, using each of the display arrays. Performance measures, including task time, editing time, number of edits completed, and number of errors made as well as usability measures evaluating effectiveness, comfort, learning ease, time to productivity, quickness of recovery from mistakes, ease of task tracking, ability to maintain task focus, and ease of movement among sources were combined into an overall evaluation of productivity. Multi-screens scored significantly higher on every measure.

Respondents got on task quicker, did the work faster, and got more of the work done with fewer errors in multi-screen configurations than with a single screen. They were 6 percent quicker to task, 7 percent faster on task, generated 10 percent more production, were 16 percent faster in production, had 33 percent fewer errors, and were 18 percent faster in errorless production. Multi-screens were seen as 29 percent more effective for tasks, 24 percent more comfortable to use in tasks, 17 percent easier to learn, 32 per cent faster to productive work, 19 percent easier for recovery from mistakes, 45 percent easier for task tracking, 28 percent easier in task focus, and 38 percent easier to move around sources of information.

Respondents were divided into three competency groups. The low competence group was significantly less able than the high competence group when both were working with a single screen but achieved near parity with single screen, high competence respondents when working in multi-monitor displays. The high competence group reasserted the initial difference when they moved to multiple screens. The low competence group increased the number of edits they completed but did not markedly increase the speed of their work when using multi-displays. High competence respondents increased both the number of edits and the speed of their performance when they moved to multi-displays.

Given the overwhelming consistency of both the performance and usability measures, multiple monitor configurations are recommended for use in any situation where multiple screens of information are an ordinary part of the work. There will be measurable gains in productivity and the work will be judged as easier to do. Multiple monitors are also recommended as cost effective where multi-screen tasks represents as little as 15 percent of the work for the highly competent, 17 percent for entry level competence and 21 percent for the general work force.

With the convergence of the technology of operating systems, display boards, and LCD monitors, these gains in productivity predict multi-monitor displays as a standard of the workplace.

Acknowledgments

This study began with Don Lindsay roaming the halls of the Department of Communication at the University of Utah looking for someone to talk to about productivity and multiple monitors. He stopped at my office in a moment of serendipity. Together we began to think through the conditions under which multiple monitor displays could be tested across performance and usability. Neil Rennert, Director of Research at ATI Technologies Inc., invited a proposal that he and Richard Mulcahy, Director of Marketing at ATI reviewed, made suggestions, and ultimately recommended. The sponsorship of ATI was joined by NEC Mitsubishi under the direction of Christopher Connery, Director of Product Marketing.

Janet Colvin and Nancy Tobler joined the research team as we moved into the final design and data collection phases. They were responsible for a good portion of the data collection and managing the experimental protocol. Nancy designed the questionnaires that Janet analyzed and conducted the analysis of the open-ended responses.

Kathleen Hom was the project ethicist. She ensured that the protocol was strictly adhered to and that IRB regulations were met.

Rebecca DaPra trained as an observer/facilitator and stepped in whenever help was needed.

Ethan Trump, with Jessica Sturm serving as on-screen talent, produced the training videos in record time and in the highest quality. My thanks to Paul Rose, video studio supervisor, for his support.

Katie Register managed all of the respondent contact and made sure appointments were made and met. Jennifer Tucker served as back-up and as a consultant in the task design.

The study could not have been done without the help and support of Ann Darling, Chair of the Department of Communication, who was generous in making the extensive and excellent facilities of the department available to the study.

The study greatly benefited from the 10 respondents who pre-tested all of the protocol materials and helped us bring the procedures into smooth working order.

Finally, my thanks to the 108 each of whom spent two hours or so working their way through the simulations. They were, to a person, the most valuable asset of the project.

James A. Anderson
Principal Investigator
Salt Lake City, Utah
July 18, 2003

Integrated three screen display illustration courtesy of Don Lindsay; used by permission.
All other illustrations by James A. Anderson.

Table of Contents

List of Figures

List of Tables

Productivity and Multi-Screen Displays

James A. Anderson, Ph. D., F.I.C.A., Principal Investigator

Janet Colvin

Nancy Tobler

Background and Introduction

With the advent of Windows 98 operating system, the PC platform has been able to support multi-monitor display configurations. Initially, multi-screen configurations found use in computer gaming, which has been the engine for most innovations in computer display, and in graphic design. As processor speed and memory capacity has increased and become less expensive, the office has found that it can support more open applications, so that multi-tasking could be a reality not just a term. The problem has been the management of the computer desktop. Even with increased monitor size, the single screen presented fundamental problems with window placement, stacking and tracking windows, multiple applications on the task bar, and the like. These problems have limited the increases in productivity theoretically possible with increased processor speed and memory capacity. The multi-screen display has provided some solutions.

Multi-screen Solutions

The multi-screen is a display configuration that can range from a fully integrated set of liquid crystal displays to a simple, physical arrangement of CRT monitors. Each screen or monitor in a multi-screen display is connected to the same computer through its own display port and is treated by the operating system both as an unified, boundaried space and as a connected or extended desktop. For example, an application will maximize to the boundaries of its "home" single screen but can also be "windowed" across all screens. Multi-screens or multi-monitor configurations (multi-screen will be used here as the general term for both) allow the user to place different windows on different screens or to spread a single application across all available screens (theoretically unlimited but usually 2-5).

Task efficiencies that can be produced from multi-screen displays are evident in a number of application conditions. Consider the simple example of transferring edits from one text draft to another. With the single screen, there are two common ways that this work is done. In the first case, the documents are opened full screen one on top of the other. The editor switches between the documents by clicking on the appropriate task bar location or by using the key combination of Alt + Tab. The switch returns the editor to the last insertion point in the working document. Problems occur when transporting an edit from the source document to precisely locate the position of placement in the destination document. This problem often entails switching back and forth between the two documents to find the exact position.

The other typical method is to "window" the two documents and arrange them in a half display side by side or top and bottom. The benefit of this approach is that the editor can see some of both texts at the same time. With most monitor sizes, however, the editor still cannot see the entire page or must reduce the view substantially, making reading difficult. Scrolling to reveal or straining to see slows the processes down.

In a multi-screen configuration, the two documents can be arranged in a parallel configuration so that both texts are in full view without any reduction or the need for horizontal scrolling. The editor can quickly confirm the placement of edits through direct observation.

Spreadsheets offer another example of where efficiencies can be gained. The typical monitor can show about 15 columns of a numerical spreadsheet. Spreadsheets that contain more than 15 columns have to be scrolled horizontally. In doing so the data processor often loses sight of the row titles, making data entry difficult. Large, complex graphical displays of data cannot be seen in their entirety.

With multiple screens, the large spreadsheet can be displayed across all the screens, a physical solution similar to unrolling the sheet on a larger table. The data entry operator can always track the proper row of entry and complex graphics can be accessed completely and manipulated with a full view of the results.

But even seemingly single screen tasks can be enhanced by multi-screen displays. For example, a graphics editor might have half a dozen sub menus opened on the graphics palette. It is a common problem that as the graphic design develops, the sub menus have to be moved, sized, or minimized to clear the space. In multi-screen, the sub-menus can parked on a separate screen and accessed there or brought over as needed or preferred.

Multi-Screen Management Software

Multi-screen management software adds another potential set of efficiencies. Multi-screen management software allows the user to instantly transport application windows to different screens, maximize applications across all displays, open child windows (e.g., multiple spreadsheets or tool and property sub-menus) on different displays, and to switch between virtual desktops (e.g., from a text editing setup to a graphics design set up).

The simplification of the “grab and drag” tasks of arranging windows across screens should offer some small benefits, but the real strengths of screen management software comes in the organizational properties that such software provides. Contemporary office work often requires multiple switching between tasks. The completion of a final report may require text elements, spreadsheet tables and/or graphs and graphical design elements all concurrently in process. Each of these tasks can be setup on its own desktop in the most efficient manner and accessed there.

Productivity and Multi-screen Displays

There is a small but growing multiple-monitor computer display industry that is a sub-set of the computer display industry as a whole. This industry includes vendors of multiple port video cards, multiple-screen computer displays, multiple-monitor software utilities, multiple-monitor applications and equipment. Flat panel LCD monitor manufacturers are also considered part of this market as the small form of the flat panel easily allows the physical placement of multiple monitors on a single desk–an option that bulky CRT monitors have not provided. While users of multi-screens attest to increased productivity, there has been no systematic study of productivity increases across ordinary office tasks using multi-screens and the software that manages them.

Productivity testing involves the reproduction of an ordinary work site, plausible and recognizable work tasks, and reasonable conditions of work. Participants are asked to complete a series of tasks in which performance differences are expected to appear across different display configurations. During the completion of the tasks, respondent practices are observed, time to completion measures and performance data are collected. Following participation in the work scenarios, respondents are asked to rate the display configuration for acceptability on an appropriate questionnaire and are debriefed concerning their experience.

Productivity testing itself is a combination of usability testing and performance testing. In usability testing, a sample group is asked to perform a set of tasks and *subjectively rate the* *ease of use* of a piece of hardware or software. In performance testing, automated tools collect facts about what the users *actually did* and *how long it took* them to do it. Because usability without increased performance or increased performance without adequate usability will not sustain overall increases in productivity, authentic measures of productivity must involve both.

Theoretical Basis for Increases in Productivity

There appears to be two bases in theory for predicting increases in productivity: rather straight forward notions of task efficiencies and somewhat more complex relationships between physical configurations and cognitive processing. We will take a look at each:

Task Efficiencies

Efficiencies are clearly to be gained in any task where the individual has to access more than one screen of information simultaneously. In the writing of this document, to use a simple example, the author was managing a text editing application and a graphics editing application simultaneously. As the adjacent graphic shows, the author was able to use one screen for the writing task and a separate screen to track and edit the JPEG files to be inserted. Certainly toggling back and forth between the applications is not particularly difficult, but it is slower than a flick of the eyes to the other screen.

The other advantage would appear to be a savings in setup time, particularly with a large number of short length tasks that have different formatting requirements. Moving for example from editing a spreadsheet using corrections from a memo that would require a side by side format to editing one text document from another that might be most effectively done in a top to bottom format.

For this study then, our expectations would be that for simulated office tasks that call for accessing multiple screens of information, total task time will be shorter with less time spent on setup when these tasks are completed using a multi-screen configuration than when using a single screen configuration. We also expect self-reports of higher effectiveness, greater productivity, and task focus.

Cognitive Processing

The physical placement of content appears to have a concomitant effect of anchoring the material in mental processing. Two cognitive abilities are central to the tasks of manipulating multiple documents simultaneously: the ability to distinguish between the documents and the ability to track location both individually and relatively to one another.

In single screen displays, it is quite easy to lose the identity of the document that is currently operational. This loss is particularly easy when the documents are nearly the same with relatively few distinguishing differences. Multi-screen displays, however, provide a concrete anchor for the identity of the document. Documents are in place and stay in place through out the editing process.

Tracking is the other important ability that is aided by multi-screen configurations. The single screen forces dislocations either through replacement or a reduced view. In multi-screens, place is always in view, which means that cognitive effort need not be expended to relocate where the editor is in each document. Again the writing of this document provides a good, if simple, example. The problem for the author is to match one of the dozens available illustrations with the written text. Toggling back and forth between the text and the index sheet means remembering the exact text while sifting through the many possible pictures. In multi-screen both text and pictures are immediately available for inspection.

Given these presumed advantages of multi-screen configurations, our expectations would include higher accuracy in multi-screen editing and self-reports of greater comfort, easier tracking, easier movement from source to source, and quicker recovery from mistakes for multi-screens over single screens.

Comparing Single and Multi-Screens over Performance and Usability

Overview

In order to test our theoretical suppositions concerning the efficiencies and efficacies of multi-screen configurations, an experimental comparison was devised using simulated office tasks. Three blocks of three tasks each were developed. Each block contained a text editing task (TXT), a spreadsheet editing task (SST) and a slide presentation editing task (PPT). Each task was designed to use six windows of information: two windows concerned the administrative, data collection, and simulation management of the experiment per se and four windows were components of the task. A seventh window provided navigational information that governed the entire session and the hyperlinks for the various files required.

Each of the 108 respondents completed a different block in each of the three configurations: single screen (SS), multi-screen (MS), and multi-screen assisted by multi-screen management software (HV)[1]. The order of tasks was the same in each block: text, spreadsheet and slide. An equal number of respondents (36 per block x configuration combinations)

[1] Hydravision, ATI screen management software, was used, hence the HV acronym.

completed each block to control for possible task by configuration differences. Screen configurations and tasks were used as "within subjects" factors in the analysis.

Strong order effects were to be expected as respondents learned how the task was to be performed. To control for these effects, an equal number of respondents (12 per each of the 9 block x configuration x order combinations) started the task set with a different configuration in the first position. Table 1 present the rotation of tasks and configurations.

Table 1

Order			**First**			**Second**			**Third**	
			Task			**Task**			**Task**	
Start	#Rs	Text	Spread	Slide	Text	Spread	Slide	Text	Spread	Slide
Single	12	GS	CR	MDY	SR	CS	WP	HV	PR	MDK
Single	12	HV	PR	MDK	GS	CR	MDY	SR	CS	WP
Single	12	SR	CS	WP	HV	PR	MDK	GS	CR	MDY
Multi	12	GS	CR	MDY	SR	CS	WP	HV	PR	MDK
Multi	12	HV	PR	MDK	GS	CR	MDY	SR	CS	WP
Multi	12	SR	CS	WP	HV	PR	MDK	GS	CR	MDY
HV	12	GS	CR	MDY	SR	CS	WP	HV	PR	MDK
HV	12	HV	PR	MDK	GS	CR	MDY	SR	CS	WP
HV	12	SR	CS	WP	HV	PR	MDK	GS	CR	MDY

Text tasks: Graduate studies, Screen Report, Hydravision
Spreadsheet tasks: Candidate Rankings, Products by Region, Customer Survey
Slide Tasks: Multi-Desk, Multi- Display, Window Placement

Table 1: Starting rotation of tasks and configurations

This procedure was repeated for each of the task sets. Order effects were, therefore, balanced across all configurations. In this manner, each respondent completed all 9 tasks in blocks of three and experienced all three screen configurations addressing them in one of three orders.

Finally, to get some sense of an "optimal" number of monitors, the multi-screen configuration was further divided into one with two monitors and one with three monitors. Half of the respondent pool (54) worked the tasks in a 2-monitor setup and half in a 3-monitor setup. This "monitor condition" was used as a "between-subject" factor in the analysis.

Tasks

All three tasks were based on the same conceptual framework. A destination text, spreadsheet, or slide presentation had been previously prepared and sent out for review or error correction. The copy edits and corrections had been returned to the respondent whose job was to make the changes on the destination file.

Text Tasks

The text files were prepared using Microsoft Word© with "track changes" enabled. The application was configured with only the menu bar and the reviewing toolbar showing. The toolbar had only the "Show," "Previous", and "Next" buttons on it. The task files consisted of the destination document on which all changes were to be made and two source documents (Mulcahy Edit and Tobler Edit) from which the changes were to be drawn. Each of the source documents had between 8 and 10 edits to be completed, including a requirement to open a graphics file, copy a logo appropriate to the text, and to paste the graphic at the bottom of the last page. (Appendix A contains a printed version of the tasks.)

The three texts were entitled *Graduate Studies*, *Hydravision*, and *Screen Report*.

Graduate Studies was a promotional description of the Communication graduate program at the University of Utah; *Hydravision* was a promotional description of the screen management software that the respondents were using or about to use; and *Screen Report* was a promotional description of this study. These texts were chosen because they represented familiar materials for the respondents and could be readily adapted by the respondents. The three texts were well populated with position markers such as paragraphing, headings, and graphics to assist the respondent in tracking location from one document to another.

Spreadsheet Tasks

The spreadsheet files were prepared using Microsoft Excel© and Microsoft Word©. Each spreadsheet was designed to cover approximately one and a half screens (average of 33 rows by 25 columns). Each of the data sets had summary information that was dynamically linked to a bar chart. Corrections were provided to the respondents in the form of a “Corrections Memo” simulating an e-mail addressed to them. Sixteen corrections were listed for the respondent to enter. After the corrections were made, the respondent was to copy the bar chart and paste it into a designated location in a “Final Report.” The Final Report was accessed by a hyperlink on the instructions page.

The spreadsheet files were titled *Candidate Rankings*, *Customer Survey*, and *Products by Region.* The *Candidate Rankings* file simulated the data collected on four candidates for a job position in human resources. *Customer Survey* simulated the responses of customers to 24 questions for each of four products. *Products by Region* simulated the data for material and labor costs and sales revenue for four products in 8 regions. Position markers included different colored blocks of columns and common color for the summary rows.

Slide Tasks

The slide files were prepared using Microsoft PowerPoint©. PowerPoint has a rather limited editing handling protocol (as compared with most word processing). Edits were identified in comments and placed in the source documents in color coded type. Each slide task had between 11 and 17 edits, including navigating to a graphics page, selecting a logo, and pasting the graphic into a new slide.

Each slide task consisted of 6 slides representing a complete segment drawn from a much larger presentation on the ATI software *Hydravision*©. The slides had been professionally prepared and each was a combination of graphics and text. The files were entitled *Multi-Desk, Multi-Display,* and *Window Placement*.

Data Collection

Data were collected in six ways: a paper and pencil intake questionnaire, automated time reports and automated usability questionnaire, stop watch measurements, task observations, and open-ended, end-of-testing questions. A description of each follows:

Intake Questionnaire

A single page intake questionnaire asked respondents to record their experience levels with computers, with the various applications used in the study, and with multiple screens. It also queried job experience and hours of work. A copy of the questionnaire is in Appendix B

Automated Time Report and Usability Questionnaire

An Excel spreadsheet was devised to collect the respondent's ID number, the time spent reading instructions, the total time spent on the task, and the responses to each of 8 usability questions.

Respondents would enter their ID number at the start of a task, click on the instructions “Start Box” that would record the current time, access the instruction files through a hyperlink, return to the Time Stamp to click on the instructions “Done Box” that would record the current time, time spent on instructions was then recorded as the difference between the start and done times. They would then click on the task “Start Box” to record the current time and access the task files through a hyperlink. When the task was completed they would click on the task “Done Box” to record the current time, total task time was recorded as the difference between the two times. Respondents would then scroll down to the usability questions. Each question text was followed by a 10 position slider that recorded the degree of agreement (disagreement) with the statement. A “Done” button at the bottom of the questionnaire posted a unique file with all the data recorded.

The usability questionnaire recorded the respondents self reports on their effectiveness, comfort, ease of learning, productivity, mistake recovery, task tracking, task focus, and ease of movement across sources.

Stop Watch Measurements

Stop watch timing was initiated at the start of the actual editing task. Each task had its own marker events for the start and completion of editing. An observer/facilitator (O/F) was seated next to the respondent started the watch on the initiation event and stopped it on the completion event. The time values were recorded on the task observation sheet in minutes and seconds.

Task Observations

As noted, an observer/facilitator (O/F) was seated with each respondent during the entire time of the session. As an observer (facilitation practices are described under "Protocol"), the O/F was responsible for stop watch data, recording the correct completion of each edit, recording any missed edits and errors in editing or changes otherwise introduced into the source documents, recording any comments about the task or the screen configuration, and any unusual practices in the editing task that appeared worthy of notice. All notations were recorded on the task observation sheet. The number of edits completed, number of editing errors, and number of missed edits became primary data on productivity and accuracy.

Post-Session Questions

At the completion of all the tasks, the observer/facilitator asked four questions: "Focusing on single screen versus multiple screens, what did you think about that difference" "Focusing on multiple screens with Hydravision and multiple screens without Hydravision, what did you think about that difference?" "Focusing on the tasks and the different screen configurations, did any task seem easier or harder in a given screen configuration?" Focusing on the experiment itself, was there anything that bothered you or that we should do differently?" A summary of the respondent's answers was recorded on the task observation sheet.

Protocol

This study was conducted under the supervision of the Institutional Review Board (IRB) at the University of Utah.[2] It was conducted in the Department of Communication's research facilities.

Sampling

Announcements of the project were made to advanced undergraduate and graduate level classes in the Department of Communication. Flyers were distributed throughout the department. Faculty and staff were notified through snowball sampling—individual respondents would identify other potential respondents who would then be notified. Snowball sampling was also used to identify adults not associated with the University. All respondents were volunteers. The volunteers completed a registration form, providing their contact information. Each respondent was individually contacted and an appointment made to participate in the study.

The sample totaled 108 respondents equally divided between the 2-monitor and 3-monitor conditions (54 respondents each). The majority of the sample was composed of advanced undergraduate and graduate students at the University of Utah. University of Utah students are older (average undergraduate age is 25; average graduate is 32), more likely to be married, and married with children, and more likely to have a full time job outside of school than the traditional undergraduate student.

---

[2] Institutional Review Boards are federally mandated in the United States for the protection of human subjects and are under the supervision of the US Department of Health and Human Services

Testing Procedures

When the respondent appeared at the appointed time, s/he was greeted by project O/F assigned to that time slot. The respondent was given a short description of the study, a signed copy of the informed consent document, and the intake questionnaire to complete. The respondent was then shown one of three 5-minute training videos, SS, MS, or HV depending on the initial configuration of the task. The training video demonstrated a set of editing procedures appropriate to each task in the block and to the specific screen configuration.

At the conclusion of the video, the O/F described the screen configuration that was in use, the tasks to be done, and the role the O/F would play in the process. When all questions were answered, the respondent was asked to navigate to the first Time Stamp screen to begin the block session. At the completion of the respondent's reading of the instructions, the O/F provided a brief summary and asked for questions. When those were answered the task itself was initiated. The O/F assisted the respondent in finding and opening the task files. As those files were being opened, the O/F pointed out the destination document and the source documents, indicated the specifics of the task and repeated the relevant training video information. Again the O/F answered any questions. When the respondent initiated the editing task, the stop watch was started. Respondents were given 5 minutes to complete the task, although time was added to allow the completion of an edit in progress.

The O/F recorded each edit as it was made. Errors and missed edits were also recorded. The O/F offered no advice or instruction on how to do the task, while the task was in progress, but did answer any questions directly without elaborating or instructing ("Is this the destination document?" "Yes."). At the conclusion of the task, the stopwatch was stopped, the time recorded, and the respondent immediately directed back to the Time Stamp. The respondent checked the task "Done Box" completed the usability questionnaire, and posted the file. The O/F then helped the respondent close all task files and return to the navigation screen to start the next task of the block.

At the conclusion of all three tasks in the configuration block, the respondent was informed that s/he would now change configurations and that there was a training video available for that configuration. The appropriate video was shown while the O/F changed the configuration. In changing from SS to MS, the O/F simply enabled the other monitors by changing the control panel settings. Going in the other direction the additional monitors were disabled. No physical changes were made to the setup. Going from MS to HV involved activating *Hydravision.* Going from SS to HV involved changing the settings and activating *Hydravision. Hydravision* was deactivated in the MS and SS conditions.

At the conclusion of all three task blocks the post-session questions were asked and answers recorded. Each session took approximately 90 minutes. Respondents were paid $20 for their time.

Project activities were under the supervision of a project ethicist whose responsibility was to ensure that all procedures were followed by the O/F and other project staff. The project ethicist made random visitations and observed entire sessions. Her final report noted no violations.

Facilities

Testing was done in the University of Utah, Department of Communication interaction laboratory. This testing facility has the look and feel of a living room (albeit one with a large one-way mirror and video cameras in the corners) with couches, easy chairs and a large television set. Two work tables were added for each of the testing stations.

Each testing station was configured with a new PC computer with a clean install of Windows XP and Microsoft Office Suite. The computers were based on the Intel Pentium 4 chip running at 1.8 GHz, with 512MB DDR SDRAM, a 40 GB, 7200rpm Ultra ATA hard drive, standard keyboard, and two button wheel Mouse. Monitors were NEC Mitsubishi Multisync LCD 1855NX, an 18 inch liquid crystal display. Display boards were ATI Radeon 9000 AGP with two monitor ports and ATI Radeon 7000 PCI with a single port. One station had two monitors arranged in a slight V with the right hand monitor having the Taskbar; the other had three in a triptych arrangement with the center monitor having the Taskbar.

The training videos were made at the University of Utah Department of Communication video recording study on digital video by a professional videographer. The digital tapes were transferred to individual VHS tapes for the project.

Data Handling: Performance Measures

This section reports the variables used, their definitions, and error corrections applied.

Basic Variables and Their Definitions

Five variables used to test performance were collected automatically or through direct observation. The variables, their definitions and method of collection are reported below:

Task Time: One of two basic time units. Task Time is the lapsed time from the respondent's checking of the task "Start" Box on the Time Stamp to the Respondent's checking of the "Done" box on the Time Stamp. Task Time includes setup time and edit time plus any time spent in meeting project requirements (navigating to and from the Time Stamp, for example). Task Time was an automated data collection.

Edit Time: The other basic time unit. Edit Time is the stopwatch recorded time from the first editing marker event to the last editing marker event. It represents the amount of time actually on task and has no other time component. The time was recorded by the O/F assigned to the respondent.

Number of Correct Edits: The number of correctly executed edits as observed and recorded by the O/F. Each task was broken down into the distinct subtasks called for in the editing. For example, a deletion or an insertion was considered a single edit. A deletion and an insertion (delete one word and insert another) was considered two edits. Navigating to another source document, acquiring an element and inserting or pasting that element were each considered a single edit. These definitions varied slightly across the tasks. In the spreadsheet tasks all corrections called for the substitution of a correct number for the one entered. Technically, this is a deletion and an insertion, but the task is a simple type over. One does not actually delete and insert. Similarly, in the slide task, there was a single substitution (the word "the" for the word "your") and in the text task there was a letter substitution ("n" for "s") that were both counted as a single edit because the actual practice was a type over. Each of these edits were listed for each task on the task observation sheets. The O/F checked off each edit as it was completed or recorded an error or miss as described below.

Number of Errors: The O/F recorded an error when the edit called for was completed incorrectly. An error was defined as any event that would have to be "found" and "corrected" by another editing process. In the text task, errors included such events as incorrect placement, incomplete deletion, over-deletion, but not misspelling (as that would be automatically flagged). In the spreadsheet tasks, errors included entering the wrong value and incorrect placement (wrong row or column). In the slide tasks, errors included text editing errors and graphic selection and placement errors.

Number of Missed Edits: The O/F recorded a missed edit when the respondent skipped a complete edit (partial edits were considered errors).

Derived Variables and Their Definitions

Five performance variables were derived through calculations using the basic variables as factors. Those variables and their definitions are:

Proportion of Edits Completed: The number of correct edits divided by the total number of edits required by the task.

Accuracy: The number of correct edits minus the number of errors and missed edits. Accuracy is a performance cost measure. Inaccurate editing increases costs as the task has to be redone. The greater the inaccuracy the less confidence can be given to the original work and the more care required in the re-editing.

Proportion of Accurate Edits: The accuracy coefficient (number of correct edits minus the number of errors and missed edits) divided by the number of edits required.

Time per Edit: Edit time divided by the number of correct edits. This measure can be used to project the time required for larger tasks.

Time per Accurate Edit: Edit time divided by the accuracy coefficient (number of correct edits minus the number of errors and missed edits).

Single Screen Proficiency: The number of edits completed were summed and averaged across all three tasks in the single screen configuration and were divided into four roughly equal sized groups. Edit times summed and averaged across all three tasks in the single screen configuration were also divided into four roughly equal groups. These group values were cross multiplied producing a discontinuous score from 1 to 16. The distribution was divided into 3 roughly equal sized groups representing high, moderate, and low overall proficiency to correspond to the three groups of the Expertise variable.

Block Variables and Their Definitions

Eight block variables were constructed from the task values. Block variables were summed across all nine tasks (3 task types by 3 cases of each type) used in a given screen configuration. Because all 9 tasks appeared in equal number for each screen configuration the values are directly comparable. Block variables represent the performance characteristics across a period of various work requirements.

Block Task Time: Task time was summed across each of the three task types (text, spreadsheet, slide) in a screen configuration (SS, MS, HV). Block Task Time represents the ordinary office routine of going from task to task over the course of a day.

Block Edit Time: Edit times summed across each task in a screen configuration.

Block Number of Edits: The sum of the completed edits over each task in a screen configuration.

Block Number of Errors: The sum of the editing errors over each task in a screen configuration.

Block Number Misses: The sum of the missed edits over each task in a screen configuration.

Block Accuracy: The sum of the Accuracy scores over each task in a screen configuration.

Block Time per Edit: The quotient of Block Edit Time divided by Block Number of Edits.

Block Time per Accurate Edit: The quotient of Block Edit Time divided by Block Accuracy.

Data Entry and Correction

The automated procedures of the Task Time variable created a unique file for each task by configuration combination that required the Task Time values and the task/configuration descriptors. Each respondent generated nine such files. The observed editing time, edit performance (edits completed, errors, and missed edits), and the order in which the task was completed were entered by hand into these files. These files were concatenated into a single spread sheet for each respondent. The derived variable manipulations were made and the entire data set was entered into an SPSS data file with each row containing all of the data from a single respondent (108 respondents by 191 values/descriptors). The hand entered data were then verified in what approximated a double entry system.

The study generated nearly 5,000 basic performance values and another 9,000 derived values. With this number of data points the possibility of at least some error in the data base is high. All variables were checked for missing or out of range entries. Twenty-three cases (1 percent) had missing or out of range data in either the Task Time or Edit Time variable. Out of range Task Time entries were truncated to 600 seconds. Missing Task Time entries were recorded as Edit Time plus 60 seconds (an approximate average of the difference across tasks). Missing Edit Time values were recorded as Task Time minus 60 seconds. The very small number of corrections indicated that they would have little artificial effect on the analysis but would preserve the integrity of the data set.

More complex corrections were called for in the derived variables. The major difficulties occurred in the division by the number of edits or number of accurate edits when the divisor was zero or 1. There were 16 such cases distributed (less than 1 percent) over all task and configurations. These cases created a balloon value that distorted the mean and distribution of the variable. To control for these anomalies each of the task/configuration derived variables were examined. An anomaly was defined as being at the 98th percentile and at least 35 points above the 97th percentile. These anomalies were truncated to the 97th percentile values. This truncation was further limited in that the Time per Accurate Edit had to be higher than the Time per Edit value when the accuracy coefficient was lower than the number of edits value. The next higher score value was used.

The rationale for these corrections is straight forward. Some correction has to be applied in the division by zero case as that calculated value makes no sense. Our observations indicated that it took a few respondents several minutes to execute the first edit in their first experience as they acclimated to the testing situation. It is unreasonable to presume that this circumstance is representative of an actual time per edits, but is rather a function of the restricted event space or the conditions of the simulation. In a typical example, a data set would show 97 percent of the values at 48 or lower and then a single value of 200 or more. Using the truncation policy preserved the character of the performance's contribution to the mean and variance without allowing it to unduly distort the value of either. Again the very small percentage of corrections should show no effect on the relative standings of the configurations but did allow more realistic values in each.

Data Handling: Questionnaires

Two questionnaire instruments were used in this study: an intake questionnaire that queried respondents on their computer, application, and multi-screen experience and a usability questionnaire administered after every task performance. An open ended interview based on four questions followed the testing session.

Intake Questionnaire

The intake questionnaire was a paper and pencil device composed of 6 sections:

Computer Expertise: A four point scale ranging from zero (None) to 3 (Advanced).

Application Expertise: A four point scale ranging from zero (None) to 3 (Advanced).

Block Expertise: The average of the three Application Expertise measures divided into three roughly equal groups. Cut points were (1) less than 1.67, (2) equal to 1.67, and (3) 2.00 or greater. Corrections were made for anomalous cases (described in the Performance by Expertise section).

Time Spent on Text, Spreadsheet, and Slide Applications: In hours per week from zero to ten.

Level of Application Use: A three point scale from one (Personal) to three (Professional).

Multi-screen Experience: A “yes” “no” item followed by the number of monitors used (1-6).

Current Job Situation: Number of hours per week on the job and job title.

Data were hand entered with double entry verification.

Usability Questionnaire

Each task performance was immediately followed by a usability questionnaire that was part of the Time Stamp file. The questionnaire used a 10-point slider to register the self-reported position between the poles of Strongly Disagree and Strongly Agree. As reported above, the items recorded the respondents’ self reports on their effectiveness, comfort, ease of learning, productivity, mistake recovery, task tracking, task focus, and ease of movement across sources.

Data were recorded by the same procedures used in collecting the time data and directly entered into the data bas.

Interviews

Respondents were asked to compare single and multi-screens, multi-screens with software and without, task difficulty in different configurations, and to comment on the protocol itself. A summary of each response was recorded by the O/F and entered verbatim into the data file.

The three test related questions, as reported above, were: “Focusing on single versus multiple screens, what did you think about the difference?” “Focusing on multiple screens using Hydravision and multiple screens without Hydravision, what did you think about that difference?” “Focusing on the tasks and the different screen configurations, did any task seem easier or harder in a given screen configuration?” In general, the interviews were conducted without additional probes. However, on occasion, participant’s responses to a questions led to a more in-depth discussion of that issue. Interviews typically lasted five minutes or less.

Interview comments were first parsed into thought units. A thought unit was a phrase that could stand alone and be understood. The free-standing thought unit was further parsed into a comment about a single task or screen configuration. For example, “I liked the multi monitors better” was a single unit about a screen configuration, but the phrase “All three applications were easier in multi-screen” parsed into 6 thought units, one for each application and 3 for the multi-screen configuration. If a screen configuration was implied, it was coded to that category. Using these rules, 1,293 thought units were identified, with 797 thought units on screen configurations.

The screen configuration thought units were then coded in the following seven category system to expand understanding of screen configurations. Once categorized, the thought units were then evaluated by positive, negative, or neutral valence. Table 2 presents the coding categories and an exemplar response for each. Appendix C provides all comments.

Table 2

Coding Categories	Examples
1. Usability	"Multi-screen is easier to use" "In single screen, it was harder to remember"
2. Affect	"Multi-screen was fun" "Multi-screen was overwhelming."
3. Experience	"I use multi-screen at work" "I am use to single screen."
4. Familiarity	"Once you got use to Hydravision" "I caught on quickly to multi-screen"
5. Comparative	"Hydravision was better than multi-screen" "Multi-screen was much easier than single screen."
6. Anticipated Usability	"I can see how I would use multi-screen at work" "If I were doing many tasks, HV would be helpful."
7. Cognitive Framing	"Hydravision was like having a book open," "Single screen is like shuffling papers."
Single Screen	
Single Screen Positive	"I like single screen best."
Single Screen Negative	"I felt stifled by single screen."
Single Screen Neutral	"I have always used single screen."
Multi-screen	
Multi-screen Positive	"I like multi-monitors better."
Multi-screen Negative	"Space may be a problem with multi-monitors."
Multi-screen Neutral	"I use multi monitors now."
Hydravision	
Hydravision Positive	"I liked HV. It organized things a lot better."
Hydravision Negative	"I don't feel it really helps that much."
Hydravision Neutral	"Hydravision would take some getting use to."

Usability comments were those made about how the screen configuration worked.
Affect comments were those that referred to emotions connected with a screen configuration experience.
Experience comments were those that indicated a prior experience with a screen configuration.
Familiarity was experience during the study.

Comparative was used to identify instances when the comment directly compared two screen configurations.
Anticipated Usability category included comments on projected use in the future with a screen configuration.
Cognitive Framing included those comments which the participant described how they "saw" the screen configuration.

Table 2: Coding categories and examples for interview responses.

Analysis and Results: Performance Data

This section reports the data transformations, statistical design, analytic procedures and findings for the 12 basic and derived performance variables, the 8 block variables, and the analysis of performance over expertise.

Data Preparation

In each of the 12 basic and derived variables, data were reorganized from their original task-specific entry into a task-type centered entry that distributed both order of performance and specific task in balanced numbers through out the data. Each task-type data set had an equal number of the three tasks and three orders. For example, in the text task, the data set had 36 respondents from each of the *Graduate Studies, Hydravision,* and *Screen Report* tasks. Within those specific tasks 12 respondents each had done the task in the first order, 12 in the second, and 12 in the third accounting for all 36. This reorganization was done in the SPSS data file using an independent verification of all syntax and a random hand check of each data set. All findings are reported by task type: slide, spreadsheet, and text.

Statistical Design

All respondents did all task-types in all screen configurations (a different version of the task type was used in each configuration). All performance variables are "within subjects" or "repeated measures" variables. This design controlled for inter-subject differences based on differential experience or expertise in computer or program use that most likely would have been introduced in an independent groups design. Skill and experience levels were held constant by having the same respondent perform in all three configurations providing a "fair" comparison.. The two "within" variables in this design, then, were task types (**Tasks**) and configurations (**Screens**). The task types were **Slide, Spreadsheet,** and **Text**. The three configurations were single screen (**SS**) multi-screen (**MS**) and multi-screen with *Hydravision* (**HV**).

The testing conditions of a two-monitor or three-monitor station was a "between subjects" or "independent groups" factor in the design. Half of the respondents (54) went through the protocol in each of these **conditions**. Figure 1 diagrams the design.

Figure 1

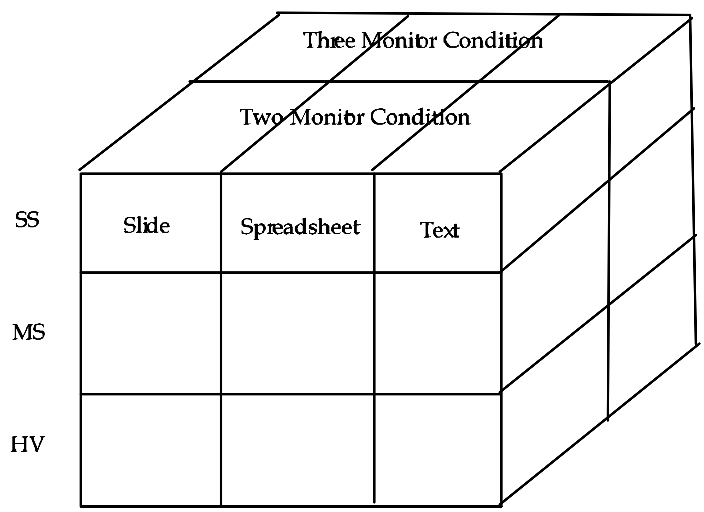

Figure 1: Statistical design for each performance variable, tasks by screens by conditions.

Each variable (except the block time variables) was analyzed using this classic "Type III" design using the General Linear Model as formulated in SPSS. An Alpha of .05 was set as the decision criterion for significance.

Results: Performance Basic and Derived Variables

The results for each of the 12 variables is presented in turn. Each section starts with tests of significance in the three-factor (tasks by screens by conditions), each of the two-factor (screens by tasks, screens by conditions, tasks by conditions) and main effects (screens, tasks, and conditions). A table of the means, standard deviations, and confidence intervals by cell is then presented followed by tables of means and standard deviations for each significant condition. The reader is reminded that significant interactions at one level confound the analysis of the next lower level (three-factor confounds two-factor confounds main effects). The results will be discussed only to the lowest non-confounded level.

Task Time

The Task Time variable measured the amount of time from the opening of the first task file to the clicking of the task "Done" box on the time stamp. It represented the total work time required to do the task. Table 3 presents the analysis of variance results.

Table 3

Task Time Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks by Condition	2.41	4, 424	.049
Screens by Condition	.069	2, 212	.933
Tasks by Condition	1.246	2, 212	.290
Screens by Task	7.120	4, 212	.000
Screens	9.643	2, 212	.000
Tasks	243.130	2, 212	.000
Conditions	.092	1, 106	.763

Table 3: Analysis of variance results for Task Time.

Table 4 presents the means and standard deviations for tasks and screens by condition for the Task Time variable. Table 5 presents a comparison of the SS means with the MS means for the two monitor conditions. Table 6 presents a comparison of the SS means with HV means for the monitor conditions. In both of the latter tables significant differences are noted.

Table 4

Task Time

Monitors	Screens	Tasks	Mean	Standard Error	95% Confidence Interval	
					Lower Bound	Upper Bound
Two Monitors	Single	Slide	398.892	10.621	377.835	419.949
		Spreadsheet	331.774	9.546	312.849	350.699
		Text	401.767	8.009	385.888	417.645
	Multi-screen	Slide	388.468	9.110	370.408	406.529
		Spreadsheet	292.214	10.486	271.424	313.004
		Text	374.446	10.329	353.967	394.925
	Hydravision	Slide	372.825	10.638	351.734	393.916
		Spreadsheet	306.189	9.137	288.074	324.304
		Text	380.814	8.950	363.069	398.559
Three Monitors	Single	Slide	391.690	10.621	370.633	412.747
		Spreadsheet	343.443	9.546	324.518	362.368
		Text	398.371	8.009	382.493	414.250
	Multi-screen	Slide	393.375	9.110	375.314	411.436
		Spreadsheet	286.981	10.486	266.191	307.771
		Text	389.523	10.329	369.044	410.002
	Hydravision	Slide	381.620	10.638	360.529	402.711
		Spreadsheet	288.405	9.137	270.289	306.520
		Text	397.876	8.950	380.131	415.622

Table 4: Conditions by screen configurations by tasks means, standard errors, and confidence intervals for Task Time.

Table 5

	Task	Single Mean	Multi Mean	Difference	Percent Change	Significant
Two Monitors	Slide	398.892	388.468	10.424	3	No
	Spreadsheet	331.774	292.214	39.56	12	Yes
	Text	401.767	374.446	27.321	7	Yes
Three Monitors	Slide	391.690	393.375	-1.685	-0	No
	Spreadsheet	343.443	286.981	56.462	16	Yes
	Text	398.371	389.523	8.848	2	No

Table 5: Comparison of SS screen Task Time means with MS Task Time means, difference, percent of change, and significance for each monitor condition.

Table 6

	Task	Single Mean	Hydravision	Difference	Percent Change	Significant
Two Monitors	Slide	398.892	372.825	26.067	7	Yes
	Spreadsheet	331.774	306.189	25.585	8	Yes
	Text	401.767	380.814	20.953	5	Yes
Three Monitors	Slide	391.690	381.620	10.070	3	No
	Spreadsheet	343.443	288.405	55.038	16	Yes
	Text	398.371	397.876	0.495	0	No

Table 6: Comparison of SS screen Task Time means with HV Task Time means, difference, percent of change, and significance for each monitor condition.

The significant screens by tasks by conditions interaction indicates that the differences among the various means were not consistent over the various factors. Inspection of Tables 4 through 6 shows that task time decreased in 11 of the 12 comparisons when moving from a single to either multi-screen condition and was slightly higher in one slide task. Consistent and large differences were found with the spreadsheet task. All of these time reductions were significant. Three of the four slide task comparisons show time reduction, but only one was significant. The text task showed significant time reductions in both MS and HV configurations in the two monitor setup but not for the three monitor setup.

None of the difference between two monitor and three monitor conditions were significant. However there was a consistent pattern of differences across tasks. The three monitor condition showed nearly significant reductions in the spreadsheet tasks but the direction was reversed for both the slide and text tasks.

Edit Time

The Edit Time variable measured the time lapsed between the first task marker event and the last task marker event. It can be considered on-task time. Table 7 presents the analysis of variance results.

Table 7

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks by Condition	2.732	4, 424	.029
Screens by Condition	.013	2, 212	.987
Tasks by Condition	.643	2, 212	.527
Screens by Task	9.915	4, 424	.000
Screens	21.254	2, 212	.000
Tasks	200.681	2, 212	.000
Conditions	.026	1, 106	.872

Table 7: Analysis of variance results for Edit Time.

Table 8 presents the means, standard deviations, and confidence intervals for tasks and screens by condition for the edit time variable. Table 9 presents a comparison of the SS means with the MS means for the two monitor conditions. Table 10 presents a comparison of the SS means with HV means for the two and three monitor conditions. Significant differences are noted in these latter tables.

Table 8

Edit Time

Monitors	Screens	Tasks	Mean	Standard Error	95% Confidence Interval	
					Lower Bound	Upper Bound
Two Monitors	Single	Slide	275.444	6.165	263.221	287.668
		Spreadsheet	243.037	7.326	228.512	257.562
		Text	297.481	3.724	290.099	304.864
	Multi-screen	Slide	266.796	6.148	254.608	278.984
		Spreadsheet	213.056	7.721	197.747	228.364
		Text	279.759	6.467	266.937	292.581
	Hydravision	Slide	261.611	7.112	247.511	275.712
		Spreadsheet	217.648	7.361	203.054	232.243
		Text	280.352	5.006	270.426	290.277
Three Monitors	Single	Slide	272.833	6.165	260.610	285.057
		Spreadsheet	253.389	7.326	238.864	267.914
		Text	293.648	3.724	286.265	301.031
	Multi-screen	Slide	275.148	6.148	262.960	287.336
		Spreadsheet	209.296	7.721	193.988	224.605
		Text	279.074	6.467	266.252	291.896

	Hydravision	Slide	270.889	7.112	256.788	284.989
		Spreadsheet	204.389	7.361	189.794	218.983
		Text	285.389	5.006	275.463	295.315

Table 8: Conditions by screen configurations by tasks means, standard errors, and confidence intervals for Edit Time.

Table 9

	Task	Single Mean	Multi Mean	Difference	Percent Change	Significant
Two Monitors	Slide	275.444	266.796	8.648	3	No
	Spreadsheet	243.037	213.056	29.981	12	Yes
	Text	297.481	279.759	17.722	6	Yes
Three Monitors	Slide	272.833	275.148	-2.315	-1	No
	Spreadsheet	253.389	209.296	44.093	17	Yes
	Text	293.648	279.074	14.574	5	Yes

Table 9: Comparison of SS screen Edit Time means with MS Edit Time means, difference, percent of change, and significance for each monitor condition.

Table 10

	Task	Single Mean	Hydravision	Difference	Percent Change	Significant
Two Monitors	Slide	275.444	261.611	13.833	5	No
	Spreadsheet	243.037	217.648	25.389	10	Yes
	Text	297.481	280.352	17.129	6	Yes
Three Monitors	Slide	272.833	270.889	1.944	1	No
	Spreadsheet	253.389	204.389	49.00	20	Yes
	Text	293.648	285.389	8.259	3	Yes

Table 10: Comparison of SS screen Edit Time means with HV Edit Time means, difference, percent of change, and significance for each monitor condition.

Again, the significant three-factor interaction requires analysis at the cell level. Eleven of the 12 comparisons between single screen and multi-screen configurations showed reductions in editing time. These differences were significant in eight of these comparisons. Only the slide task failed to show significant or consistent reductions with the three-monitor MS condition showing a reversal. The spreadsheet tasks showed the largest reductions of time across both monitor conditions. Slide and text editing was done more quickly in the two-monitor condition;

spreadsheet editing was faster in the three-monitor condition. None of the differences were significant, although nearly so in the spreadsheet task.

Number of Edits

The "Number of Edits" variable gave a count of the number of edits correctly entered by the respondent. This measure is a typical measure of productivity (number of units produced). Table 11 presents the analysis of variance results.

Table 11

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks by Condition	1.330	4, 424	.258
Screens by Condition	.192	2, 212	.826
Tasks by Condition	.390	2, 212	.678
Screens by Task	4.796	4, 424	.001
Screens	25.541	2, 212	.000
Tasks	585.306	2, 212	.000
Conditions	.287	1, 106	.594

Table 11: Analysis of variance results for Number of Edits.

The results in Table 11 indicated that the differences among screen configurations changed over tasks. The lack of a significant three-factor interaction or any of the two-factor interactions involving the condition of two or three monitors signals that the configuration means and the task means remained consistent over the monitor conditions. Although the three-factor interaction was not significant, Table 12 presents the cell means and the single screen multi-screen comparisons to keep the data record consistent for the reader. This table will be followed by a table of means summed over the non-significant factor of conditions.

Table 12

Number of Edits

Monitors	Screens	Tasks	Mean	Standard Error	95% Confidence Interval	
					Lower Bound	Upper Bound
Two	Single	Slide	10.741	.446	9.856	11.626

Monitors		Spreadsheet	17.148	.395	16.365	17.932
		Text	11.389	.464	10.469	12.309
	Multi-screen	Slide	11.852	.492	10.877	12.827
		Spreadsheet	17.630	.285	17.064	18.195
		Text	13.500	.442	12.623	14.377
	Hydravision	Slide	11.741	.435	10.878	12.603
		Spreadsheet	17.722	.353	17.022	18.422
		Text	14.037	.384	13.277	14.798
Three Monitors	Single	Slide	11.037	.446	10.152	11.922
		Spreadsheet	16.852	.395	16.068	17.635
		Text	11.796	.464	10.876	12.717
	Multi-screen	Slide	11.130	.492	10.154	12.105
		Spreadsheet	18.278	.285	17.712	18.843
		Text	13.852	.442	12.975	14.729
	Hydravision	Slide	12.259	.435	11.397	13.122
		Spreadsheet	17.944	.353	17.244	18.644
		Text	14.352	.384	13.591	15.112

Table 12: Conditions by screen configurations by tasks means and standard errors for Number of Edits.

Table 13

	Task	Single Mean	Multi Mean	Difference	Percent Change
Two Monitors	Slide	10.741	11.852	1.111	10
	Spreadsheet	17.148	17.630	0.482	2
	Text	11.389	13.500	2.111	18
Three Monitors	Slide	11.037	11.130	0.093	1
	Spreadsheet	16.852	18.278	1.426	8
	Text	11.796	13.852	2.056	17

Table 13: Comparison of SS screen Number of Edits means with MS Number of Edits means, difference, percent of change for each monitor condition.

Table 14

	Task	Single Mean	Hydravision	Difference	Percent Change
Two Monitors	Slide	10.741	11.741	1.00	9
	Spreadsheet	17.148	17.722	0.574	3
	Text	11.389	14.037	2.648	23
Three Monitors	Slide	11.037	12.259	1.222	11
	Spreadsheet	16.852	17.944	1.092	6
	Text	11.796	14.352	2.556	21

Table 14: Comparison of SS screen Number of Edits means with HV Number of Edits means, difference, percent of change for each monitor condition.

Table 15

Number of Edits

Configuration	Tasks	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Single	Slide	10.889	.316	10.263	11.515
	Spreadsheet	17.000	.279	16.446	17.554
	Text	11.593	.328	10.942	12.243
Multi-screen	Slide	11.491	.348	10.801	12.180
	Spreadsheet	17.954	.202	17.554	18.353
	Text	13.676	.313	13.056	14.296
Hydravision	Slide	12.000	.308	11.390	12.610
	Spreadsheet	17.833	.250	17.338	18.328
	Text	14.194	.271	13.657	14.732

Table 15: Screen configurations by tasks means, standard errors, and confidence intervals for Number of Edits.

Table 16

Task	Single Mean	Multi-screen	Difference	Percent Change	Significance
Slide	10.889	11.491	0.602	6	No (α = .15)
Spreadsheet	17.00	17.954	0.954	6	Yes
Text	11.593	13.676	2.083	18	Yes

Table 16: Comparison of SS screen Number of Edits means with MS Number of Edits means, difference, percent of change, and significance.

Table 17

Task	Single Mean	Hydravision	Difference	Percent Change	Significance
Slide	10.889	12	1.111	10	Yes
Spreadsheet	17.00	17.833	0.833	5	Yes
Text	11.593	14.194	2.601	22	Yes

Table 17: Comparison of SS screen Number of Edits means with HV Number of Edits means, difference, percent of change, and significance.

Multi-screen configurations show a consistent increase in the number of edits completed over single screen. This advantage is significant in five of the six comparisons. In the lone non-significant condition, the MS mean is .02 below the upper bound limit of the SS confidence interval and is matched with a significant difference in the HV multi-screen condition. It is likely that the multi-screen advantage is being masked by sample conditions.

Proportion of Edits Completed

The Proportion of Edits Completed variable represents the proportion of the task that could be completed within the time limits imposed. This variable is a derivative of the Number of Completed Edits and has to duplicate the results from that variable. The variable does, however, provide a common base for comparing the effectiveness of different screen configurations across tasks where the number of edits can vary by task demand.

Because the Tasks by Screens interaction was significant, screen means need to be analyzed for each task. Table 18 presents the means, standard errors and confidence intervals for each screen configuration and each task.

Table 18

Screens	**Tasks**	**Mean**	**Std. Error**	**95% Confidence Interval**	
				Lower Bound	**Upper Bound**
Single	Slide	.771	.021	.729	.813
	Spreadsheet	.895	.015	.866	.924
	Text	.696	.019	.657	.734
Multi-Screen	Slide	.805	.021	.764	.846
	Spreadsheet	.945	.011	.924	.966
	Text	.821	.019	.784	.858
Hydravision	Slide	.840	.017	.805	.874
	Spreadsheet	.939	.013	.913	.965
	Text	.852	.016	.821	.883

Table 18: Means, standard errors, and confidence intervals for SS, MS, and HV configurations over tasks for the Proportion of Edits Completed.

Inspection of Table 18 shows that all multi-screen conditions result in a greater proportion of the task being completed (a duplicate of edits completed). These differences are significant in every case but in the slide task for MS configuration (α=.10). None of the MS to HV comparisons were significant.

Number of Editing Errors

Editing errors counted the number of incorrectly entered edits. Editing errors was distinguished from missed edits as errors generally require effort to correct. There would be a higher cost savings if the number of errors could be reduced. Table 19 presents the analysis of variance results for this variable.

Table 19

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks by Condition	.476	4, 424	.754
Screens by Condition	.423	2, 212	.655
Tasks by Condition	1.037	2, 212	.161
Screens by Task	.277	4, 424	.893
Screens	3.698	2, 212	.026
Tasks	14.720	2, 212	.000
Conditions	.009	1, 103	.925

Table 19: Analysis of variance results for Number of Errors.

The main effects of screens and tasks were each significant with no other significant findings. Table Eighteen presents the cell means for the data record. Table 20 presents a comparison of SS means with MS configuration for the two monitor conditions. Table 21 presents a comparison of the SS means with HV configuration means for the two monitor conditions. Table 22 presents the screen means for the main effect of screens, and Table 23 presents the task means for the main effect of tasks.

Table 20

Number of Errors

Monitors	Screens	Tasks	Mean	Standard Error	95% Confidence Interval	
					Lower Bound	Upper Bound

Two Monitors	Single	Slide	.241	.084	.074	.407
		Spreadsheet	.593	.156	.283	.902
		Text	.389	.097	.196	.582
	Multi-screen	Slide	.204	.086	.033	.375
		Spreadsheet	.556	.147	.264	.847
		Text	.167	.060	.049	.285
	Hydravision	Slide	.130	.078	-.024	.283
		Spreadsheet	.519	.229	.064	.973
		Text	.111	.051	.010	.212
Three Monitors	Single	Slide	.370	.084	.204	.537
		Spreadsheet	.630	.156	.320	.939
		Text	.296	.097	.104	.489
	Multi-screen	Slide	.241	.086	.070	.412
		Spreadsheet	.370	.147	.079	.662
		Text	.148	.060	.030	.266
	Hydravision	Slide	.352	.078	.198	.506
		Spreadsheet	.389	.134	.124	.654
		Text	.167	.051	.066	.268

Table 20: Conditions by screen configurations by tasks means, standard errors, and confidence intervals for Number of Errors.

Table 21

	Task	**Single Mean**	**Multi Mean**	**Difference**	**Percent Change**
Two Monitors	Slide	.241	.204	0.037	15
	Spreadsheet	.593	.556	0.037	6
	Text	.389	.167	0.222	57
Three Monitors	Slide	.370	.241	0.129	35
	Spreadsheet	.630	.370	0.26	41
	Text	.296	.148	0.148	50

Table 21: Comparison of SS screen Number of Errors means with MS Number of Errors means, difference, percent of change.

Table 22

	Task	**Single Mean**	**Hydravision**	**Difference**	**Percent Change**
Two Monitors	Slide	.241	.130	0.111	46
	Spreadsheet	.593	.519	0.074	12
	Text	.389	.111	0.278	71
Three Monitors	Slide	.370	.352	0.018	5
	Spreadsheet	.630	.389	0.241	38
	Text	.296	.167	0.129	44

Table 22: Comparison of SS screen Number of Errors means with HV Number of Errors means, difference, percent of change.

Table 23

Screens	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	.420	.053	.314	.525
Multi-screen	.281	.049	.184	.378
Hydravision	.278	.039	.200	.356

Table 23: Means, standard errors, and confidence intervals for SS. MS, and HV configurations over all tasks and conditions for Number of Errors.

Table 24

Tasks	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Slide	.256	.034	.188	.324
Spreadsheet	.509	.067	.376	.643
Text	.213	.031	.152	.274

Table 24: Means, standard errors, and confidence intervals for slide, spreadsheet, and text tasks over all screens and conditions for Number of Errors.

The absence of interaction effects means that the main effects of screens and tasks are consistent over all other factors and can be interpreted directly. Taking tasks first, respondent made significantly more errors in the spread sheet task than in the slide or text tasks. Observational data indicate that most of these spreadsheet errors were location errors (wrong column or row). In screens, the single screen configuration had significantly more errors than either of the two multi-screen configurations. The Hydravision configuration scored lower, but not significantly lower than the multi-screen configuration.

The cell mean comparisons show very large percentage changes indicating relatively large reduction in errors for MS and HV conditions. The reader is cautioned that the number of total errors is small, which increases the effect on percent of change. Nonetheless, the differences are large enough to be indicative of the sort of error reduction one might expect from different screen configurations in different kinds of editing tasks.

Number of Missed Edits

Missed edits were those observed by the O/F to be both skipped by the respondent and followed by a completed edit (correct or erroneous). Edits that were not completed by the five minute time limitation were not counted as missed. Missed edits were considered different from errors as the search and correction protocols would be different for each. Table 25 presents the analysis of variance for this variable.

Table 25

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks by Condition	3.310	4, 424	.011
Screens by Condition	.519	2, 212	.596
Tasks by Condition	.663	2, 212	.516
Screens by Task	.860	4, 424	.488
Screens	1.425	2, 212	.243
Tasks	1.778	2, 212	.172
Conditions	1.356	1, 106	.247

Table 25: Analysis of variance results for Number of Missed Edits.

The significant three-factor interaction directs us to interpret the means at the cell level. Table 26 presents the means, standard errors, and confidence intervals for the cell means. Table 27 presents a comparison of SS means with MS means and Table 28 provides a comparison of SS means with HV means.

Table 26

Number of Missed Edits

Monitors	Screens	Tasks	Mean	Standard Error	95% Confidence Interval	
					Lower Bound	Upper Bound
Two Monitors	Single	Slide	.370	.086	.199	.542
		Spreadsheet	.241	.122	-.001	.482
		Text	.500	.110	.282	.718

	Multi-screen	Slide	.333	.087	.160	.506
		Spreadsheet	.444	.134	.178	.711
		Text	.537	.113	.313	.761
	Hydravision	Slide	.389	.079	.232	.545
		Spreadsheet	.278	.110	.060	.496
		Text	.148	.085	-.021	.317
Three Monitors	Single	Slide	.185	.086	.014	.357
		Spreadsheet	.407	.122	.166	.649
		Text	.241	.110	.022	.459
	Multi-screen	Slide	.315	.087	.142	.488
		Spreadsheet	.130	.134	-.137	.396
		Text	.481	.113	.258	.705
	Hydravision	Slide	.148	.079	-.008	.305
		Spreadsheet	.259	.110	.041	.477
		Text	.389	.085	.220	.558

Table 26: Conditions by screen configurations by tasks means, standard errors and confidence intervals for Number of Missed Edits.

Table 27

	Task	**Single Mean**	**Multi Mean**	**Difference**	**Percent Change**	**Significant**
Two Monitors	Slide	.370	.333	0.037	10	No
	Spreadsheet	.241	.444	-0.203	-84	No
	Text	.500	.537	-0.037	-7	No
Three Monitors	Slide	.185	.315	-0.130	-70	No
	Spreadsheet	.407	.130	0.277	68	Yes
	Text	.241	.481	-0.240	-100	Yes

Table 27: Comparison of SS screen Number of Missed Edits means with MS Number of Missed Edits means, difference, percent of change, and significance.

Table 28

	Task	**Single Mean**	**Hydravision**	**Difference**	**Percent Change**	**Significant**
Two Monitors	Slide	.370	.389	-0.019	-5	No
	Spreadsheet	.241	.278	-0.037	-15	No
	Text	.500	.148	0.352	70	Yes
Three Monitors	Slide	.185	.148	0.037	20	No
	Spreadsheet	.407	.259	0.148	36	No
	Text	.241	.389	-0.148	-61	No

Table 28 Comparison of SS screen Number of Missed Edits means with HV Number of Missed Edits means, difference, percent of change, and significance.

Three of the 12 comparisons were significant with two showing a reduction in misses for multi-screens and one showing an increase. These results indicate that the number of misses is not consistently related to screen configurations. An inspection of the data indicates that 80 percent of the respondents had no misses in the completion of their tasks. This large percentage suggests that the phenomenon of misses is more likely an individual skill issue.

Accuracy

Accuracy is a constructed variable based on the number of completed edits minus the number of error and the number of misses. The rationale for this measure is that missed work and incorrect work requires more time and money to correct than simple unfinished work. While the previous analyses of edits and errors indicate an advantage for multi-screen configurations, it is possible that the location of these measures may result in a different outcome. That possibility suggests that should the same advantage appear in Accuracy, it is a confirmation rather than a replication. Table 29 presents the analysis of variance results.

Table 29

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks by Condition	2.065	4, 424	.085
Screens by Condition	.026	2, 212	.974
Tasks by Condition	3.028	2, 212	.697
Screens by Task	3.850	4, 424	.004
Screens	22.610	2, 212	.000
Tasks	357.961	2, 212	.000
Conditions	.410	1, 106	.523

Table 29: Analysis of variance results for Accuracy

The three-factor interaction and the two-factor interactions involving the number monitors were not significant, but the two-factor screens by task interaction was. Table 30 presents the means, standard errors, and confidence intervals for the cell values; Table 31 presents a comparison of SS and MS means; Table 32 presents a comparison of SS and HV means, all for the data record.

Table 30

Accuracy (Number of Completed Edits minus Errors and Misses)

Monitors	Screens	Tasks	Mean	Standard Error	95% Confidence Interval	
					Lower Bound	Upper Bound
Two Monitors	Single	Slide	10.130	.483	9.173	11.087
		Spreadsheet	16.315	.528	15.269	17.361
		Text	10.500	.502	9.506	11.494
	Multi-screen	Slide	11.315	.514	10.296	12.333
		Spreadsheet	16.630	.498	15.643	17.617
		Text	12.796	.497	11.810	13.782
	Hydravision	Slide	11.222	.479	10.199	12.097
		Spreadsheet	16.926	.511	15.913	17.939
		Text	13.778	.430	12.925	14.631
Three Monitors	Single	Slide	10.481	.483	9.525	11.438
		Spreadsheet	15.815	.528	14.769	16.861
		Text	11.259	.502	10.265	12.254
	Multi-screen	Slide	10.574	.514	9.555	11.593
		Spreadsheet	17.815	.498	16.828	18.802
		Text	13.222	.497	12.236	14.208
	Hydravision	Slide	11.704	.479	10.754	12.653
		Spreadsheet	17.333	.511	16.321	18.346
		Text	13.796	.430	12.943	14.650

Table 30: Conditions by screens by tasks means, standard errors and confidence intervals for Accuracy.

Table 31

	Task	Single Mean	Multi Mean	Difference	Percent Change
Two Monitors	Slide	10.130	11.315	1.185	12
	Spreadsheet	16.315	16.630	0.315	2
	Text	10.500	12.796	2.296	22
Three Monitors	Slide	10.481	10.574	0.093	1
	Spreadsheet	15.815	17.815	2.000	13
	Text	11.259	13.222	1.963	17

Table 31: Comparison of SS screen Accuracy means with MS Accuracy means, difference, and percent of change.

Table 32

	Task	Single Mean	Hydravision	Difference	Percent Change
Two Monitors	Slide	10.130	11.222	1.092	11
	Spreadsheet	16.315	16.926	0.611	4
	Text	10.500	13.778	3.278	31
Three Monitors	Slide	10.481	11.704	1.223	12
	Spreadsheet	15.815	17.333	1.518	10
	Text	11.259	13.796	2.537	22

Table 32: Comparison of SS screen Accuracy means with HV Accuracy means, difference, and percent of change.

Because the three-factor interaction was not significant and the two-factor screens by task interaction was, the data are best analyzed by collapsing monitor conditions and looking at the screens means by task. Table 33 presents that information.

Table 33

Screens	Tasks	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Single	Slide	10.306	.341	9.629	10.982
	Spreadsheet	16.065	.373	15.325	16.805
	Text	10.880	.355	10.177	11.583
Multi-screen	Slide	10.944	.363	10.224	11.665
	Spreadsheet	17.222	.352	16.524	17.920
	Text	13.009	.352	12.312	13.707
Hydravision	Slide	11.491	.328	10.841	12.141
	Spreadsheet	17.130	.361	16.413	17.846
	Text	13.787	.304	13.184	14.390

Table 33: Means, standard errors, and confidence intervals for SS, MS, and HV configurations by tasks over Accuracy.

Inspection of Table 33 shows that multi-screen configurations resulted in higher accuracy scores that were significantly higher in all but the SS to MS slide task comparison ($\alpha = .125$). In addition, the HV text scores was significantly higher than the MS text score, although the other two comparisons were not significant and their direction mixed.

Proportion of Accurate Edits

The Proportion of Accurate Edits is fully derivative of Accuracy and, consequently, does not add to the weight of evidence, but as with the Proportion of Completed Edits, it provides a common base by which to compare the effectiveness of different screen configurations across tasks where the means differ because of differences in task demands. The analysis of variance replicated that of Accuracy as expected. Table 34, therefore, presents the means, standard errors, and confidence intervals for SS, MS, and HV configurations by task for Proportion of Accurate Edits.

Table 34

Screens	Tasks	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Single	Slide	.732	.024	.684	.779
	Spreadsheet	.846	.020	.807	.884
	Text	.653	.021	.612	.695
Multi-screen	Slide	.768	.023	.723	.812
	Spreadsheet	.906	.019	.870	.943
	Text	.781	.021	.740	.822
Hydravision	Slide	.805	.020	.765	.844
	Spreadsheet	.902	.019	.864	.939
	Text	.827	.018	.792	.862

Table 34: Means, standard errors, and confidence intervals for SS, MS, and HV configurations by Tasks over Proportion of Accurate Edits..

Again, the results (absolute and significant) for Accuracy are duplicated with multi-screen configurations showing a higher percentage of correct edits for all tasks. Hydravision shows that advantage for slide and text tasks but not for spreadsheet tasks.

Time per Completed Edit

Time per Completed Edit is the editing time divided by the number of completed edits. It represents the flow of work over time and can be used to craft estimates of work completion over jobs of varying length. Table 35 presents the analysis of variance results for Time per Completed Edit.

Table 35

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks by Condition	1.322	4, 424	.261
Screens by Condition	.701	2, 212	.497
Tasks by Condition	.488	2, 212	.615
Screens by Task	5.742	4, 424	.000
Screens	23.452	2, 212	.000
Tasks	282.492	2, 212	.000
Conditions	.006	1, 106	.940

Table 35: Analysis of variance results for Time per Completed Edit.

None of the multi-factor interactions involving Condition were significant. The two-factor Screens by Tasks interaction was significant pointing to a differential effect of screen configurations across tasks. Tables 36 through 38 provide the cell means comparisons that contribute to the data record.

Table 36

Time per Edit

Monitors	Screens	Tasks	Mean	Standard Error	95% Confidence Interval	
					Lower Bound	Upper Bound
Two Monitors	Single	Slide	28.167	1.644	24.908	31.426
		Spreadsheet	15.120	.886	13.363	16.878
		Text	28.739	1.523	25.719	31.759
	Multi-screen	Slide	25.299	1.742	21.846	28.752
		Spreadsheet	12.525	.680	11.177	13.873
		Text	22.479	1.352	19.799	25.158
	Hydravision	Slide	25.223	1.533	22.183	28.262
		Spreadsheet	12.911	.777	11.370	14.453

		Text	21.371	.926	19.535	23.207
Three Monitors	Single	Slide	28.081	1.644	24.823	31.340
		Spreadsheet	16.023	.886	14.265	17.780
		Text	28.360	1.523	25.340	31.380
	Multi-screen	Slide	28.667	1.742	25.214	32.119
		Spreadsheet	11.913	.680	10.564	13.261
		Text	22.425	1.352	19.745	25.105
	Hydravision	Slide	24.223	1.533	21.183	27.262
		Spreadsheet	12.044	.777	10.502	13.585
		Text	20.925	.926	19.089	22.761

Table 36: Conditions by screen configurations by tasks means, standard errors and confidence intervals for Time per Completed Edit.

Table 37

	Task	**Single Mean**	**Multi Mean**	**Difference**	**Percent Change**
Two Monitors	Slide	28.167	25.299	2.868	10
	Spreadsheet	15.120	12.525	2.595	17
	Text	28.739	22.479	6.26	22
Three Monitors	Slide	28.081	28.667	-0.586	-2
	Spreadsheet	16.023	11.913	4.11	26
	Text	28.360	22.425	5.935	21

Table 37: Comparison of SS screen Time per Completed Edit means with MS Time per Completed Edit means, difference, and percent of change.

Table 38

	Task	**Single Mean**	**Hydravision**	**Difference**	**Percent Change**
Two Monitors	Slide	28.167	25.223	2.944	10
	Spreadsheet	15.120	12.911	2.209	15
	Text	28.739	21.371	7.368	26
Three Monitors	Slide	28.081	24.223	3.858	14
	Spreadsheet	16.023	12.044	3.979	25
	Text	28.360	20.925	7.435	26

Table 38: Comparison of SS screen Time per Completed Edit means with HV Time per Completed Edit means, difference, and percent of change.

Table 39 presents the screen configuration means for each task in order to investigate the significant Screens by Task interaction.

Table 39

Screens	Tasks	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Single	Slide	28.124	1.162	25.820	30.428
	Spreadsheet	15.572	.627	14.329	16.814
	Text	28.549	1.077	26.414	30.685
Multi-screen	Slide	26.983	1.232	24.541	29.425
	Spreadsheet	12.219	.481	11.266	13.172
	Text	22.452	.956	20.557	24.347
Hydravision	Slide	24.723	1.084	22.573	26.872
	Spreadsheet	12.478	.550	11.388	13.567
	Text	21.148	.655	19.850	22.446

Table 39: Time per Completed Edit means, standard errors and confidence intervals for each screen configuration by task.

The data in Table 39 show a consistent advantage for multi-screen configurations across all tasks in terms of shorter average time per edit. These differences are significant for all but the SS to MS comparison for the slide task ($\alpha = .37$). There were no significant differences between MS and HV means, although the pattern of HV being more effective in slide and text tasks was repeated. In terms of absolute values, multi-screen configurations (MS and HV combined) result in a savings of 2.2 seconds per slide edit, 3.2 seconds per spreadsheet edit and 6.7 seconds per text edit.

Time per Accurate Edit

Time per Accurate Edit is a ratio of editing time divided by the number of completed edits minus the number of errors and missed (not unfinished) edits. This variable can be considered the time it takes to turn in a perfect performance. Table 40 presents the analysis of variance results.

Table 40

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks by Condition	1.356	4, 424	.248
Screens by Condition	.193	2, 212	.824

Tasks by Condition	.539	2, 212	.584
Screens by Task	4.674	4, 424	.001
Screens	24.132	2, 212	.000
Tasks	172.003	2, 212	.000
Conditions	.041	1, 106	.839

Table 40: Analysis of variance for Time per Accurate Edit.

None of the interactions involving the number of monitors was significant. The two-factor Screens by Task interaction was significant, however. Tables 41-43 present the comparisons of the cell means for the data record. Table 44 presents the indicated comparison of the screen configuration means by task.

Table 41

Time per Accurate Edit

Monitors	Screens	Tasks	Mean	Standard Error	95% Confidence Interval	
					Lower Bound	Upper Bound
Two Monitors	Single	Slide	30.976	2.208	26.599	35.353
		Spreadsheet	16.856	1.232	14.413	19.298
		Text	32.496	1.989	28.551	36.440
	Multi-screen	Slide	27.649	2.104	23.477	31.821
		Spreadsheet	14.744	1.185	12.395	17.093
		Text	23.941	1.582	20.805	27.076
	Hydravision	Slide	27.099	1.790	23.550	30.649
		Spreadsheet	14.546	1.240	12.089	17.004
		Text	21.871	1.045	19.799	23.943
Three Monitors	Single	Slide	31.420	2.208	27.043	35.797
		Spreadsheet	17.681	1.232	15.239	20.123
		Text	30.389	1.989	26.445	34.333
	Multi-screen	Slide	30.516	2.104	26.344	34.688
		Spreadsheet	12.520	1.185	10.171	14.870
		Text	23.974	1.582	20.839	27.110
	Hydravision	Slide	25.668	1.790	22.118	29.217
		Spreadsheet	13.206	1.240	10.748	15.663
		Text	22.028	1.045	19.956	24.100

Table 41: Conditions by screen configurations by tasks means, standard errors and confidence intervals for Time per Accurate Edit.

Table 42

	Task	Single Mean	Multi Mean	Difference	Percent Change
Two Monitors	Slide	30.976	27.649	3.327	11
	Spreadsheet	16.856	14.744	2.112	12
	Text	32.496	23.941	8.555	26
Three Monitors	Slide	31.420	30.516	0.904	3
	Spreadsheet	17.681	12.520	5.161	29
	Text	30.389	23.974	6.415	21

Table 42: Comparison of SS screen Time per Accurate Edit means with MS Time per Accurate Edit means, difference, and percent of change.

Table 43

	Task	Single Mean	Hydravision	Difference	Percent Change
Two Monitors	Slide	30.976	27.099	3.887	11
	Spreadsheet	16.856	14.546	2.31	14
	Text	32.496	21.871	10.625	33
Three Monitors	Slide	31.420	26.197	5.223	17
	Spreadsheet	17.681	13.206	4.475	25
	Text	30.389	22.028	8.361	28

Table 43: Comparison of SS screen Time per Accurate Edit means with HV Time per Accurate Edit means, difference, and percent of change.

Table 44

Screens	Tasks	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Single	Slide	31.198	1.561	28.103	34.293
	Spreadsheet	17.268	.871	15.542	18.995
	Text	31.442	1.407	28.653	34.232
Multi-screen	Slide	29.083	1.488	26.132	32.033
	Spreadsheet	13.632	.838	11.971	15.293
	Text	23.958	1.118	21.740	26.175
Hydravision	Slide	26.384	1.266	23.874	28.893
	Spreadsheet	13.876	.877	12.138	15.614
	Text	21.950	.739	20.485	23.415

Table 44: Time per Accurate Edit means, standard errors and confidence intervals for each screen configuration by task.

The pattern of differences in Table 44 reflects the recurring theme of multi-screen configurations showing superior performance over the single screen configuration. All differences are significant with the exception of the SS to MS comparison in the slide task (α = .163). HV times were significantly smaller than MS times in the text task and approached significance in the slide task (α = .07). The spreadsheet task showed a marginal reversal as has been the pattern. The multi-screen advantage (MS and HV combined) averaged 3.5 seconds in the slide and spreadsheet tasks and 8.5 seconds in the text editing tasks.

Block Variables

Block variables are times and counts summed across the tasks in a given screen configuration. They were devised to give some sense of the differences among screen configurations across a varied workday. The unit variables are the same as the basic task variables: task time, edit time, number of edits, number of errors and number of misses. The reader is reminded that each screen configuration block has each of the nine separate tasks and each of the three orders of those tasks in equal proportion. Any task or order effects are, therefore, equally distributed across screen configurations. Further, the total number of edit events is the same for each configuration so the values are directly comparable. Figure 2 shows the basic Conditions by Screens mixed design used in these analyses.

Figure 2

	Single Screen	Multi-Screen	Hydravision
Two Monitors			
Three Monitors			

Figure 2: Statistical design for all block performance variables.

Block variables are presented in the order of time, number, and the ratio of time over number.

Block Task Time

Block Task Time is the sum of the Task Time values for the three tasks performed under a given screen configuration. It represents both work and transition time to answer the question of whether efficiencies are different over a varied work period. Because times are summed over tasks, the task variable drops out of the analysis, giving a screens by conditions analysis. Screens is the within-subjects variable (SS, MS, HV) and conditions, the number of monitors at the station, is the between-subjects variable. Table 45 presents the analysis of variance

Table 45

Factors	F-Test	Degrees of Freedom	Significance
Screens by Condition	.069	2, 212	.933
Screens	9.643	2, 212	.000
Conditions	.092	1, 106	.763

Table 45: Screens by Conditions analysis of variance for Block Task Time.

The interaction effect was not significant indicating that the main effects were consistent. The main effect of Screens was significant. Table 46 presents the three configuration means, standard errors and confidence intervals.

Table 46

Screens	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	1132.969	15.604	1102.032	1163.906
Multi-Screen	1062.504	17.922	1026.971	1098.037
Hydravision	1063.865	17.067	1030.027	1097.702

Table 46: Means, standard errors, and confidence intervals for screen configurations over Block Task Time.

Inspection of Table 46 shows that Block Task Times were significantly longer for the single screen configuration than for either multi-screen configuration, with a difference of 70 and 69 seconds respectively. There was no significant difference between multi-screen configurations.

Block Edit Time

Block Edit Time is the sum of the editing times for the three tasks in a given screen configuration. It represents the "on-task" time required to complete work during a varied task routine. The analysis is a two-factor mixed design with screens the repeated measure and conditions the between factor. Table 47 presents this analysis.

Table 47

Factors	F-Test	Degrees of Freedom	Significance
Screens by Condition	.013	2, 212	.987
Screens	21.254	2, 212	.000
Conditions	.026	1, 106	.872

Table 47: Screens by Conditions analysis of variance for Block Edit Time.

Again the interaction was not significant and the main effect of screens was indicating a significant difference between the configuration means. Table 48 presents those means, standard errors and confidence intervals.

Table 48

Screens	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	817.917	9.270	799.537	836.296
Multi-Screen	761.565	11.890	737.991	785.139
Hydravision	760.139	11.294	737.748	782.530

Table 48: Means, standard errors, and confidence intervals for screen configurations over Block Edit Time.

As with Block Task Time, editing times for the single screen configuration were significantly longer than either of the two multi-screen configurations with a difference of 56 and 57 seconds respectively. There was no significant difference between multi-screen configurations.

Block Number of Edits

This variable is the sum of all completed edits over the nine tasks within a given screen configuration. It represents completed work. Table 49 presents the analysis of variance results.

Table 49

Factors	F-Test	Degrees of Freedom	Significance
Screens by Condition	0.192	2, 212	.826
Screens	25.541	2, 212	.000
Conditions	0.287	1, 106	.594

Table 49: Screens by Conditions analysis of variance for Block Number of Edits.

The main effect of screens was the only significant effect found in this analysis indicating that there is a linear structure of differences between means across conditions. Table 50 presents the comparison of the configuration means.

Table 50

Screens	Mean	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	39.481	.721	38.052	40.911
Multi-screen	43.120	.670	41.792	44.449
Hydravision	44.028	.604	42.829	45.226

Table 50: Means, standard errors, and confidence intervals for screen configurations over Block Number of Edits.

Single screen blocks showed significantly fewer edits than multi-screen blocks in either MS or HV configuration. The difference between MS and HV screens was not significant but repeated the pattern shown in editing time

Block Number of Errors

Table 51 presents the analysis of variance for the number of errors made across tasks in a given screen configuration.

Table 51

Factors	F-Test	Degrees of Freedom	Significance
Screens by Condition	.423	2, 212	.655
Screens	3.698	2, 212	.026
Conditions	.009	1, 106	.925

Table 51: Screens by Conditions analysis of variance for Block Number of Errors.

The main effects of Screens was significant and all others were not. Table 52 provides the comparison among configuration means.

Table 52

Screens	Mean	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	1.259	.159	.943	1.575
Multi-screen	.843	.146	.552	1.133
Hydravision	.833	.118	.600	1.067

Table 52: Means, standard errors, and confidence intervals for screen configurations over Block Number of Errors.

The pattern of significant advantage for multi-screen configurations is repeated in these results as is the non-significant advantage for Hydravision.

Block Number of Misses

Table 53 presents the analysis of variance for the number of missed edits within an editing performance across tasks for a given screen configuration.

Table 53

Factors	F-Test	Degrees of Freedom	Significance
Screens by Condition	.519	2, 212	.596
Screens	1.425	2, 212	.243
Conditions	1.356	1, 106	.247

Table 53: Screens by Conditions analysis of variance for Block Missed Edits.

The analysis of variance over Block Missed Edits showed no significant effects. However, there appears to be some information to be gained in the inspection of the cell means. Table 54 provides those data.

Table 54

Conditions	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Two Monitors	Single	1.111	.201	.713	1.509
	Multi-screen	1.315	.229	.860	1.770
	Hydravision	.815	.185	.447	1.183
Three Monitors	Single	.833	.201	.435	1.231
	Multi-screen	.926	.229	.471	1.381
	Hydravision	.796	.185	.429	1.164

Table 54: Means for each screen configuration by number of monitors for Block Missed Edits.

The non-significant findings plus the variations in the cell means suggest that the contention that the number of misses is not related to screen configuration is supported here. (Why if it were associated, for example, would the SS mean for the 3 monitor station be that much lower than the SS mean for 2 monitor station? Only one monitor is turned on at each station.) The non-significant pattern of the HV advantage is repeated, however, and this is the strongest of the 2-monitor/3-monitor comparisons that we have seen. This is suggestive that other measures may have value in future research.

Block Accuracy

Block Accuracy is the sum of the Accuracy values (edits minus errors and misses) over tasks within a given screen configuration. Table 55 presents the analysis of variance for this variable.

Table 55

Factors	F-Test	Degrees of Freedom	Significance
Screens by Condition	.026	2, 212	.974
Screens	22.610	2, 212	.000
Conditions	.410	1, 106	.523

Table 55: Screens by Conditions analysis of variance for Block Accuracy.

The main effect of Screens was significant. Table 56 presents the means for the three configurations.

Table 56

Screens	Mean	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	37.250	.807	35.650	38.850
Multi-screen	41.176	.806	39.578	42.774
Hydravision	42.407	.744	40.932	43.883

Table 56: Means for each screen configuration for Block Accuracy.

The single screen configuration showed significantly fewer accurate edits than either of the multi-screen conditions. Repeated again is the non-significant advantage for Hydravision.

Block Time per Edit

Block Time per Edit is constructed by dividing Block Edit Time by Block Number of Edits. Using the summed values to calculate the variable means that the base is equivalent across all configurations. Table 57 presents the analysis of variance for this variable.

Table 57

Factors	F-Test	Degrees of Freedom	Significance
Screens by Condition	.489	2, 212	.614
Screens	23.337	2, 212	.000
Conditions	.018	1, 106	.892

Table 57: Screens by Conditions analysis of variance for Block Time per Edit.

The main effect of Screens was significant. Table 58 presents the comparison of the means

Table 58

Screens	Mean	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	21.956	.670	20.627	23.286
Multi-screen	18.728	.690	17.360	20.096

Hydravision	18.043	.567	16.918	19.167

Table 58: Means for each screen configuration for Block Time per Edit.

The pattern typical of these block variables is repeated here: The single screen configuration has significantly longer times for each edit than either of the multi-screen configurations. Hydravision shows a small and non-significant advantage.

Block Time per Accurate Edit

Block Time per Accurate edit is the quotient of editing time divided by number of accurate edits. Error controls described above were applied to three out of range cases. Table 59 presents the analysis of variance results.

Table 59

Factors	F-Test	Degrees of Freedom	Significance
Screens by Condition	.210	2, 212	.811
Screens	21.150	2, 212	.000
Conditions	.026	1, 106	.872

Table 59: Screens by Conditions analysis of variance for Block Time per Accurate Edit.

The main effect of Screens was significant. Table 60 presents the comparisons of the configuration means.

Table 60

Screens	Mean	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	23.793	.839	22.130	25.455
Multi-screen	20.067	.885	18.312	21.822
Hydravision	18.869	.626	17.628	20.111

Table 60: Means for each screen configuration for Block Time per Accurate Edit.

The single screen configuration shows significantly longer accurate editing times than either multi-screen configuration. HV times are non-significantly lower than MS repeating what has become the standard pattern of these block variables.

Analysis and Results Performance by Expertise

Analysis: Performance by Expertise

Observations of respondent performance by the observer/facilitators clearly suggested that respondents with different prior levels of competence performed differently in the single-to-multi-to-Hydravision screen configurations. It is also clear from the analyses just reported that multi-screen configurations generally result in shorter work times and more productivity across all respondents. The question remains as to how that advantage might be differentiated across expertise. In order to investigate this difference, the ipsative measure of Application Expertise (see definitions above) was crafted into a three level variable that divided the total respondent group into three roughly equal groups. The variables of Block Editing Time and Block Number of Edits were used in these analyses. Block variables are called for to preserve the controls for order. The correlation analysis showed that all time variables correlate highly and all number variables correlated highly (average $r > .80$), but time and number variables function differently from one another. We will learn about the same information regardless of which time or number variable we use, but we will learn the most in comparing them. Table 61 presents the correlations for Number of Edits and Editing Time for each of the screen configurations.

Table 61

		Number of Edits Single	Number of Edits Multi-	Number of Edits Hydra-

		Screen	screen	vision
Editing Time Single Screen	Pearson Correlation	-.642		
	Sig. (2-tailed)	.000		
	N	108		
Editing Time Multi-screen	Pearson Correlation		-.698	
	Sig. (2-tailed)		.000	
	N		108	
Editing Time Hydravision	Pearson Correlation			-.617
	Sig. (2-tailed)			.000
	N			108

Table 61: Correlations between Number of Edits and Editing Time for each screen configuration.

The significant correlations in Table 61 show that time and number are negatively correlated. The more time a respondent takes to complete the task the less the respondent gets done. This is a strong indication of the impact of expertise accounting for some 40 percent of the variance between these two variables for each screen configuration.

Given the justification provided by the correlation analysis, Block Editing Time and Block Number of Edits were analyzed over the combined expertise variable. Figure 3 presents

Figure 3

	Screen Configuration		
Expertise	Single Screen	Multi-Screen	Hydravision
Low	Editing Time Number of Edits		
Moderate		Editing Time Number of Edits	
High			Editing Time Number of Edits

Figure 3: Analysis of variance design for performance variables over screens by expertise.

the mixed design used, where screens is the "within" variable and expertise is the "between" variable.

Results: Performance by Expertise

The expertise variable was a self-rating by respondents on their relative competence in each of the three applications used in this study. To construct the expertise groups the mean of the three expertise scores was taken and certain adjustments made. As is typical in these constructed groups, the middle group represents characteristics of both of the other groups. For this analysis, the point of equivocality were respondents who scored a "1,2,3" for the three application scores. These respondents, by rule, were to be coded as high expertise, but they had reported themselves as also low and moderate. Each of these respondents were sorted by their position on the single screen expertise score. Those above the mean were coded as "high;" those below as "moderate."

Expertise and Block Editing Time

Table 62 presents the analysis of variance for Block Editing time over screens and reported expertise.

Table 62

Factors	F-Test	Degrees of Freedom	Significance
Screens by Expertise	2.878	4, 210	.024
Screens	23.118	2, 210	.000
Expertise	6.117	2, 105	.003

Table 62: Analysis of variance for Block Editing Time over screens and Reported Expertise.

The screens by expertise interaction was significant indicating a differential effect of screen configurations over expertise. Table 63 presents the means, standard errors, and confidence intervals for those cells.

Table 63

Expertise	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
High	Single	798.941	16.377	766.469	831.414
	Multi-screen	696.235	19.574	657.424	735.047
	Hydravision	725.618	19.628	686.699	764.536
Moderate	Single	812.185	18.378	775.746	848.625
	Multi-screen	761.148	21.965	717.595	804.701
	Hydravision	751.407	22.026	707.734	795.081
Low	Single	834.936	13.929	807.317	862.555
	Multi-screen	809.064	16.648	776.053	842.074
	Hydravision	790.128	16.694	757.026	823.229

Table 63: Means, standard errors, and confidence intervals for Block Editing Time over screens and Reported Expertise.

The data in Table 67 show multi-screen configurations resulting in faster editing times over all comparisons, although the difference is not significant in the low expertise SS to HV comparison. High expertise respondents showed the greatest gains in both MS and HV configurations.

Expertise and Block Number of Edits

Table 64 presents the analysis of variance for block Number of Edits over screens and reported expertise.

Table 64

Factors	F-Test	Degrees of Freedom	Significance
Screens by Expertise	.539	4, 210	.707
Screens	22.275	2, 210	.000
Expertise	4.522	2, 105	.013

Table 64: Analysis of variance for Block Number of Edits over screens and Reported Expertise.

The screens by expertise interaction was not significant indicating that the effects of screens was relatively consistent over expertise. The main effects of screens and expertise were significant. Table 65 provides the Block Number of Edits means, standard errors, and confidence intervals for the screens by expertise cells.

Table 65

Expertise	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
High	Single	41.471	1.245	39.002	43.939
	Multi-screen	45.206	1.169	42.887	47.524
	Hydravision	45.500	1.064	43.390	47.610
Moderate	Single	40.889	1.397	38.119	43.659
	Multi-screen	43.185	1.312	40.583	45.787
	Hydravision	44.556	1.194	42.187	46.924
Low	Single	37.234	1.059	35.134	39.334
	Multi-screen	41.574	.995	39.603	43.546
	Hydravision	42.660	.905	40.865	44.454

Table 65: Means, standard errors, and confidence intervals for Block Number of Edits over screens and Reported Expertise.

All levels of expertise were more productive in multi-screen configurations (MS and HV) than in single screen. Low expertise respondents report the highest gains with a nearly 15 percent increase in productivity for the Hydravision condition.

Summary: Performance by Expertise

Table 66 presents the comparisons of single screen to multi-screen and single screen to Hydravision for the two expertise analyses.

Table 66

Analysis	Screens	Level	Single Means	Multi Means	Difference	Percent Change	Sig.
Time over Expertise	Multi-screen	High	798.94	696.24	102.71	12.86	*
		Mod	812.19	761.15	51.04	6.28	*
		Low	834.94	809.06	25.87	3.10	0.10
	Hydra-vision	High	798.94	725.62	73.32	9.18	*
		Mod	812.19	751.41	60.78	7.48	*
		Low	834.94	790.13	44.81	5.37	*
Number over Expertise	Multi-screen	High	41.47	45.21	3.74	9.01	*
		Mod	40.89	43.19	2.30	5.62	0.18

		Low	37.23	41.57	4.34	11.66	*
	Hydra-vision	High	41.47	45.50	4.03	9.72	*
		Mod	40.89	44.56	3.67	8.97	*
		Low	37.23	42.66	5.43	14.57	*

*Significant at .05

Table 66: Reported Expertise comparisons between SS and MS and SS and HV means over Block Edit Time and Block Number of Edits.

Table 66 gives a clear analysis of the differential effect of screens over competence pointed to in the correlation analysis. For respondents who rate themselves high in application expertise, multi-screen configurations increase both speed and productivity but have the strongest effect over speed. High expertise respondents show 4 times the gains of low expertise respondents. For respondents who rate themselves low in application expertise, multi-screen configurations also increase both speed and productivity but have the strongest effect over number of edits performed. In fact, low expertise respondents in multi-screen perform better than high expertise respondents do in single screen. The effect, then, is that low expertise respondents are able to produce nearly 15 percent more during the same time frame of effort. High expertise respondents show a 9 percent gain in productivity and a 13 percent gain in speed.

One way to translate these relative gains is to calculate the time to completion for low and high groups. There are an average of 50 edits (1/3 at 48, 1/3 at 50, 1/3 at 52) in the task blocks. By dividing the average number of edits into the average editing time and multiplying by the total number of edits required, the time to completion can be estimated. For low expertise respondents, average single screen time per edit is 22.43 seconds estimating a time of completion at 18.69 minutes. When they move to multi screens with Hydravision their completion time drops to 15.43 minutes. For high expertise respondents, average single screen time per edit is 19.27 seconds estimating a completion time at 16.05 minutes. When they move to multi-screens, their completion time drops to 12.83 minutes.

Low expertise respondents benefit in productivity with multi-screens to overtake high expertise respondents in single screen. High expertise respondents benefit both in productivity and speed to maintain their advantage over low expertise respondents when they move to multi-screen. The result is that one should expect a 17 percent time savings for low expertise and a 20 percent time savings for high expertise.

Analysis and Results: Usability Data

Analysis

Data from the usability questionnaires that were collected at the end of every task performance (9 questionnaires per respondent) were analyzed two ways: (a) in a tasks by screens repeated measures design that examined differences across tasks and screens for each of effectiveness, comfort, ease of learning, productivity, mistake recovery, task tracking, task focus, and ease of movement across sources; (b) an items by task block screens by level of proficiency mixed design that looked at the eight topics with the responses averaged over the three task types in a task block by a proficiency measure based on task time to completion and number of edits completed.

In order to determine if respondent perceptions of usability differed across screens and task types, a comparison of the three screen configurations and three task types was conducted separately for each of the eight items. Figure 4 presents the design as replicated across each item. This design allows the analysis of the relationship between screen configuration and task on each of the respondent's judgments of usability. Based on our initial suppositions, it was hypothesized that multi-screen configurations would score higher on each item than the single screen. The comparison of multi-screens with and without screen management software was considered exploratory and no hypotheses were developed.

Figure 4

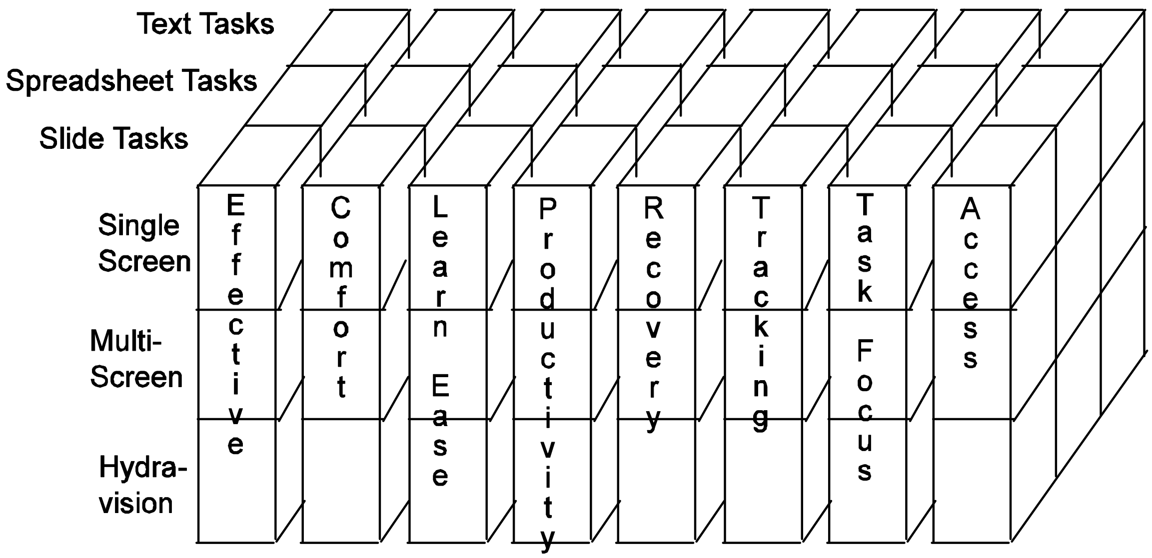

Figure 4: Analysis of Variance design for screens by task type for each item.

Results

Screens by Tasks for each Item

Overview

As hypothesized, multi-screen configurations scored significantly higher in usability than the single screen on every measure in every task. HV means were generally not significantly different from MS means on all measures but varied in direction of difference across tasks. In the item by item report that follows, the tasks by screens by levels analysis of variance results are first presented followed by a table of the cell means. Tables of row (screens), or column (tasks) means are presented as appropriate. Significant results are reported.

Item One: I can effectively complete the tasks using this display configuration.

Table 67 presents the analysis of variance results for tasks by screens for item one.

Table 67

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks	.926	4, 428	.449
Screens	60.228	2, 214	.000
Tasks	39.44	2, 214	.000

Table 67: Analysis of variance results for item one—effectiveness.

The screens by task interaction was not significant indicating that screen differences were consistent over task. To keep the data record complete, Table 68 presents the cell means, standard errors, and confidence intervals for this item. Table 69 presents the data for screens and Table 70 presents the results for tasks.

Table 68

Tasks	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Slide	Single	5.833	.214	5.410	6.257
	Multi-screen	7.926	.181	7.567	8.285
	Hydravision	7.713	.230	7.257	8.169
Spreadsheet	Single	7.102	.223	6.659	7.544
	Multi-screen	8.759	.137	8.488	9.030
	Hydravision	8.648	.154	8.344	8.953
Text	Single	6.315	.228	5.863	6.767
	Multi-screen	8.231	.209	7.817	8.646
	Hydravision	8.231	.166	7.903	8.560

Table 68: Cell means, standard errors, and confidence intervals for item one—effectiveness.

Table 69

Screens	Means	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	6.417	.181	6.058	6.776
Multi-screen	8.306	.146	8.016	8.595
Hydravision	8.198	.155	7.890	8.505

Table 69: Item one means standard errors and confidence intervals for screen configurations over all tasks.

Single screen configurations were considered significantly less effective than either multi-screen configuration. The HV mean was slightly but not significantly lower than the MS mean.

Table 70

Tasks	Mean	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Slide	7.157	.137	6.885	7.429
Spreadsheet	8.170	.121	7.931	8.409
Text	7.593	.144	7.308	7.877

Table 70: Item one means, standard errors, and confidence intervals for tasks over all screens.

Respondents felt significantly less effective in the slide task than any other and significantly more effective in the spreadsheet task than any other.

Item Two: I feel comfortable using this display configuration to complete the tasks.

Table 71 presents the analysis of variance results for tasks by screens for item two.

Table 71

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks	3.583	4, 428	.007
Screens	38.230	2, 214	.000
Tasks	31.921	2, 214	.000

Table 71: Analysis of variance results for item two—comfort in using.

The screens by task interaction was significant indicating that screen differences were not consistent over task. In order to trace the source of this interaction, Table 72 presents the means, standard errors, and confidence intervals for item two.

Table 72

Tasks	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Slide	Single	5.815	.221	5.377	6.253
	Multi-screen	7.935	.180	7.579	8.291
	Hydravision	7.602	.220	7.166	8.037
Spreadsheet	Single	6.926	.240	6.450	7.402
	Multi-screen	8.694	.140	8.418	8.971
	Hydravision	8.509	.173	8.167	8.852
Text	Single	6.880	.227	6.430	7.329
	Multi-screen	8.019	.213	7.597	8.440
	Hydravision	7.907	.192	7.527	8.288

Table 72: Cell means, standard errors, and confidence intervals for item two—comfort in using.

Respondents were significantly less comfortable in single screen than in multi-screen or Hydravision configurations. The slide task was significantly less comfortable in all configurations.

Item Three: It was easy to learn this display configuration.

Table 73 presents the analysis of variance results for tasks by screens for item three.

Table 73

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks	4.480	4, 428	.001
Screens	32.279	2, 214	.000
Tasks	38.103	2, 214	.000

Table 73: Analysis of variance results for item three—ease of learning.

The screens by task interaction was significant indicating that screen differences were not consistent over task.. Table 74 presents the means, standard errors and confidence intervals for the cell means.

Table 74

Tasks	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Slide	Single	6.407	.224	5.962	6.852

	Multi-screen	8.065	.168	7.732	8.398
	Hydravision	7.824	.210	7.407	8.241
Spreadsheet	Single	7.491	.210	7.074	7.908
	Multi-screen	8.787	.137	8.516	9.058
	Hydravision	8.852	.131	8.591	9.112
Text	Single	7.611	.202	7.210	8.012
	Multi-screen	8.296	.180	7.939	8.653
	Hydravision	8.463	.150	8.165	8.761

Table 74: Cell means, standard errors, and confidence intervals for item three—ease of learning.

SS means were significant lower than either MS or HV means in each task. The slide task was the most difficult application site. HV means were higher than MS means in all but the slide task.

Item Four: I believe I became productive quickly using this display configuration

Table 75 presents the analysis of variance results for tasks by screens for item four.

Table 75

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks	.174	4, 428	.952
Screens	54.993	2, 214	.000
Tasks	42.917	2, 214	.000

Table 75: Analysis of variance results for item four—time to productivity.

The screens by task interaction was not significant indicating that screen differences were consistent over task. To keep the data record complete, however, Table 76 presents the cell means, standard errors, and confidence intervals for this item. Table 77 presents the data for screens and Table 78 presents the results for tasks.

Table 76

Tasks	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Slide	Single	5.704	.232	5.243	6.165
	Multi-screen	7.759	.195	7.373	8.146

	Hydravision	7.694	.223	7.252	8.137
Spreadsheet	Single	6.759	.252	6.259	7.259
	Multi-screen	8.833	.135	8.565	9.101
	Hydravision	8.620	.177	8.270	8.971
Text	Single	6.222	.216	5.794	6.650
	Multi-screen	8.148	.199	7.755	8.542
	Hydravision	8.120	.186	7.751	8.490

Table 76: Cell means, standard errors, and confidence intervals for item four—time to productivity.

Table 77

Screens	Means	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	6.228	.192	5.848	6.609
Multi-screen	8.247	.144	7.962	8.532
Hydravision	8.145	.171	7.806	8.484

Table 77: Item four means standard errors and confidence intervals for screen configurations over all tasks.

Single screen configurations scored significantly lower on quickness to productivity than either multi-screen configuration. The HV mean was non-significantly lower than the MS mean.

Table 78

Tasks	Mean	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Slide	7.052	.136	6.784	7.321
Spreadsheet	8.071	.127	7.820	8.322
Text	7.497	.132	7.235	7.759

Table 78: Item four means, standard errors and confidence intervals for tasks over all configurations.

Respondents felt significantly less quick to productivity in the slide task than any other and significantly more effective in the spreadsheet task than any other.

Item Five: Whenever I made a mistake I recovered quickly.

Table 79 presents the analysis of variance results for tasks by screens for item five.

Table 79

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks	2.412	4, 428	.049
Screens	35.084	2, 214	.000
Tasks	28.819	2, 214	.000

Table 79: Analysis of variance results for item five—speed of recovery.

The significant screens by task interaction indicates that screen differences varied by task. Table 80 presents the cell means for this item.

Table 80

Tasks	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Slide	Single	6.241	.213	5.819	6.663
	Multi-screen	7.704	.188	7.332	8.076
	Hydravision	7.537	.228	7.085	7.989
Spreadsheet	Single	7.074	.237	6.603	7.545
	Multi-screen	8.611	.137	8.339	8.883
	Hydravision	8.361	.165	8.035	8.687
Text	Single	7.111	.203	6.709	7.513
	Multi-screen	7.963	.171	7.623	8.303
	Hydravision	8.315	.149	8.019	8.611

Table 80: Cell means, standard errors, and confidence intervals for item five—speed of recovery.

SS means were significant lower than either MS or HV means in each task. HV means were lower than MS means in the slide and spreadsheet task but significantly higher in the text task.. The slide task scored significantly lower than the other two tasks.

Item Six: It was easy to keep track of my tasks.

Table 81 presents the analysis of variance results for tasks by screens for item six.

Table 81

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks	.536	4, 428	.710
Screens	58.582	2, 214	.000

Tasks	25.639	2, 214	.000

Table 81: Analysis of variance results for item six—ease of tracking.

The screens by task interaction was not significant indicating that screen differences were consistent over task. For the record, Table 82 presents the cell means, standard errors, and confidence intervals for this item. Table 83 presents the data for screens and Table 84 presents the results for tasks.

Table 82

Tasks	**Screens**	**Mean**	**Standard Error**	**95% Confidence Interval**	
				Lower Bound	**Upper Bound**
Slide	Single	4.852	.268	4.321	5.383
	Multi-screen	7.435	.258	6.924	7.946
	Hydravision	7.296	.283	6.736	7.857
Spreadsheet	Single	6.065	.290	5.491	6.639
	Multi-screen	8.250	.213	7.828	8.672
	Hydravision	8.269	.220	7.832	8.705
Text	Single	5.009	.283	4.448	5.570
	Multi-screen	7.472	.248	6.981	7.963
	Hydravision	7.583	.254	7.080	8.086

Table 82: Cell means, standard errors, and confidence intervals for item six—ease of tracking.

Table 83

Screens	**Means**	**Standard Error**	**95% Confidence Interval**	
			Lower Bound	**Upper Bound**
Single	5.309	.227	4.859	5.758
Multi-screen	7.719	.200	7.322	8.116
Hydravision	7.716	.219	7.282	8.150

Table 83: Item six means, standard errors, and confidence intervals for screen configurations over tasks.

Single screen configurations were considered significantly less effective than either multi-screen configuration or HV. The mean for HV was not significantly different from the MS mean.

Table 84

Tasks	Mean	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Slide	6.528	.192	6.148	6.908
Spreadsheet	7.528	.170	7.191	7.864
Text	6.688	.175	6.342	7.034

Table 84: Item six means, standard errors and confidence intervals for tasks over configurations.

Respondents felt slightly less effective in the slide task than in the text task but significantly more effective in the spreadsheet task than any other.

Item Seven: It was easy to remember the problem or task.

Table 85 presents the analysis of variance results for tasks by screens for item seven.

Table 85

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks	1.957	4, 428	.100
Screens	57.548	2, 214	.000
Tasks	24.053	2, 214	.000

Table 85: Analysis of variance results for item seven—task memory.

The screens by task interaction was not significant indicating that screen differences were consistent over task. Table 86 presents the cell means, standard errors, and confidence intervals for this item. Table 87 presents the data for screens and Table 88 presents the results for tasks.

Table 86

Tasks	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Slide	Single	6.111	.250	5.615	6.608
	Multi-screen	7.704	.202	7.303	8.105
	Hydravision	8.028	.192	7.648	8.408
Spreadsheet	Single	7.093	.252	6.593	7.593
	Multi-screen	8.676	.150	8.379	8.973
	Hydravision	8.630	.161	8.311	8.949
Text	Single	6.074	.262	5.555	6.593
	Multi-screen	8.259	.189	7.884	8.635

	Hydravision	8.222	.178	7.870	8.574

Table 86: Cell means, standard errors, and confidence intervals for item seven—task memory.

Table 87

Screens	Means	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Single	6.426	.205	6.020	6.832
Multi-screen	8.213	.146	7.924	8.502
Hydravision	8.293	.144	8.007	8.579

Table 87: Item seven means, standard errors, and confidence intervals for screen configurations over tasks.

Single screen configurations were considered significantly more difficult to maintain memory than either multi-screen configuration or HV. The HV was slightly but not significantly higher than the MS mean.

Table 88

Tasks	Mean	Standard Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Slide	7.281	.144	6.996	7.566
Spreadsheet	8.133	.138	7.859	8.406
Text	7.519	.147	7.227	7.810

Table 88: Item seven means, standard errors, and confidence intervals for tasks over configurations.

Respondents were significantly more able to maintain task memory in the spreadsheet task than any other.

Item Eight: It was easy to move from sources of information.

Table 89 presents the analysis of variance results for tasks by screens for item eight.

Table 89

Factors	F-Test	Degrees of Freedom	Significance
Screens by Tasks	.473	4, 428	.756
Screens	87.495	2, 214	.000

Tasks	17.511	2, 214	.000

Table 89: Analysis of variance results for item eight—ease of movement.

The screens by task interaction was not significant indicating that screen differences were consistent over task. Table 90 presents the cell means, standard errors, and confidence intervals for this item. Table 91 presents the data for screens and Table 92 presents the results for tasks.

Table 90

Tasks	**Screens**	**Mean**	**Standard Error**	**95% Confidence Interval**	
				Lower Bound	**Upper Bound**
Slide	Single	5.907	.253	5.405	6.409
	Multi-screen	8.287	.170	7.950	8.624
	Hydravision	8.148	.216	7.719	8.577
Spreadsheet	Single	6.667	.257	6.158	7.176
	Multi-screen	8.935	.143	8.652	9.219
	Hydravision	8.935	.150	8.639	9.232
Text	Single	5.981	.255	5.476	6.487
	Multi-screen	8.620	.181	8.262	8.979
	Hydravision	8.343	.183	7.979	8.706

Table 90: Cell means, standard errors, and confidence intervals for item eight—ease of movement.

Table 91

Screens	**Means**	**Standard Error**	**95% Confidence Interval**	
			Lower Bound	**Upper Bound**
Single	6.185	.208	5.773	6.597
Multi-screen	8.614	.128	8.361	8.868
Hydravision	8.475	.151	8.175	8.775

Table 91: Item eight means, standard errors, and confidence intervals for screen configurations over all tasks.

Single screen configurations were considered significantly more difficult to move across sources than either multi-screen configuration. The MS mean was slightly but not significantly higher than the HV mean.

Table 92

Tasks	**Mean**	**Standard**	**95% Confidence Interval**

		Error	Lower Bound	Upper Bound
Slide	7.448	.139	7.173	7.722
Spreadsheet	8.179	.126	7.929	8.429
Text	7.648	.145	7.360	7.936

Table 92: Item eight means, standard errors, and confidence intervals for tasks over all configurations.

Respondents reported the spreadsheet task as being significantly the easiest to move among sources.

Analysis and Results: Usability by Proficiency

This section examines the relationship between judgments of usability and empirical measures of respondent editing proficiency. It provides a parallel analysis of the performance by expertise investigation. It will describe the ways in which usability judgments are dependent upon actual editing competence.

Analysis: Usability by Proficiency

Task usability block values were created by taking the average of the responses for a given item across the slide, spreadsheet, and text tasks within a screen configuration. A level of proficiency score was developed by dividing both the block task times and block number of edits completed for the single screen into quartiles. The quartile scores were cross-multiplied giving a discontinuous set of scores from 1 to 16. The distribution of these scores divided into three nearly equal groups using the cut points of less than 4 and greater than 9. These three groupings formed the level of proficiency with scores above 9 being the highest.

The purpose of this analysis was to determine if level of proficiency in single screen editing affected the perceived usability of multi-screen configurations. Our expectation was that low skill levels would benefit most from multi-screens and report the greatest difference across the usability measures. We expected higher skill levels to report higher usability scores generally but with multi-screen being higher than single. Figure 5 presents the design. The analysis was based on a items by screens by proficiency design with items and screens being “within subjects” variables and proficiency a “between subjects” variable.

Figure 5

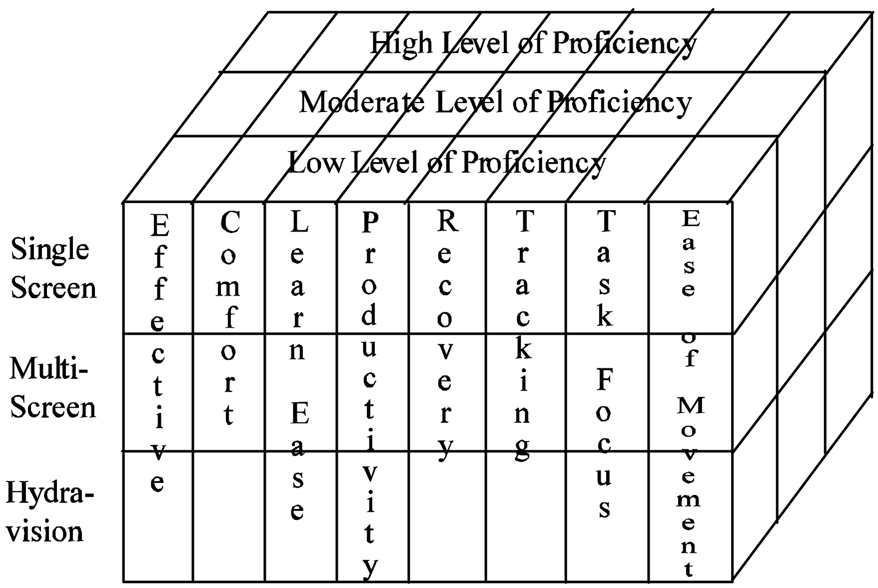

Figure 5: Analysis of Variance design for screens by items by level of proficiency.

Results: Usability by Proficiency

Table 93 presents the analysis of variance results for the comparison of the 8 usability items each averaged over the 3 tasks in a configuration block by configuration screens and performance skill levels.

Table 93

Factors	F-Test	Degrees of Freedom	Significance
Items by Screens by Proficiency	1.107	28, 1470	.320
Items by Proficiency	.583	14, 735	.880

Screens by Proficiency	4.556	4, 210	.002
Items by Screens	12.100	14, 1470	.000
Items	31.669	7, 735	.000
Screens	72.014	2, 210	.000
Proficiency	3.030	2, 105	.053

Table 93: analysis of variance over block items by screens by proficiency.

Two, two-factor interactions—screens by proficiency and items by screens—were significant indicating that the usability scores for screens varied by proficiency and that they also varied across items. Table 94 presents the screen by proficiency cell means, standard errors and confidence intervals, and Table 95 presents the item by screens cell results.

Table 94

Proficiency	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Low	Single	5.451	.275	4.907	5.996
	Multi-screen	8.024	.232	7.565	8.484
	Hydravision	8.225	.262	7.704	8.745
Moderate	Single	6.713	.267	6.182	7.243
	Multi-screen	8.292	.226	7.844	8.739
	Hydravision	7.970	.255	7.464	8.477
High	Single	7.009	.283	6.448	7.569
	Multi-screen	8.359	.239	7.886	8.832
	Hydravision	8.305	.270	7.770	8.840

Table 94: Means, standard errors, and confidence intervals for proficiency by screens.

Single screen configurations scored significantly less usable than either multi-screen configurations across all performance levels. Respondent with low performance scores report the greatest gains in usability when moving to multi-screens, but all showed significant gains.

Table 95

Item	Screens	Mean	Standard Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Effective	Single	6.419	.174	6.074	6.763
	Multi-screen	8.309	.147	8.018	8.600
	Hydravision	8.203	.156	7.894	8.513
Comfort	Single	6.549	.177	6.198	6.900
	Multi-screen	8.216	.149	7.921	8.511

	Hydravision	8.008	.168	7.676	8.340
Learning Ease	Single	7.182	.166	6.852	7.511
	Multi-screen	8.385	.130	8.126	8.643
	Hydravision	8.388	.139	8.112	8.663
Productivity	Single	6.230	.182	5.869	6.591
	Multi-screen	8.247	.143	7.964	8.530
	Hydravision	8.157	.171	7.818	8.495
Mistake Recovery	Single	6.813	.163	6.490	7.136
	Multi-screen	8.096	.134	7.831	8.361
	Hydravision	8.081	.152	7.779	8.382
Task Tracking	Single	5.311	.218	4.878	5.743
	Multi-screen	7.724	.202	7.323	8.124
	Hydravision	7.717	.221	7.279	8.156
Task Focus	Single	6.432	.195	6.046	6.818
	Multi-screen	8.211	.147	7.920	8.502
	Hydravision	8.299	.145	8.012	8.587
Accessibility of Sources	Single	6.193	.203	5.791	6.594
	Multi-screen	8.612	.129	8.357	8.868
	Hydravision	8.481	.153	8.178	8.783

Table 95: Means, standard errors, and confidence intervals for items by screens.

As has been the pattern throughout the usability analysis, single screen configurations scored significantly lower than either multi-screen configuration in each of the 8 usability items. MS and HV means were not significantly different for any item, and those differences were always small, though not consistent in direction.

Differences in items showed the effect of screen configurations. In single screen, task tracking was significantly lower than any other item and ease of learning was significantly higher than any other. In multi-screen, task tracking was also significantly lower than any other item, while accessibility was the highest, significantly higher than all but ease of learning. Hydravision means showed task tracking as significantly lower than all other items and accessibility as highest. Accessibility was significantly higher than mistake recovery, productivity, and comfort as well as task tracking.

This table can also be used to calculate the changes in respondent judgments concerning screen configuration usability by using the single screen score and the average of the two multi-screen scores. In this analysis, multi-screens are seen as 29 percent more effective 24 percent more comfortable, 17 percent easier to learn, 32 percent quicker to productivity, 19 percent easier for mistake recovery, 45 percent easier to track tasks, 28 percent better for task focus and 38 percent easier for moving among sources.

Analysis and Results: Performance and Usability

Analysis

The analysis of the usability items showed that both multi-screen configurations were considered significantly more "usable" than the single screen configuration. The design of the study allows additional information to be gleaned from this relationship through the use of multiple regression analysis. To conduct this analysis, item scores were averaged over the three tasks in a given screen configuration and regressed against the block performance variables of number of edits and editing time. These variables were selected as they were theoretically independent and empirically the least correlated.

Factor analysis of the item scores had shown the questionnaire to be functioning, as designed, as multiple elements of a single concept that we called usability. We could, therefore, have used a single usability score averaged over all items. Open-ended responses suggested, however, that subtle differences were recognized by the respondents among the screen configurations. The regression, consequently, used all 8 items as the independent variables.

The regression analysis used the stepwise method in which independent variables are entered from the most correlated to the least and eliminated if they do not significantly increase the regression coefficient. This method allows us to determine which usability characteristics might be significantly attached to the various screen configurations over the performance variables. In this manner we could discover the “best usability descriptor” for each of the screen configurations. Because of the exploratory nature of this analysis a more “relaxed” decision rule was used with .10 being the criterion for entry and .20 for exclusion.

Results

Table 96 presents the intercorrelation matrices for the single screen, multi-screen and Hydravision item sets. All of the items are significantly and highly intercorrelated, indicating the unity of the questionnaire. This high intercorrelation also suggests that it will take a relative few items to exhaust the relationship between usability and the performance variables. Nonetheless, there are some interesting differences between the matrices. The single screen matrix shows high variability across the correlations and the lowest intercorrelations, indicating usability was a less unified concept under that configuration. The Hydravision matrix is notable in that all of the correlations, save one, are at .80 or above, indicating a strong convergence of usability. The multi-screen matrix has the highest correlations, but with more variability than either the SS or HV matrices, indicating a high agreement with clear differentiation remaining. (The reader is reminded that order of experience with configurations is balanced throughout.)

Table 96

	BSSQ1	BSSQ2	BSSQ3	BSSQ4	BSSQ5	BSSQ6	BSSQ7	BSSQ8
BSSQ1	1.000	.809	.672	.810	.754	.698	.716	.650
BSSQ2	.809	1.000	.829	.835	.759	.725	.733	.682
BSSQ3	.672	.829	1.000	.771	.768	.685	.739	.667
BSSQ4	.810	.835	.771	1.000	.791	.784	.780	.777
BSSQ5	.754	.759	.768	.791	1.000	.681	.773	.599
BSSQ6	.698	.725	.685	.784	.681	1.000	.772	.687
BSSQ7	.716	.733	.739	.780	.773	.772	1.000	.600

BSSQ8	.650	.682	.667	.777	.599	.687	.600	1.000

	BMSQ1	BMSQ2	BMSQ3	BMSQ4	BMSQ5	BMSQ6	BMSQ7	BMSQ8
BMSQ1	1.000	.902	.850	.847	.801	.756	.825	.644
BMSQ2	.902	1.000	.905	.842	.850	.845	.839	.740
BMSQ4	.847	.842	.841	1.000	.835	.749	.804	.689
BMSQ5	.801	.850	.857	.835	1.000	.801	.802	.680
BMSQ6	.756	.845	.812	.749	.801	1.000	.796	.764
BMSQ7	.825	.839	.816	.804	.802	.796	1.000	.669
BMSQ8	.644	.740	.725	.689	.680	.764	.669	1.000

	BHVQ1	BHVQ2	BHVQ3	BHVQ4	BHVQ5	BHVQ6	BHVQ7	BHVQ8
BHVQ1	1.000	.880	.869	.857	.851	.841	.860	.834
BHVQ2	.880	1.000	.896	.909	.826	.871	.827	.811
BHVQ3	.869	.896	1.000	.875	.852	.832	.846	.808
BHVQ4	.857	.909	.875	1.000	.859	.837	.808	.800
BHVQ5	.851	.826	.852	.859	1.000	.811	.810	.763
BHVQ6	.841	.871	.832	.837	.811	1.000	.886	.834
BHVQ7	.860	.827	.846	.808	.810	.886	1.000	.828
BHVQ8	.834	.811	.808	.800	.763	.834	.828	1.000

Table 96: Correlation matrices for single screen, multi-screen and Hydravision block item sets.

Multiple regression analyses were conducted separately for each of the screen configuration by performance variable combinations. In both variables, a significant model could be developed for each configuration. The screen configuration models developed were different from one another and across the performance variables. Table 97 presents the descriptors in the model, the regression coefficient, the coefficient of concordance (R Squared) the F-test, degrees of freedom and significance for each analysis.

Table 97

Performance Variable	**Screens**	**Item**	**R**	**R^2**	**F-Test**	**Deg. of Freedom**	**Sig.**
Block Number Of Edits	Single	Recovered quickly	.361	.130	15.902	1, 106	.000
	Multi-screen	Productive quickly	.513	.263	37.875	1, 106	.000
	Hydra-vision	Easy to learn Effective quickly	.455	.154	13.672	2, 105	.000
Block Editing	Single	Recovered quickly	.310	.096	11.299	1, 106	.000

Time	Multi-screen	Easy to track Productive quickly	.429	.184	11.811	2, 105	.000
	Hydra-vision	Recovered quickly	.484	.235	32.482	1, 106	.000

Table 97: Regression models for performance variables and items over screen configurations.

The analysis indicates that single screen configurations are best described from a usability standpoint according to the ability to recover from mistakes. Multi-screen descriptors included quickness to productivity and the ease of tracking sources. Hydravision descriptors noted the ease of learning, quickness to effectiveness and ease of mistake recovery.

Analysis and Results: Open-Ended Interviews

Open ended responses were coded in two ways: First they were coded as positive, negative or neutral in valence toward the object of judgment be it a screen configuration or a task. Second the comments relating to screen configurations were coded according to themes or criteria of judgment. Table 98 presents the number, percent in category and total percentage of positive, negative and neutral comments for each screen configuration and each task.

Table 98

Comment Category	**Number**	**Percent in Category**	**Percent Total**
Single Screen Positive	28	22.8	2.5
Single Screen Negative	92	74.8	8.3
Single Screen Neutral	2	1.6	.2
Multi-screen Positive	313	79.0	28.2
Multi-screen Negative	52	13.1	4.7
Multi-screen Neutral	23	5.8	2.1
Hydravision Positive	191	68.0	17.2
Hydravision Negative	54	19.2	4.9
Hydravision Neutral	21	7.5	1.9
Text Positive	13	18.1	4.2
Text Negative	2	2.8	.6
Text Neutral	57	79.2	18.5
Spreadsheet Positive	34	28.6	11.0

Spreadsheet Negative	8	6.7	2.6
Spreadsheet Neutral	77	64.7	25.0
Slide Positive	4	3.4	1.3
Slide Negative	32	27.4	10.4
Slide Neutral	81	69.2	26.3
Other	184	100.0	7.03

Table 98: Number, relative and total percentage of positive, negative, and neutral comments for each screen configuration and task.

Multi-screen configurations received the greatest number of positive comments (13 times more than for single screen and twice the number for Hydravision). Multi-screen with Hydravision had 6 times the number of positive statements than single screen. Single screen had the largest number of negative comments. The themes attached to multi-screen configurations included “definitely better,” “easier to see what you were doing,” “easier to work with,” “more intuitive.” Hydravision was described as “organizing,” potentially “easier to work with, but had to learn to use it,” “clean,” “very useful on extensive tasks.” Single screen was characterized as “frustrating,” “hard to see [track],” “everything too small.” Of the tasks, slide editing was considered the most negative and spreadsheet the most positive. These attitudes correspond to performance and usability measures results.

In the second method of analysis, a seven category system was developed based on thematic analysis. Categories include, (a) usability, (b) affect, (c) experience, (d) familiarity, (e) comparing, (f) anticipated use and (g) cognitive framing. Comments were then evaluated on valence (a) positive, (b) negative, (c) neutral or (d) comparative.

Usability comments included statements such as “Multi-screen is easier to use” or “In single screen, it was harder to remember”. Affect comments were those that referred to emotions connected with a screen configuration experience. For example, “Multi-screen was fun” or “Multi-screen was overwhelming.” Experience comments were those that indicated a prior experience with a screen configuration such as “I use multi-screen at work” or “I am use to single screen." Familiarity was experience during the study. "Once you got use to Hydravision" or "I caught on quickly to multi-screen" are examples of familiarity. The Comparing category was used to identify instances when the comment directly compared two screen configurations, such as "Hydravision was better than multi-screen" or "Multi-screen was much easier than single screen." The Anticipated Use category included comments on projected use in the future with a screen configuration such as "I can see how I would use multi-screen at work" or "If I were doing many tasks at once, Hydravision would be helpful." Cognitive Framing included those comments which the participant described how they "saw" the screen configuration. For example, "Hydravision was like having a book open," or "Single screen is like shuffling papers."

Table 99 presents the number of comments in each category. The table shows that Usability accounted for the majority of comments in all three configurations, followed by Affect comments. Experience is more frequent in single and multi-screen comments. Familiarity statements are more frequent in multi-screen and Hydravision. Comparing screen configurations was most often between multi-screen and Hydravision. Anticipated usability is split between multi-screen and Hydravision. Cognitive Framing had the fewest occurrences overall with most referencing Hydravision.

Table 99

Category	**SS**			**MS**			**HV**		
	Count	Row %	Col %	**Count**	Row %	Col %	**Count**	Row %	Col %
Usability	**99**	19.3	80.5	**249**	48.6	63.7	**164**	32.0	59.2
Affect	**11**	10.6	8.9	**63**	60.6	16.1	**30**	28.8	10.8
Experience	**6**	37.5	4.9	**10**	62.5	2.6	**0**	0.0	0.0
Familiarity	**2**	3.3	1.6	**31**	51.7	7.9	**27**	45.0	9.7
Compare	**3**	6.1	2.4	**19**	38.8	4.9	**27**	55.1	9.7
Anticipated Usability	**0**	0.0	0.0	**17**	40.5	4.3	**25**	59.5	9.0
Cognitive Framing	**2**	25.0	1.6	**2**	25.0	.5	**4**	50.0	1.4

Table 99: Category frequencies and percentages for each screen configuration and task.

Comments were evaluated based on the positive, negative, neutral or comparative valence of the statement. Positive comments were “I like multi-screen” or “I can see the possibility of multi-screen.” Negative comments were “I hate this” or “I don’t think I would ever use multi-screen.” Neutral comments did not indicate either a positive or negative valence. For example, “I have always used single screens” or “Hydravision would take some getting used to.” Comparative comments included statements which had either a direct or implied comparison between two screen configurations. For example, “I think multi-screen and Hydravision are about the same” or “Single screens is the least usable out of the screens.” An analysis of the comparative statements follows Table 100.

Table 100 reports the valence by screen configuration. In looking at the single screen comments only, the negative comments were the highest at 74.8 percent. Positive comments were second in frequency of the single screen comments at 22.8 percent. In looking just at multi-screen comments, the majority of comments were positive (78.8%), and negative comments were 13.3 percent. In Hydravision comments, the majority of the comments were positive (67.9%), and negative comments were 19.1%. Neutral comments are the third highest frequency in all screen configurations.

Table 100

Valence		**Single Screen**	**Multi-Screen**	**Hydra-Vision**	**Total**
Positive	**Count**	**28**	**308**	**188**	**524**
	% within Valence	5.3	58.8	35.9	100.0
	% within Screen	22.8	78.8	67.9	66.2
	% of Total	3.5	38.9	23.8	66.2
Negative	**Count**	**92**	**52**	**53**	**197**
	% within Valence	46.7	26.4	26.9	100.0

	% within Screen	74.8	13.3	19.1	24.9
	% of Total	11.6	6.6	6.7	24.9
Neutral	**Count**	**2**	**23**	**21**	**46**
	% within Valence	4.3	50.0	45.7	100.0
	% within Screen	1.6	5.9	7.6	5.8
	% of Total	.3	2.9	2.7	5.8
Comparative	**Count**	**1**	**8**	**15**	**24**
	% within Valence	4.2	33.3	62.5	100.0
	% within Screen	.8	2.0	5.4	3.0
	% of Total	.1	1.0	1.9	3.0
Total	**Count**	**123**	**391**	**277**	**791**
	% within Valence	15.5	49.4	35.0	100.0
	% within Screen	100.0	100.0	100.0	100.0
	% of Total	15.5	49.4	35.0	100.0

Table 100: Valence by Screen frequency and percentages.

Statements that involved a comparison between screen configurations were examined by positive, negative and neutral valence. Each comparisons was listed by which configuration was the subject of the statement. For example, if a participant stated "Multi-screen is the same as Hydravision," the statement would first be coded as multi-screen and then comparison and then neutral.

Using this coding method, the single screen configuration received only three comparisons with multi-screen and Hydravision. Two were negative and one was neutral. In multi-screen compared to single screen, the majority were positive toward multi-screen and none were negative toward multi-screen. In comparing multi-screen and Hydravision, neutral valence was the most frequent. In Hydravision comparisons to single screen, all comments were positive toward Hydravision. In comparison between Hydravision and multi-screen, the majority of the comments presented the two as equal.

Finally, Tables 101 to 103 report the category by valence results for single screen, multi-screen, and Hydravision configurations respectively. In single screen, the majority of the comments (99) come in the usability category (78.8 % of all comments) with most having negative valence (84.8% of Usability). Affect is a distant second with 11 comments, 10 of which are negative. In both multi-screen and Hydravision, Usability is again the highest (209 and 124 respectively), but the valence is reversed. Positive comments account for 83.9 percent of the Usability statements in multi-screen and 75.6 percent in Hydravision. Affect was again second most frequent statement for both multi-screen and Hydravision with the greater majority in each being positive.

Table 101

Category	**Single Screen Valence**					
		Positive	Negative	Neutral	Compar-ative	Total
Usability	**Count**	**19**	**78**	**2**	**0**	**99**
	% within Category	19.2	78.8	2.0	0.0	100.0
	% within Valence	67.9	84.8	100.0	0.0	80.5
	% of Total	15.4	63.4	1.6	0.0	80.5
Affect	**Count**	**1**	**10**	**0**	**0**	**11**
	% within Category	9.1	90.9	0.0	0.0	100.0
	% within Valence	3.6	10.9	0.0	0.0	8.9
	% of Total	.8	8.1	0.0	0.0	8.9
Experience	**Count**	**6**	**0**	**0**	**0**	**6**
	% within Category	100.0	0.0	0.0	0.0	100.0
	% within Valence	21.4	0.0	0.0	0.0	4.9
	% of Total	4.9	0.0	0.0	0.0	4.9
Familiarity	**Count**	**2**	**0**	**0**	**0**	**2**
	% within Category	100.0	0.0	0.0	0.0	100.0
	% within Valence	7.1	0.0	0.0	0.0	1.6
	% of Total	1.6	0.0	0.0	0.0	1.6
Compare	**Count**	**0**	**2**	**0**	**1**	**3**
	% within Category	0.0	66.7	.0.0	33.3	100.0
	% within Valence	0.0	2.2	0.0	100.0	2.4
	% of Total	0.0	1.6	0.0	.8	2.4
Anticipated Usability	**Count**	**0**	**0**	**0**	**0**	**0**
	% within Category	0.0	0.0	0.0	0.0	0.0
	% within Valence	0.0	0.0	0.0	0.0	0.0
	% of Total	0.0	0.0	0.0	0.0	0.0
Cognitive Processing	**Count**	**0**	**2**	**0**	**0**	**2**
	% within Category	0.0	100.0	0.0	0.0	100.0

	% within Valence	0.0	2.2	0.0	0.0	1.6
	% of Total	0.0	1.6	0.0	0.0	1.6
Total	**Count**	**28**	**92**	**2**	**1**	**123**
	% within Category	22.8	74.8	1.6	.8	100.0
	% within Valence	100.0	100.0	100.0	100.0	100.0
	% of Total	22.8	74.8	1.6	.8	100.0

Table 101: single screen category by valence frequencies and percentages.

Table 102

Category	**Multi Screen Valence**					
		Positive	Negative	Neutral	Compar-ative	Total
Usability	**Count**	**209**	**34**	**6**	**0**	**249**
	% within Category	83.9	13.7	2.4	0.0	100.0
	% within Valence	67.9	65.4	26.1	0.0	63.7
	% of Total	53.5	8.7	1.5	0.0	63.7
Affect	**Count**	**52**	**11**	**0**	**0**	**63**
	% within Category	82.5	17.5	0.0	0.0	100.0
	% within Valence	16.9	21.2	0.0	0.0	16.1
	% of Total	13.3	2.8	0.0	0.0	16.1
Experience	**Count**	**2**	**3**	**5**	**0**	**10**
	% within Category	20.0	30.0	50.0	0.0	100.0
	% within Valence	.6	5.8	21.7	0.0	2.6
	% of Total	.5	.8	1.3	0.0	2.6
Familiarity	**Count**	**20**	**1**	**10**	**0**	**31**
	% within Category	64.5	3.2	32.3	0.0	100.0
	% within Valence	6.5	1.9	43.5	0.0	7.9
	% of Total	5.1	.3	2.6	0.0	7.9
Compare	**Count**	**11**	**0**	**0**	**8**	**19**
	% within Category	57.9	0.0	0.0	42.1	100.0
	% within Valence	3.6	0.0	0.0	100.0	4.9
	% of Total	2.8	0.0	0.0	2.0	4.9
Anticipated Usability	**Count**	**13**	**3**	**1**	**0**	**17**
	% within Category	76.5	17.6	5.9	0.0	100.0
	% within Valence	4.2	5.8	4.3	0.0	4.3
	% of Total	3.3	.8	.3	0.0	4.3%
Cognitive Processing	**Count**	**13**	**3**	**1**	**0**	**17**
	% within Category	76.5	17.6	5.9	0.0	100.0
	% within Valence	4.2	5.8	4.3	0.0	4.3
	% of Total	3.3	.8	.3	0.0	4.3
Total	**Count**	**1**	**0**	**1**	**0**	**2**

	% within Category	50.0	0.0	50.0	0.0	100.0
	% within Valence	.3	0.0	4.3	0.0	.5
	% of Total	.3	0.0	.3	0.0	.5

Table 102: Multi-screen category by valence frequencies and percentages.

Table 103

Category	**Hydravision Valence**					
		Positive	Negative	Neutral	Compar-ative	Total
Usability	**Count**	**124**	**36**	**4**	**0**	**164**
	% within Category	75.6	22.0	2.4	0.0	100.0
	% within Valence	66.0	67.9	19.0	0.0	59.2
	% of Total	44.8	13.0	1.4	0.0	59.2
Affect	**Count**	**24**	**5**	**1**	**0**	**30**
	% within Category	80.0	16.7	3.3	0.0	100.0
	% within Valence	12.8	9.4	4.8	0.0	10.8
	% of Total	8.7	1.8	.4	0.0	10.8
Experience	**Count**	**0**	**0**	**0**	**0**	**0**
	% within Category	0.0	0.0	0.0	0.0	0.0
	% within Valence	0.0	0.0	0.0	0.0	0.0
	% of Total	0.0	0.0	0.0	0.0	0.0
Familiarity	**Count**	**2**	**9**	**16**	**0**	**27**
	% within Category	7.4	33.3	59.3	0.0	100.0
	% within Valence	1.1	17.0	76.2	0.0	9.7
	% of Total	.7	3.2	5.8	0.0	9.7
Compare	**Count**	**10**	**2**	**0**	**15**	**27**
	% within Category	37.0	7.4	0.0	55.6	100.0
	% within Valence	5.3	3.8	0.0	100.0	9.7
	% of Total	3.6	.7	0.0	5.4	9.7
Anticipated Usability	**Count**	**24**	**1**	**0**	**0**	**25**
	% within Category	96.0	4.0	0.0	0.0	100.0
	% within Valence	12.8	1.9	0.0	0.0	9.0
	% of Total	8.7	.4	0.0	0.0	9.0
Cognitive Processing	**Count**	**4**	**0**	**0**	**0**	**4**
	% within Category	100.0	0.0	0.0	0.0	100.0
	% within Valence	2.1	0.0	0.0	0.0	1.4
	% of Total	1.4	0.0	0.0	0.0	1.4
Total	**Count**	**188**	**53**	**21**	**15**	**277**
	% within Category	67.9	19.1	7.6	5.4	100.0
	% within Valence	100.0	100.0	100.0	100.0	100.0
	% of Total	67.9	19.1	7.6	5.4	100.0

Table 103: Hydravision category by valence frequencies and percentages.

Discussion: Performance

This section first considers the central question of the effectiveness of multiple screens, briefly looks as the differences among tasks, the differences between conditions, then considers the interaction between screens and tasks, and finally examines the circumstances under which particular screen configurations should be adopted.

Screens

The effect of screen configurations can be most clearly seen by reviewing the results from the block variables. In every analysis, except for the non-significant misses, multi-screen configurations scored significantly better than the single screens. Respondents were able to get on task quicker, do the work faster, and get more of the work done with fewer errors in multi-screen configurations than with a single screen. The gains are solid: 6 percent quicker to task, 7 percent faster on task, 10 percent more production, 16 percent faster in production, 33 percent fewer errors, 18 percent faster in errorless production. Equally impressive is that these gains are achieved by turning on a monitor and five minutes of training.

The value added by the screen management tool, Hydravision is subtle. It did not reach significance, but it was consistent and showed its greatest strength in controlling errors and containing missed edits. Very little of the features of this software were used in this study, because of the nature of the tasks and measurements involved. Respondents used only the multiple desktop feature consistently as they were required by the protocol. They were instructed on other features but generally used procedures with which they were experienced.

Observational notes and responses from the interviews point to Hydravision as an organizing device that allowed respondents to better track the requirements of the task.

Tasks

Without question, the most difficult task for most respondents was the slide editing task. Respondents reported the least experience with the application (mean of 1.25 on a scale of 0-3). Further observations showed respondents with little experience in dealing with graphics editing. A common failing was the inability to recognize differences among slides that had common background but substantial content differences. Interview responses indicated that respondents had little or no experience with editing graphics and were frustrated by the awkwardness of the application's editing protocols. When the slide task was removed from analysis the efficiencies of time to task and time through task rose to 9 and 10 percent respectively.

The spreadsheet task for its part showed the shortest times to completion and the greatest proportion of edits completed but also the greatest number of errors. Respondents reported slightly more experience with the spreadsheet application (m=1.40) than with the slide application. Observational notes show that respondents benefited from the spreadsheet application's ease of editing. When errors were made, they were generally entries in the wrong cell. Most commonly an entire row of entries were shifted up or down, accounting for the relatively large number of total errors in the individual edits.

The text task showed the fewest errors but also the lowest proportion of completed edits. Respondents indicated substantial experience with the application (m=2.22), but few had experience with editing across screens (most work from paper corrections to a screen). Observations indicate that the visual task of locating place from one screen to another was the key difficulty.

Conditions

Conditions represent whether a respondent completed the study using two monitors or three monitors. Interestingly, the study was designed to "naturally" fit a three monitor display, but the three monitor condition consistently showed no advantage over the two monitor condition. Anecdotally, multi-screen users consider the three-monitor display to be optimum, but it did not show here. Observations and comments from interviews suggest that the size of the monitor interacts with the optimal number of screens. Drawing on the comments of one respondent, a highly experienced graphics editor, the 18-inch monitors were too large for a three screen display as one could not keep the entire display within the field of vision. It may be very useful to advance this study with one that uses a three 15-inch monitor configuration.

Screens by Task

The lesson learned in the screen by task interaction is that there appears to be an optimum level of experience with a task that maximizes the immediate effect of the adoption of multi-screens. Too little as in the slide task, and the inexperience is an overburden on the multi-screen effect. Too much, as in the text task, and the productive methods of single screen editing prove a worthier competitor to reduce the size of the effect. Both of these conditions are functions of testing protocol. Respondents given the regular experience of editing slide presentations would eliminate many of their difficulties, and respondents given the regular experience in multi-screen editing would return the competition to a level field.

The greatest proportion of our respondents (95 %) work only in single screen whether at home, at school, or in the office. As multi-screens were more effective than single screen across all tasks on measures of both time and production, it is clear that there is little learning curve in the adoption of multi-screen configurations. The short run benefits should be immediate and the long term gains substantial.

Performance Considerations for Adoption of Multi-screens

This study was designed using simulated office tasks that involved the application of multiple sources of information to a final product. It was, therefore, specifically designed to be responsive to the characteristics of multi-screen displays. The evidence it generated and the recommendations provided here presume similar circumstances—work that involves the integration of multiple sources. In those circumstances, the evidence speaks clearly and convincingly that multi-screen configurations are preferable and make good economic sense.

But not all work involves multiple sources of information. The question can be raised as to what proportion needs to be multi-sourced to justify the expense of adding that additional display port and monitor. The simplest way to answer that question is to extrapolate from the time per edit measure. The evidence suggests a 16 percent savings in time for the same level of production. Over a year's time, one would save $3,840 in labor costs at a $12 per hour clerical wage. Cost for upgrading computers vary by platform, region, and industry. Done at the authors' location with PCs, the upgrade would cost approximately $800 (adding a $75 PCI display card, a $600 LCD monitor, and the installation labor). The break even point is approximately 21 percent. If more than 21 percent of the work involves the use of multiple sources of information, upgrading to multiple screens is cost effective. The reader is also reminded that the break even point will be lower (less than 20%) with a less experienced (rather than diversely experienced) work force and even lower (less than 17 %) with a highly experienced work force.

Discussion: Usability

Usability results showed the consistency of a mantra: Multi-screens either with or without management software are reported as significantly more usable than single screens on measures of effectiveness, comfort, learning ease, time to productivity, mistake recovery, task tracking, task focus, and ease of source movement. Slide tasks were considered the most difficult; spreadsheet tasks the easiest. Further, the least proficient of the respondents moved immediately to the level of the most proficient in their evaluations. They were not intimidated by the introduction of multi-screen displays. Respondents characterized multi-screens according to their quickness to effectiveness and productivity, ease of task tracking, ease of learning, and ease of recovery from mistakes. The open-ended interview data confirmed the positive response to multi-monitor displays. Those data showed overwhelmingly more positive comments for multi-screen and for Hydravision than for single screen and indicated that both multi-screen and Hydravision would be more useable and more likely associated with positive affect. Unlike many technological improvements, the adoption of multi-screen configurations should be a positive experience for the workforce and highly preferred over single screen arrangements. It not only increases productivity; it makes the work judged as easier to do.

Summary and Conclusions

This study compared single screen computer display configurations with multi-screen displays without screen management software and with multi-screen displays with screen management software—ATI's Hydravision. The comparisons were made using three types of ordinary office editing tasks in slide, spreadsheet and text applications. Each task type was a multi-screen task using 7 different windows of information over the course of the task. Each task type had three different forms giving three slide tasks, three spreadsheet tasks, and three text tasks. Each respondent completed a different set of slide, spreadsheet and text tasks in each of the three configurations, single screen (SS), multi-screen (MS) and multi-screen supported by Hydravision (HV).

Performance and usability measures were collected. Performance measures included task time (total time from opening to closing of files), editing time (from first to last edit), number of edits completed, number of errors committed, and number of edits missed. Usability measures included items on screen configuration effectiveness, user comfort, ease of learning, time to productivity, ease of recovery from mistakes, ease of task tracking, ease of task focus, and ease of movement among information sources. In addition, respondents provided information on their computer knowledge, application expertise, time spent on each application, experience with multi-screens, and current job situation.

One hundred eight volunteer respondents participated in the study. They were drawn from the undergraduate, graduate, faculty, and staff of a large intermountain university with non-university personnel as well. The respondents were divided into two groups: One group had multi-screen configurations using two 18 inch LCD NEC Mitsubishi monitors; the other had its multi-screen configurations composed of three 18 inch LCD NEC Mitsubishi monitors.

Multi-screens fared significantly better than single screen on every performance measure except the non-significant misses (80 per cent of the respondents had no missed edits). Respondents got on task quicker, did the work faster, and got more of the work done with fewer errors in multi-screen configurations than with a single screen. They were 6 percent quicker to task, 7 percent faster on task, generated 10 percent more production, were 16 percent faster in production, had 33 percent fewer errors, and were 18 percent faster in errorless production.

These gains are achieved by turning on a monitor and five minutes of training. Nonetheless, some care must be taken in extrapolating these gains over three 5-minute tasks to time saved and production increases achieved over a 40 hour work week. Such gains depend on the nature of the work and the amount of time spent on task and on multi-screen tasks. There is a utility in replicating this study using the tasks integrated into a continuous work period rather than as separate episodes as done here.

Respondents considered multi-screen configurations significantly more useful than single screen on every usability measure. Multi-screens were seen as 29 percent more effective for tasks, 24 percent more comfortable to use in tasks, 17 percent easier to learn, 32 per cent faster to productive work, 19 percent easier for recovery from mistakes, 45 percent easier for task tracking, 28 percent easier in task focus, and 38 percent easier to move around sources of information. These increases were immediate post-test gains. As always long-term gains may be different.

There were no significant differences between two-monitor and three-monitor multi-screen configurations. There was some evidence, however, of a relationship between optimum monitor size and the optimum number of monitors. The gains provided by an additional monitor in a three monitor array can apparently be offset if the monitors are too large, forcing the user to physically track across the screens with head movement. It is recommended that a three-monitor array with 15 to 17-inch monitors be tested. Testing should also investigate portrait and landscape orientations within the array.

The respondent pool was divided into high, moderate, and low competence groups for both performance and usability measures. Low competence respondents were generally significantly lower in single screen performance and usability than high, but when moved to multi-screen arrays they achieved nearly the same levels of performance and usability as high competence respondents did in single screen. When high competence respondents moved to multi-screen, they restored the single screen difference between themselves and low competence respondents on performance but not usability. Low competence respondents increased the number of edits they were able to complete but did not markedly reduce the amount of time it took them to do the task. High competence respondents both increased the number of edits and reduced the amount of time it took to conduct them. Both groups were very positive in their usability judgments for multi-screens.

Given the overwhelming consistency of both the performance and usability measures, multiple monitor configurations are recommended for use in any situation where multiple screens of information are an ordinary part of the work. There will be measurable gains in productivity and the work will be judged as easier to do. In addition, because the gains are strong, multiple monitors are also recommended as cost effective where multi-screen tasks represents as little as 15 percent of the work for the highly competent, 17 percent for entry level competence and 21 percent for the general work force.

The contemporary status of computer displays is poised on a moment of convergence as operating systems can now handle multiple monitors with some routine, display boards with multiple ports are readily available and inexpensive, and LCD monitors with reduced foot print, space, and energy requirements, as well as cost are now becoming the standard. This study demonstrates that multiple monitor arrays should also be a standard of the workplace.