THESIS

An Empirical Study on the Effects of a Collaboration-Aware Computer System and Several Communication Media Alternatives on Product Quality and Time to Complete in a Co-authoring Environment

by

Charles Arthur Green

Thesis Submitted to the Faculty of the Virginia Polytechnic Institute and State University in Partial Fulfillment of the Requirements for the Degree of Masters of Science in

Industrial and Systems Engineering

R. C. Williges, Chairperson

B. H. Williges

R. D. Dryden

Blacksburg, Virginia

August, 1992
ABSTRACT

A new type of software, termed a “group editor”, allows multiple users to create and simultaneously edit a single document; this software has ostensibly been developed to increase efficiency in co-authoring environments where users may not be co-located. However, questions as to the effectiveness of this type of communication aid, which is a member of the “groupware” family of tools used for some types of computer supported cooperative work, remain. Particularly, there has been very little objective data on any group editor because of the problems inherent in evaluating writing, as well as due to the few examples of group editors that exist.

A method was developed to examine the effect of using a particular group editor, Aspects™ from Group Technologies in Arlington, Va., in conjunction with several communication media, on a simple dyad writing task. Six dyads of college students familiar with journalistic writing were matched on attributes of dominance and writing ability and were asked to write short news articles based on short video clips in a balanced two factor within-subject analysis of variance design. Six conditions were tested based on communication media: audio only, audio plus video, and face-to-face; each of these with and without the availability of the group editor. Constraints inherent in the task attempted to enforce consistent document quality levels, measured by grammatical quality and content quality (correctness of information and chronological sequencing). Time to complete the articles was used as a measure of efficiency, independent from quality due to the consistent quality levels of the resulting work.

Results from the time data indicated a significant effect of communication media, with the face-to-face conditions taking significantly less time to complete than either of the other media alternatives. Grammatical quality of the written articles was found to be of consistent high quality by way of computerized grammar checker. Content quality of the documents did not significantly differ for any of the conditions. A supplemental
Latin square analysis showed additional significant differences in time to complete for trial means (a practice effect) and team differences. Further, significantly less variance was found in certain conditions which had the group editor than in other conditions which did not. Subjective data obtained from questionnaires supported these results and additionally showed that subjects significantly preferred trials with the group editor and considered them more productive.

The face-to-face conditions may have been more efficient due to the nature of the task or due to increased communication structure within dyads due to practice with the group editor. The significant effect of Team Differences may have been due to consistent style differences between dyads that affected efficiency levels. The decreased variability in time to complete in certain group editor conditions may have been due to increased communication structure in these conditions, or perhaps due to leveling effects of group writing as opposed to individual writing with team member aid.

These hypotheses need to be tested with further study, and generalizability of the experimental task conditions and results from this particular group editor need to be established as well. Face-to-face conditions clearly resulted in the most efficient performance on this task. The results obtained concerning the group editor suggest possible efficiency or consistency benefits from the use of group editors by co-authoring persons when face-to-face communication is not practical. Perhaps group editors will become a useful method for surrogate travel for persons with disabilities.
ACKNOWLEDGEMENTS

I would like to thank my major advisor, Dr. Robert Williges, and the other members of my committee, Beverly Williges and Dr. Robert Dryden, for giving me the professional and personal support necessary for the completion of this work, as well as for helping me to survive two years of graduate school. I would also like to thank my parents, Dr David Green and Mrs. Theodora Green, for providing the means for me to complete my degree and the encouragement to get me through it. I would like to thank Dr. Walter Wierwille for sitting in my defense at the last moment. Lastly, I would like to thank my friends here at the university for their encouragement throughout.
# Table of Contents

Abstract ....................................................................................... ii
Acknowledgements ....................................................................... iv
Table of Contents ......................................................................... v
List of Figures .............................................................................. viii
List of Tables ............................................................................... ix
Chapter 1 — Introduction .......................................................... 1
  Purpose ...................................................................................... 3
  Literature Review ...................................................................... 5
    Group Interactions ................................................................. 7
    CSCW Studies ....................................................................... 9
    Communications Studies ....................................................... 17
    Writing studies ..................................................................... 20
Chapter 2 — Method ................................................................. 22
  Task ....................................................................................... 22
  Questions ............................................................................... 23
  Task Pretest .......................................................................... 25
Treatment Conditions ............................................................... 26
Subjects ..................................................................................... 28
Procedure ................................................................................... 28
  Pretests .................................................................................. 28
  Training ............................................................................... 31
  Practice ............................................................................... 31
  Sequence .............................................................................. 32
  Exit Questionnaire ............................................................... 33
Apparatus .................................................................................... 34
Video Viewing Area .............................................. 34
2 Offices ................................................................ 34
Monitoring .............................................................. 35
Statistical Analyses .................................................. 36
Chapter 3 — Results ................................................ 39
Dependent Measures ............................................... 39
Analyses of Time Data ............................................ 40
Analysis of Questionnaire Data ............................... 45
Chapter 4 — Discussion ............................................ 53
Conclusions ............................................................. 57
References .............................................................. 59
Appendices .............................................................. 65
Appendix I. Task Pretest Fact Sheets .......................... 65
Fact Sheet 1 ............................................................. 65
Fact Sheet 2 ............................................................. 66
Fact Sheet 3 ............................................................. 67
Appendix II. Questions For the Video Clips .................. 68
Practice Question Set (1) .......................................... 68
Question Set 1 .......................................................... 69
Question Set 2 .......................................................... 70
Question Set 3 .......................................................... 71
Question Set 4 .......................................................... 72
Question Set 5 .......................................................... 73
Question Set 6 .......................................................... 74
Appendix III. Participant's Informed Consent ............... 75
Appendix IV. Instructions ........................................... 77
Training ................................................................. 77
Task Pretest ............................................................ 79
Trials .................................................................. 79
Appendix V. IRB Protocol .............................................. 82
Appendix VI. Exit Questionnaire ................................. 85
Appendix VII. Additional Charts and Analyses .............. 88
Vita .................................................................... 94
List of Figures

Figure 1. Procedure for the Study. ................................................. 34
Figure 2. Experimental Apparatus. ............................................... 35
Figure 3. Experimental Design. ................................................... 37
Figure 4. Interaction Graph for Task Time. ................................. 43
Figure 5. Graph of Trial Means. ................................................... 45
Figure 6. Aspects Preferences Chart. .......................................... 46
Figure 7. Essential to Work Feature Opinions Chart. .................... 47
Figure 8. Easy to Use Feature Opinions Chart. ............................. 48
Figure 9. Detrimental Feature Opinions Chart. ............................. 48
Figure 10. Aspects™ General Questions Chart. ............................. 50
List of Tables

Table 1. CSCW Studies. ................................................................. 17
Table 2. Task Pretest Results. ........................................................ 26
Table 3. Pretest Scores and Matching of Subjects to Dyads. ............ 30
Table 4. Trial Presentation Order. .................................................. 38
Table 5. Intercorrelation Matrix. .................................................... 41
Table 6. Main Analysis of Time Data. ............................................. 42
Table 7. LSD tests for Communication Media Effect. ....................... 42
Table 8. Comparisons of Variance for Treatment Conditions. ............ 43
Table 9. Latin Square ANOVA. ....................................................... 44
Table 10. Aspects Opinions ANOVA. .............................................. 46
Table 11. Essential to Work Feature Opinions ANOVA. ...................... 49
Table 12. Essential to Work Feature Opinion LSD Tests. ................... 49
Table 13. Kolmogorov-Smirnov Analysis on Several Questions. ....... 51
CHAPTER 1 — INTRODUCTION

Collaboration, communication, and automation have recently become recognized as some of the key facets in successful work in our information-oriented society. Automation of much information work has already been accomplished, with most information workers using computers to aid them. Communication technologies have also received consideration, with sophisticated audio communication schemes and widespread use of electronic mail becoming prevalent. Video communication technology has been much less successful. There has been little focus from a technological viewpoint on collaboration and teamwork, although it has been well recognized from an organizational stand that these concepts are critical to successful work.

Computer Supported Cooperative Work (CSCW) is a new area of research that can provide insights into the dynamics necessary to provide work environments that consider not just automation and communication aspects of work, but also collaboration. CSCW is computer support for multiple persons to communicate and work together in an automated environment; however, there is much debate on the boundaries of this field as well as over a core definition (Bannon and Schmidt, 1991). Many computer applications can be considered CSCW applications, including electronic mail, bulletin board systems, networks, and various forms of shared window systems that support multiple users. This last group of applications are often referred to as groupware (GW). Both the terms CSCW and groupware are at present not well defined.

GW can be any software application that supports groups of people working together. Gibbs (1989) defines groupware as software systems which support two or more, possibly simultaneous, users working on a common task and which provide an interface to a shared environment. Humans can communicate through these systems either synchronously or asynchronously, where asynchronous communication refers to the ability to communicate at different points in time, such as is the case with most e-mail
systems. Synchronous communication refers to the ability of a system to allow its users to communicate in real time.

This study takes an interest in a subset of GW, those systems that allow synchronous manipulation of a single object (document, drawing, etc.). In particular, a software product developed by Group Technologies of Arlington, Va. called Aspects™ is used to evaluate co-authoring (multiple users collaborating in the production of a written document) in several different communication environments. This evaluation is of interest for several reasons. Firstly, little objective, statistically relevant research has been conducted on co-authoring systems due to the complexities of evaluating writing and the lack of available systems that support co-authoring. Secondly, there have been few studies on the relationship between various communication media and GW, again due to the lack of availability of GW systems and the dearth of focus on communications media and GW. Lastly, objective research in general on GW systems is needed as these systems and the work methods that need to be developed for them are newly emerging as technologically feasible and useful aspects of the modern workplace.
Purpose

This study attempts to answer several questions, particularly relating to group writing or co-authoring activities using a group editor. Firstly, a question that arises in any study involving a GW product is whether the software will affect the quality of work produced. GW is designed to aid groups working together, and a fundamental question for any type of GW is if it will at the minimum not create difficulties in the process of group writing. It is almost an implicit assumption in studies of GW products that the product will in some way aid the group writing process. By quality it is meant grammatical measures of written work in this context, as well as measures of the correctness of the information in the written work. The null hypothesis for this question states that the quality of the documents as measured by grammatical means and by correctness of and chronological sequencing of its informational content (independent of time pressure) will not differ under any of the experimental conditions. If the null hypothesis is accepted for this question, then the content validity of questions depending on time to complete the task data is heightened; this would avoid questions concerning the relationship between the dependent measures.

The second question this study attempts to address concerns media environments for group writing. If the GW product is treated as a component of such an environment, similar to audio communication, asynchronous text communication (E-Mail), or video communication, what combinations of these media will provide the most efficient performance for a given type of group work? This question is addressed in the described study for several combinations of the aforementioned media, in the context of a writing task. The combinations used are audio only, audio plus video, and face-to-face; each of these with and without the GW for a total of six conditions. The reason for the exclusion of textual communication was twofold. One, the GW used here supports that communication, and to maintain conditions that would be appropriate to real world usage
it would be unreasonable to limit the usage of the GW product. Two, all of the communication media involved in this study are synchronous, and adding an asynchronous communication media would confound the media results with a factor of synchronicity. Other media not investigated in this study, such as audiographic communications, are not as widely used as the aforementioned media. Furthermore, there are questions as to whether any of these other media would have utility in a group writing situation. Performance is measured by time taken to complete the task. The null hypothesis for this question states that there will be no significant differences in time to complete for any of the experimental conditions.

A final question concerns characteristics (ease of use, essentialness, etc.) of the features and operations of the text processing feature of the Aspects™ GW package, as measured by subjects' perceptions of the operations of the GW separate from and in conjunction with the various media environments explored in this study. Important information regarding the usability and perceived utility of the GW alone and in different media contexts may suggest directions to take in improving the group user interface.
Literature Review

There are several studies in the literature that relate to the areas mentioned above. These studies can be categorized into four major areas: social psychological work on groups, communications media studies, work involving methods of evaluating writing, and studies specifically involving CSCW and GW. Vaske and Grantham in their 1990 book describe some social psychological considerations in group work. Kraemer and Pinsonneault (1990) review a group of studies related to Group Communication Support Systems (GCSS) also with a bent toward social psychological issues. Stefik, Bobrow, Foster, Lanning, and Tatar (1987) discuss an early manifestation of group work applications with their description of WYSIWIS (What You See Is What I See), a general theoretical model for GW; the authors also consider some social psychological aspects of GW. Several authors including Chapanis (1975, 1976, 1977) describe work with various communications media and tasks. Lastly, there are two general groups of work of particular interest: research on shared windows, research on shared workspaces, and several examples of systems that combine both of the former approaches (Tang and Minneman, 1990; Ishii, 1990; Watabe, Sakata, Maeno, Fukuoka, and Ohmori, 1990; Smith, O'Shea, O'Malley, Scanlon, and Taylor, 1991; Gale, 1991; Neale, Jenkins, and Green, 1992). Many of the systems used closely resemble the software package used in this study, with the Neale et al. study (1992) actually using the same system.

The paper by Stefik et al. (1987) describes the application of WYSIWIS to the design of a shared window system, Boardnoter. This system’s implementation caused some of the dictatorial edicts of the WYSIWIS idea (that all users must see exactly the same scene) to be repealed. Firstly, it was found that display of all person’s pointing devices was distracting; therefore, only the pointing devices that were in use for editing or for telepointing (pointing with the specific purpose of showing others) were shown. This same concept is utilized in Aspects™, the software used for the study described in
the methodological section of this paper. Secondly, it was decided that only relatively large portions of individual data input streams were to be displayed to all users during input operations; display of very fine detail of this stream caused performance to lag considerably. Aspects™ follows this suggestion. Thirdly, the portrayal of selection of an object by an individual was identical to the portrayal of selection of an object by another individual; this was solved by distinguishing between a local selection and a selection by another group member. Again, Aspects™ shares this feature. Other issues were described concerning this system, but they do not directly apply to the study described in the latter section of this paper.

Olson, Mack, and Wellner (1990) describe a certain type of GW as a “group editor.” This type of software allows multiple users to work on the same object (e.g. text document, drawing). Group editing software’s goal is to improve speed and quality of writing (Dalton, 1987, reviews five available platforms). Olson (1989) states that these goals may not be initially achievable due to barriers in behavioral change of GW users. Various researchers use different syntax to describe the same set of functions or characteristics of a given piece of software. The important aspect of a group editor is that the whole group, either face-to-face or in remote locations, have synchronous access to the same information object. This architecture follows the WYSIWIS (what you see is what I see) principle stated in the previous paragraph (Stefik et al. 1987). This type of group editor contrasts with those that allow serial access to the work object; for example, Timbuktu™, or Cognoter (Grief and Sarin, 1987). Even those systems that afford synchronous access to the work object might not allow simultaneous access (two or more people entering or editing information at the same time). Aspects™, the group editor used in this study and also in the Neale et al. (1992) study is a group editor that allows simultaneous access to work objects.
Group Interactions. Groups are composed of individuals; however, group processes impose their ambiance on individuals that cause interactions, produce certain communication structures, and produce an environment that is absent when individuals are not in a group. Groups also produce other aspects of architecture, the first being pacing. Stefik et al. (1987) state that the order in which individuals take turns in a group and the spontaneity of their contributions both affect pacing. The pacing used, therefore, affects each individual's cognitive processing style, which is different than when they are not in a group. For instance, a group's pace that is slower than some individual's cognitive processing rate might produce boredom or non-participation with that individual. In addition, the pace must also foster learning in all group members for successful participation. In summary, there are many group dynamics that act upon a group. The depth of that analysis is beyond the scope of this paper, but several more important social psychological aspects of group work will be outlined to clarify their roles in GW and to ensure that they are considered in the methodology of this study.

Many philosophers and social psychologists have asked the question, who produces more, the group or the individual. The group may be more efficient by dividing the labor or more time consuming by the effort of duplication (Collins and Guetzkow, 1964). The outcome, however, is highly dependent on the nature of the task and the environment in which the task is performed. The group environment produces obstacles originating in both the task (stimuli external to the group) and the interpersonal environment (expectations about other group members). These factors affect both individual and group productivity.

Communication plays a critical role in group production. Members of a group must create other member's attention, represent ideas, persuade others, resolve conflicts, and carry out goals. McGrath (1990) proposes that there are four stages of group task activity: 1) Formation of a goal or intention and initial selection of a performance
strategy; 2) problem solving; 3) conflict resolution; and 4) execution, performance, and goal attainment. Gale (1991) and Smith et al. (1991) provide some results linking these types of behaviors to various media, and these will be discussed further in the section relating to studies in CSCW.

Vaske and Grantham (1990) point out several different classifications of work groups that affect groupware, namely by type of group, task performed by the group, and cohesiveness of the group. Firstly, the type of work carried out by the group is classified into two general divisions. Divisive work is work wherein members of a group perform different, complementary tasks. Unitary work describes work wherein group members all perform the same task. They claim that groupware is only useful for divisive groups.

However, in the case of co-authoring and decision making (evident in the task described later in this paper, which is a unitary task), each author may be performing the same task, yet they may need to collaborate to ensure that each person's portion of the final document complements the other's. Vaske and Grantham (1990) also talk somewhat about group size in cooperative work. They note that for a disjoint task (performance determined by best member) a larger group (they speculate up to 30 persons) is always better. For a conjoint task (performance determined by worst member) smaller groups are better. Unfortunately, it may be hard to classify certain groups as disjoint or conjoint or tasks as divisive or unitary, and specific numbers for group size are not given for the former classification. Further, it is often difficult to determine who is the “best” or “worst” group member, and thus determine productivity of the group. Lastly, group cohesiveness is discussed. The authors claim that a high level of cohesiveness enhances whatever norms are already present. The example they give is that of a group of secretaries: if the group is highly cohesive and is predisposed to idle conversation, then a great degree of this will take place; if the group is predisposed to serious concentration on work activities, then this will be the predominant activity. In less cohesive groups,
neither extreme is often seen. The study described in the latter chapters of this document and many of the studies reported in this section use unfamiliar groups which have a low level of cohesiveness, therefore eliminating to a large degree conflicting norms of participants.

Kraemer and Pinsonneault (1990) report some interesting findings in their literature review of studies relating to GCSS (Group Communication Support Systems). These systems include GW and specifically, group editors. It was found that there was a positive correlation between these systems and group depth of analysis for a broad variety of decision tasks. One particular example of this was with the use of the Arctic survival task (Turoff and Hiltz, 1982; Eady and Lafferty, 1975). Participation of group members was found to increase and domination of members was found to decrease with GCSS, although this was only said to apply to groups that had not functioned together for long periods of time. Further, use of these systems was found to continue to have this effect, if the reverse trends had not been present before the studies. GCSS were also said to enhance the groups’ current structure with regards to consensus and cooperation; that is, if the study was done with new groups, there would be a low level of consensus. This last finding is highly correlated with the level of satisfaction with the group process that members had. Lastly, these systems had several different effects on the decisions reached by the groups. GCSS were found to increase decision time, possibly due to their effect on the preceding variables. GCSS seem to negatively affect decision confidence with already established groups, while increasing confidence with new groups. GCSS seem to improve decision quality in almost all cases.

CSCW Studies. There have been some interesting developments in shared desktop and window systems. Tang and Minneman (1990) discuss VideoDraw, a shared desktop system. Their system allows users to see hand drawings and hand movements. Although this is not a computer system per se, it does approach the idea of the shared
workspace. The most innovative development in this system is the use of overlaid video signals to enable the display of multiple hands and paper drawings. Watabe et al. (1990) discuss MERMAID, a fully shared workspace system with voice, video, and multimedia computer sharing. As with the Tang et al. paper, this is a discussion of a system development, with no objective research on the impact of this system as compared to other systems. This system is unique in its combination of multi-media document sharing with video and still (facsimile) transmission, along with group editing facilities. The authors note the various uses of floor control in their system; a facility for a moderator to control the floor was used with meetings which had clear authority figures (e.g. board meetings), facilities for first-come-first-serve floor control were used in meetings of equals, and a free editing facility was used for brainstorming. The chief drawback with this system is the large amount of capital needed for various types of specialized equipment and communications channels. The facility for group editing (editing by multiple persons simultaneously) of a document distinguishes this system, but the cost and difficulty of implementation of the system detract from this feature.

TeamWorkstation, a similar system, (Ishii, 1990) again is a discussion of a system development. This system stands partway between VideoDraw and MERMAID; it includes the availability for video overlay, but does not allow group editing. It also has video and audio transmission, but does not support multi-media or any other type of shared document; although users can see other users’ documents and overlay them. Again, specialized equipment and communications channels are required for its use. The paper does discuss some theoretical issues in group work. Ishii discusses the idea of “seams”: seams refer to boundaries in passing information between differing modalities (e.g. the boundary between computer documents and paper documents; between individual work and cooperative work; between video, audio, and face-to-face media; and between synchronous communication and asynchronous communication). Aspects™
does not attempt to bridge these information gaps; it concentrates on the group editing of shared documents, a facility MERMAID shares. Unfortunately, few of these papers actually objectively evaluate the combinations of different media in their respective systems, or compare their systems (other than anecdotally) to other systems.

A recent addition to the work on TeamWorkstation by Ishii and Miyake (1991) does provide some objective statistical data concerning their system. In this study, two different modes of workspace were compared: an overlaid picture mode and a tele-desk mode. The overlaid picture mode is an analog combination of video signals from the two workstations, whereas the tele-desk mode is a digital representation of the signals from the workstations. These two modes were compared in a mixed factor design with other factors being order of mode usage and instructor in a calligraphy task. The overlay mode was found to be subjectively preferable, and also reduced task time.

A most useful paper was written by Lauwers and Lantz (1990). This paper focuses entirely on GW systems; specifically, multi-user collaborative systems designed for other than meeting room use. The authors make the distinction between Collaboration-Transparent systems, which are operating systems that attempt to make existing application software into multi-user groupware, and Collaboration-Aware systems, which are applications specifically designed for group use. This latter category contains software such as Aspects™, MERMAID and other group editors. The authors conclude that Collaboration-Transparent systems are virtually impractical due to computer performance and architecture limitations, and that focus on Collaboration-Aware systems is necessary. They bring up several issues of interest, although mainly on a systems design level. Firstly, the question of spontaneous interactions (what to do about latecomers to a conference who need a history of changes; this includes problems relating to the identification of the latest version of a document) implicates large amounts of computer space and power to keep records of changes. Secondly, the question of
workspace management (which windows are shared, which are private) brings forth the need for separate window identification methods. Thirdly, the ever present question of floor control (discussed earlier in MERMAID, Watabe et al., 1990) always raises a problem relative to who is allowed to edit documents at what time during a group conference. Fourth, annotation and telepointing (see discussion on Boardnoter, Stefik et al., 1987) again becomes a problem related to the confusion of multiple pointers. Specifically, it is difficult to determine who has control of what pointer. It should be noted that Aspects™ solves this problem with different pointer icons for each user. Lastly, performance lags are discussed relative to the number of users and the computational intensity of shared applications and systems. This article suggests design alternatives from anecdotal experience with these type of collaboratory systems; it does not present objective research on the issues.

There is one research issue which has received little attention in the literature. This is the issue of the interaction of groupware systems with other communication media. Ishii (1990) in his discussion of TeamWorkstation, mentions that the video and audio media help with the establishment of floor control (who has the permission to edit) and other social protocol issues. Tang et al. (1990) mention the usefulness of the display of hand gestures in floor control. Watabe et al. (1990) cite anecdotal results concerning both audio and video channels in MERMAID: the audio channel was the most frequently used, and the video channel “enhances the visual effectiveness of conferees”. Exactly what this means is difficult to determine, but it is further said that video helps participants become more familiar with one another and accelerates informal discussions. Gale (1991) and Smith et al. (1991) also consider these issues, but sufficient data for statistically different results was not collected. In these papers some anecdotal information has been presented on new systems that combine various communication media, but no objective research with enough data to provide statistically significant
differences has been conducted on the value of the various media combinations in specific task environments.

SharedARC (Smith et al., 1991) is a GW system which involves a shared computer environment. This system was tested with various communication modes to ascertain effects on a standardized problem solving task, the so called “running man in the rain” task. This task, similar to non-computer tasks, has an objective solution, and no preset time limit. The task affords both negotiation and collaboration between team members in pursuit of the best solution. Two person teams (dyads) were assembled for the study by Smith et al. (1991), with the team members meeting for the first time in the experimental setting. A short training session was given to insure system operability.

Multiple communication modes were used in this study in addition to the computer system, including audio communication, face-to-face communication, and a method of video communication called the “video tunnel”. The “video tunnel” was a closed circuit video communication system that used mirrors to ensure straight lines of eye contact through the video media. The video data from the tunnel, along with other video data, were used to help assemble a set of measures of work activities. These measures included a categorization of team activities, levels of discourse, and patterns of eye contact. The team activities were categorized into interface, task, and social activities, goal related versus specific activities, and error recovery activities.

Several statements were made about the resulting data from the SharedARC study. Firstly, the video communication media was used in spurts, primarily for goal related discussion. This activity included frequent eye contact and gesturing, including “finger telepointing” wherein subjects attempted to use their finger instead of or in combination with the system’s telepointing device, even though the system did not allow view of the finger usage by the other participant. Video may also have helped in task related non-interface work; subjects who did not have the use of the “video tunnel” had
increased interface time possibly showing a decreased collaboration effect compared to those subjects who had the use of the “video tunnel”. Further, Smith et al. (1991) stated that video may have aided participants in task division, and may have reduced “social distance”, typified by the ability to “point over one another’s shoulder”. The developers of SharedARC promote more studies of this nature done with repeated measures designs and larger numbers of subjects to provide more statistical power.

Gale (1991) authored a study that considered combinations of various media in conjunction with a GW system consisting of a shared whiteboard. A shared whiteboard is a computer “window” with multi-user access. Specifically, the media combinations offered were whiteboard, whiteboard plus audio, and whiteboard plus audio and video. Three tasks were used in this study; two writing tasks, assembling a report and a creative writing task; and a third scheduling task. Several interesting results were reported. Firstly, there were no statistically significant time differences resulting from the writing tasks, only from the scheduling task. This was thought to be due to large amounts of variance in the data. The quality of the work produced in all conditions was said to be similar. Some subjective data provided more clear statements. The computer whiteboard plus audio was most favored by the participants, with the whiteboard said to provide a good medium for structuring and summarizing information. The writing activities worked best in a partitioned whiteboard workspace. The audio channel was used primarily for control of the work activities, not for information transmission. Video, when available, was used for subjective assessment of the readiness of group members for work activities. For this purpose, the video channel was used quite heavily. Gale (1991), in discussion, states that tasks used in evaluating GW systems should be more sensitive to social interaction. Also, the large amount of individual differences in these studies should be noted and taken into account in pursuit of statistically significant differences in the data.
A methodological paper addressing the issues involved in empirical studies of GW and specifically co-authoring systems provides insights regarding difficulties in these types of experiments (Neale et al., 1992). Here, four conditions were used to attempt to evaluate the effects of a group editor on writing style and time to complete. These conditions were: free group editing with members of a dyad in separate locations; dyad separated with only one member able to edit; dyad face-to-face with one computer; and one subject with the computer. The task used, termed the "Clue" task, is very similar to the task which was used in the study described in chapter two of this document. The Neale et al. (1992) paper points out difficulties in three areas relevant to this type of study: population variability, sensitive dependent measures, and tasks that provide negotiation, cooperation, complexity, collaboration, and discussion. The findings in this study were similar to those of the Gale (1991) study relative to the variability found when attempting to assess writing task results: these large variabilities are a critical issue addressed in the methodology of the research described in chapter two of the study presented herein.

Several characteristics of a writing task create measurement difficulties. Grammar is difficult, albeit not impossible, to measure objectively, and style can only be assessed subjectively. There are difficulties in using the "Video Clue II™" writing task if a within-subject design is used. Comparing different video clips, even with the same actors and other similarities, is questionable. Furthermore, when groups consist of strangers, repeated conditions may introduce variability created by the development of a group relationship over time. Likewise, there may be a confounding effect of task learning.

Neale et al. (1992) suggested that one method of decreasing the effect of individual and group variability when repeated task runs are not possible is to select participants from a relatively homogeneous population. For an authoring type task, this
could be achieved by selecting subjects enrolled in an advanced composition course. Various types of blocking methodologies can also be used to decrease variability. Factors to block on include writing ability as defined by standardized writing tests and typing ability (i.e., for time measures).

A summary table of studies with either statistically significant results or great relevance to the purpose of this document is shown in Table 1.
Table 1. CSCW Studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Subjects</th>
<th>Measures</th>
<th>Significant Results</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ishii et al.</td>
<td>Mixed</td>
<td>2 Instructors; 4 students each</td>
<td>Survey; time</td>
<td>Overlay preferable; reduced time</td>
<td>Calligraphy Task</td>
</tr>
<tr>
<td>Smith et al.</td>
<td>Between</td>
<td>3 Dyads per condition</td>
<td>Work activity measures from video analysis</td>
<td>Anecdotal: Video for coordination</td>
<td>Running Man in Rain task; suggest use of within design</td>
</tr>
<tr>
<td>Gale et al.</td>
<td>Within</td>
<td>5 trials of groups of 4 for each task from a pool of 12 total subjects</td>
<td>Survey; time</td>
<td>Whiteboard plus Audio preferred</td>
<td>From Observation, Video good for assessment of team members and Audio for coordination; Individual differences notably high; Several tasks, including creative writing</td>
</tr>
<tr>
<td>Neale et al.</td>
<td>Between</td>
<td>4 Dyads per condition</td>
<td>Grammar; time</td>
<td>none</td>
<td>Variability with grammar measurement; suggest within design</td>
</tr>
</tbody>
</table>

**Communications Studies.** In order to fully understand relations between conferencing media such as audio and video and media such as collaboration-aware application software (group editors included; see Lauwers and Lantz, 1990), some review of articles discussing the relative merits of audio and video conferencing has been undertaken. Much of this information has been gathered from a book by Johansen, Vallee, and Spangler (1979) which reviewed many studies related to teleconferencing.
Several articles dispute differences between face-to-face and video channels of information, and the value of video communication channels. The issue of video as compared to other media is prevalent in the literature, with mostly negative or no-difference results reported. This group of results would seem to differ from the subjective (anecdotal) evidence offered by the Smith et al. (1991) and Ishii et al. (1990) GW studies that contended that video offered organizational and “effectiveness” benefits to conferees. Champness and Reid (1970) determined that information exchange was equivalent in video conferencing as opposed to face-to-face. Larimer and Sinclair (1969) found that students taught by video media did not perform as well as those taught face-to-face. In other reviews of communications literature (Williams, 1977) video was said not to aid group work overall, due to its addition to task time caused by high levels of superfluous social interaction. Further, video and to some degree audio media inspire fear when combined with computers (Olsen, 1989). In contrast to these articles, Westrum (1972) determined that increased task complexity increased the probability that video teleconferencing would be the medium of choice. Also contrary to the many negative findings concerning video, however, Egido (1990) reports that video may add some subjective improvement if cheap and high bandwidth; this adds some credence to the GW studies’ tentative findings concerning video. Another voice in support of video comes from Scherer (1982) who says eye contact was found to be important to interpersonal interaction, and that this could help to specify the type of video necessary for a given situation.

Chapanis and his colleagues conducted a series of studies (Chapanis, 1975; Weeks and Chapanis, 1976; Chapanis, Parrish, Ochsman, and Weeks, 1977) which used a variety of interesting tasks and found conflicting results concerning video. The tasks used differed in two separate dimensions. First, there were groups of cooperative tasks, and conflictive tasks (see Sheffield 1989 for a similar conflictive task). These included
the class scheduling, issue ranking, and budget negotiation tasks. Secondly, there were spatial and verbal tasks (see Wickens, 1980, for a discussion of spatial and verbal perceptual channels). These included the geographic orientation, doctor location, and assembly tasks. These different types of task allowed examination of two new variables objectively, task orientation and communication channel.

Chapanis and his colleagues found several interesting results in their studies concerning the comparison of video communication with other communication media. It would be expected that more communication would take place in face-to-face situations, and several studies found that in the face-to-face condition, more words were used and more unique words were used when solving factual-type problems (Chapanis et al., 1977). Subjects in the voice-only condition took longer (about 14% longer) to solve the problems, but the use of voice-only media resulted in the reduction of words and unique words used. In addition to these findings, Short, Williams, and Christie (1976) found similar results in an extensive review of literature. Very few other studies have shown any differences between communication with channels carrying only audio information and those channels carrying audio and visual information. Further, video has in the past proved detrimental to problem solving (Short, 1974). The study by Weeks and Chapanis (1976) determined that for cooperative or confictive tasks there was no significant difference between the two media. An earlier study by Chapanis (1975) determined little differences using either a spatial or a verbal task between the two media, although it was found that both of these conditions took significantly less time than a text-only condition. Reid (1976) found similar results to those of Chapanis. Chapanis also found that there were other significant measures besides time. The 1977 study found seven valuable linguistic measures, including the aforementioned ratio of unique words to words (type token ratio).
An experiment by Brown, Geller, Goodnow, Hoecker, and Wish (1980) had some significant findings, and demonstrated a useful experimental task. The most relevant aspect of this study is the use of groups; the number of groups used has implications for power requirements in other group studies. They compared a voice-switched single participant view videoconference to a full group view situation. Two groups of three participants used full synchronous audio communication. They used subjective measures, but they also used an objective measure of solution quality resulting from their experimental task. The task used was ranking, first individually then as a group, a list of 15 survival items in order of survival importance under specified desert or subarctic conditions. Expert opinion scores were relied upon to provide the “best” ranking or objective score for a given ordering; the measure of group effectiveness was the difference between the average of the participants’ six pre-discussion individual scores subtracted from their group score. With both measures, the group picture mode was significantly better than the participant view mode. Questions from a post-conference questionnaire (the subjective measure) included “To what extent did today’s discussion have the feel of a face-to-face meeting?” and “During which session [conferencing condition] did people seem closer?” The task mentioned here was the “survival” task, taken from an educational text (Eady and Lafferty, 1975) and has been used in other studies. It should be noted that eight groups were used for each condition in the study, and that this number of groups provided sufficient statistical power for the experiment.

Writing studies. Several interesting points concerning writing are relevant to any research that uses writing tasks. Many interesting points concerning the natural use of computer writing tools came out of a research report by Guerrero (1989). He states that teachers in the New York City school system believed that collaborative learning promoted the goals of problem solving and report production, and that certain types of software aided in that process. Further, he states that most collaborative work occurred
naturally in pairs; this is a good indication that studies of collaboration-aware software using dyads are relevant to real world activities. Further, the students who used these systems received no special training on social skills for collaboration, again supporting a lack of such training in studies of GW. One other interesting facet of the report mentioned that little or no role or task differentiation was observed in the group activities. This supports using similar subjects to be members of a dyad, as the task requirements would be equal for both participants; additionally, this supports the use of a unitary writing task in this type of research.

Objective measures of the quality of written documents would be very valuable in any research using a writing task. Unfortunately, there are very few methods of evaluating writing, and the only objective measures are grammatical. Fagan, Jensen, and Cooper (1985) along with Cooper and Odell (1977) give a layman’s guide to evaluating writing, but their language and their field (Rhetoric) are basically subjective, and could not be called upon to provide quality measures for the research described later in this document.
CHAPTER 2 — METHOD

Task

In CSCW studies, where teamwork plays a significant role, the task used is an important part of a successful experimental methodology. As mentioned in the Neale et al. (1992) study, there are several requirements for tasks used in GW studies. These requirements include negotiation, collaboration, cooperation, complexity, and homogeneity of data (objectiveness). By using a variant of the “Clue” writing task used in the Neale et al. study (1992), these requirements could be satisfied for a writing task. Negotiation came about when members of a dyad did not agree on what they observed in the video that provides the task information (e.g., what were the characters searching for in the fortune cookies?). Collaboration could be observed where participants remember certain pieces of information from the video that their partners did not, and information was communicated about those events. Cooperation is evident in any group writing task, when participants edit each other’s work. Organizational problems relating to the creation of a written document provided complexity, as did the act of writing prose itself. Lastly, the objective nature of the data to be included in the prose report on the video provided homogeneity in the written documents; there were definite correct answers to the questions that were required to have been answered in the document.

As was stated above, the task used for this study was a variant of the original “Clue” writing task. The “Video Clue II™” writing task used in the Neale et al. (1992) study conforms to the aforementioned task requirements. In that task, subjects were required to have at least one hour of Macintosh™ experience and were disqualified if they had ever played Video Clue™, Video Clue II™, or Eyewitness™. Subjects were instructed that they would be viewing a video clip that contained various actors, objects, and locations, and then they would be required to write a report on a computer discussing their observations and providing specific details. Subjects were given two minutes to
familiarize themselves with the particular actors' pictures and names that were on a set of cards. Next, subjects viewed an introductory video clip to introduce them to the type of video presentation that would be used in the experiment. This clip also showed the actors and weapons.

The subjects then viewed a practice scene, and they were told to identify actors, weapons, rooms, and other details contained in the video clip. Five practice questions were administered to familiarize the subjects to the level of detail they would be required to recall.

Subjects then viewed a five minute scene for the experimental condition. Subjects were told that they could take notes. The subjects then were instructed to concentrate on writing a prose document that included the facts requested (i.e., 20 randomly ordered questions were given about the video clip), and that kept the facts in the proper chronological sequence (i.e., as they appeared in the video clip). Subjects were given ten minutes of instruction on the software.

One aspect in which the task used for this study differed from the task used in the Neale et al. (1992) study related to the use of a repeated measures design, recommended for further research by that study and several others (including Smith, 1991). Instead of each dyad watching a single scene from the video, multiple scenes were watched and reported on. The original Video Clue II™ game comes with 15 different usable scenes in three categories of complexity. In this study, the final scene from the second level of difficulty and all five scenes from the most difficult category were used. This material used as a basis for the question portion of the task, is especially useful due to this quality of multiple portions of similarly rated (as developed by Parker Brothers) difficulty. In addition, questions on each video clip were pretested to insure similar levels of difficulty.

**Questions.** The questions used for each scene were derived from information presented in the video clips. They were organized within groups relating to specific
locations where action takes place in each video clip. These questions are presented in Appendix II. Each set of 15 questions was tested against each other set to insure equal difficulty. Six individuals not participating in the experiment were be asked to view each video clip and give short answers (fill in the blank) to the appropriate set of questions. The order of evaluation of the questions for each video clip by the six volunteers was balanced with a Latin square. A single-factor within-subjects ANOVA done on the number of correct answers to each question set did not show the question sets to differ significantly in difficulty, with a Greenhouse-Geisser corrected P-Value of 0.45.

**Product.** The document that subjects were asked to produce with this variant of the “Clue” writing task was a newspaper article. This type of format was used because the “hard news” writing style is one of the most restrictive writing styles, aiding in production of comparable documents by the participating teams. All conventions of style common to journalistic writing were observed in the writing of this document. The content of this document was based on the aforementioned 15 questions given to the subjects about the scene shown for each trial. Questions were randomly ordered within sections relating to different rooms where action occurred in the video clips. As in the previous use of the “Clue” writing task, not only correct answers to the questions were required, but also the information required by the questions was mandated to have been given in correct sequence as it appeared in the video (in real world journalistic work for a chronologically organized article, sequence is determined by chronology with the addition of a “lead”; see Hough, 1988). There was a “newspaper submission” deadline of 1 hour for the work. Subjects were told that their first priority was to produce the best quality (grammatically, stylistically, and by content) document they could. Their second priority was to finish the article as quickly as possible.

Some journalistic conventions had to be violated for subjects to conform to the task requirements; these conventions and others are described in a book by Hough.
(1988). "Hard" news stories must begin with a "lead". This is a short sentence or group of sentences that answer the questions who, what, where, when, why, and how about the story. Since these facts were not always available, a full fledged lead was one of the typical areas where subjects could not fully conform to journalistic style. There are also several different formats for the body of the article, the most pertinent of which for this application was the chronological format. In this format, following the lead, events are described in more detail in chronological order. A story of this type may or may not have a conclusion, depending on the content and the writer. News stories also must follow several additional general style restraints, such as specific formats relating to quotations, use of tense, and identification of protagonists. Lastly, there are a set of required stylistic (in the sense of abbreviation, grammar, etc.; much like the Human Factor's Society's Author's Guide) conventions that must be followed; these are described in the Joint Stylebook published by the Associated Press and the United Press International.

Task Pretest. For this study, to ensure low levels of variability in the newspaper writing aspect of the task, a task pretest was done with four of the subjects who participated in the experiment (2 dyads). This number was deemed sufficient for pilot testing; one dyad would not enable comparisons between teams. The two dyads were volunteers from the six dyads involved in the study; teams two and three volunteered. The dyads were asked to write three one paragraph newspaper articles describing Aspects™ GW. Each dyad was given three fact sheets from which to produce these articles, and a computer running the off-line (not in networked, or simultaneous group editing mode) word processor capabilities of the GW to create the articles on. The fact sheets for this pretest, reproduced in Appendix I, were developed to be very similar; each has an equal number of facts, and no elaboration on these facts was included. Note that the conditions in which the dyads created these articles were the same as those for the face-to-face and no GW treatment combination of the experiment (a control condition).
Subjects were timed in their completion of the articles, and grammatical measures of their work were taken. The principle results of interest from this pretest were: a confirmation of a high level of written quality in the resulting articles as measured by grammatical means, and an estimate of the variability in time to complete an article. It was important for the two teams tested to have relatively comparable time and quality results (the latter measured by computerized grammar checker), but more importantly that each group exhibit consistent performance.

The two groups results were comparable to each other, but not consistent. The data from the task pretest is presented in Table 2. There was a difference of only two minutes between the average times for the two groups for each article, but times ranged from 8 to 29 minutes between trials. This is thought to indicate a practice effect or non-similarities in the fact sheets, which does not call into question the writing task itself. Grammatical quality for all articles was excellent, as measured by computerized grammar checker. Four of the six articles were error free; one article had a punctuation error, and one had a spelling error.

<table>
<thead>
<tr>
<th>Trial:</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team 3</td>
<td>27 min.</td>
<td>11 min.</td>
<td>16 min.</td>
</tr>
<tr>
<td>Team 2</td>
<td>29 min.</td>
<td>8 min.</td>
<td>13 min.</td>
</tr>
</tbody>
</table>

Table 2. Task Pretest Results.

**Treatment Conditions**

There were six treatment conditions in this 2x3 within-subjects design. The two levels of the GW factor were simply computers networked together using the
simultaneous writing capabilities of the GW, and computers not using the simultaneous writing capabilities of the GW. In other words, one level provided GW simultaneous writing, textual communication, and viewing of the document, and the other level mimicked a setup that did not have GW capabilities. For this latter setup, only one member of a dyad had a computer, because it was desired to have the dyad produce a single document. The other member of the dyad communicated to help produce the document through other media than the computer. The reason for this choice of levels was that this kept the study as simple as possible, and emulated real-world conditions. In actual work situations, either the workers would have GW and use it with no limitations on its capabilities, or they would not have GW and would have used normal, single-user computing aids. The second factor, that of Communication Media, had three levels. The first was to be face-to-face, the second audio only, and the third audio and video. These level choices again were made to attempt to mimic real world work situations. All video conferencing systems include audio, for reasons not the least of which is that technically if a video channel can be provided an audio channel is very easily added. Another reason for audio going with video is that those who can afford a videoconference would not accept it without audio, which minimally adds to cost of the conference. Another reason for including the audio and video channel as a treatment level is that from the preceding discussion of literature it is still unclear whether video is useful, harmful, or does not effect group work. At least for a writing task, perhaps the utility of video communication could be determined. The audio only condition mimicked a simple telephone call, very popular in group work taking place from remote sites. Finally, the face-to-face condition was included because face-to-face communication is used for most instances of group work, and certainly whenever remote sitting is not forced upon workers.
Subjects

Twelve subjects were used in this experimental design, as six dyad pairs. One of the reasons this number was selected was the benefits of using a Latin square for balancing treatment conditions and supplementary statistical analysis. Further, in previous within-subject media studies (e.g. Brown, 1980), ten or fewer subjects or groups have provided sufficient statistical power. Finally, financial and subject availability limitations dictated use of less than twenty subjects. For these reasons, six groups of two rather than twelve (which would have required 24 subjects) were used. Subjects were assigned to dyads based on the results of pretests of writing ability and dominance. To eliminate confounding due to group familiarity (see Kramer and Pinnsonault, 1990), subjects were assigned to be in dyads wherein they were unfamiliar with their partner. Subjects were asked not to communicate with each other outside of the experiment about the experimental material, to eliminate additional face-to-face or audio communication concerning the study or increasing group familiarity at differing rates between dyads. One person in each dyad was randomly assigned to be the lead typist for those experimental conditions wherein only one participant had facilities to type the document into the computer. This person remained the lead typist in all such conditions, to keep continuity for typing effects throughout the study. Lastly, subjects were selected from a homogeneous population to reduce variability due to individual differences in writing ability (Neale et al., 1992). Specifically, subjects were selected from graduates of an intermediate level journalism class, in order to ensure that all subjects were familiar and proficient with journalistic writing. This ensured a homogeneity of written product between dyads.

Procedure

Pretests. Subjects were pretested for writing ability and dominance. This testing took roughly one hour per subject. The results of these pretests were used to match
subjects so that persons in a dyad would have similar writing abilities and dominance levels. This similarity between members of a dyad was important to ensure an equal distribution of work. Without equal distribution of work, it is possible that less communication would have taken place, reducing the statistical power of the results and perhaps increasing variability in the non-GW conditions. In these latter conditions the member of the dyad with control of the written article may or may not have been the more productive member, causing increased variability resulting from the random selection of that controlling person. The procedure for this matching proceeded as follows for each pretest score: first priority was to prevent two subjects that were familiar with one another from forming a dyad. Second priority was to match those subjects with the closest scores on the dominance pretest (i.e. match the two subjects who had the highest dominance scores, match the two subjects with the next two highest dominance scores, etc., with exceptions made to prevent matches of familiar subjects). Third priority was to re-match those subjects with similar dominance scores to dyads of similar writing ability. This was a complex process, and proceeded at the experimenter's discretion to attempt to create as well matched dyads as possible, given the three aforementioned priorities. Scores of subjects and the final matching is shown in Table 3. No subjects had extreme scores; however, members of team five dropped out of the study, and had to be replaced. There was no opportunity to match these replacements, but pretests were done regardless to show any significant mismatch of the members of team five that may have affected the results of the study. A large imbalance between members of a dyad on either the traits of dominance or writing ability could have resulted in an unequal division of task labor, possibly reducing the GW and/or Media effects. Fortunately, the only area of mismatch was with dominance pretest scores, and here the amount of mismatch was only slight; team six which had been matched had the same amount of dominance mismatch. It is therefore believed that team five introduced no
more variability than other teams. The pretests used for writing ability were the writing sample and writing fluency subtests of the Woodcock-Johnson Psycho-Educational Battery (1989, 1990). This test has been extensively used for research purposes (Mental Measurement Yearbook, 1988). The pretest for dominance was the dominance subtest of the Personality Research Form (Jackson, 1987). This test also has been extensively used for research (Mental Measurement Yearbook, 1989).

Table 3. Pretest Scores and Matching of Subjects to Dyads.

<table>
<thead>
<tr>
<th>Team</th>
<th>Subject</th>
<th>Familiarity</th>
<th>Dominance</th>
<th>W. Fluency</th>
<th>W. Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>C, F</td>
<td>14</td>
<td>31</td>
<td>23.5</td>
</tr>
<tr>
<td>1</td>
<td>B</td>
<td>C</td>
<td>13</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>A, K, B</td>
<td>13</td>
<td>34</td>
<td>26.5</td>
</tr>
<tr>
<td>2</td>
<td>D</td>
<td></td>
<td>11</td>
<td>30</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>E</td>
<td>G</td>
<td>11</td>
<td>27</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>A</td>
<td>11</td>
<td>21</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>E</td>
<td>10</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>4</td>
<td>H</td>
<td></td>
<td>9</td>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>I</td>
<td></td>
<td>9</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>J</td>
<td></td>
<td>6</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>K</td>
<td>C</td>
<td>7</td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>L</td>
<td></td>
<td>4</td>
<td>25</td>
<td>29</td>
</tr>
</tbody>
</table>

Following this pretesting and matching of subjects, two dyads volunteered to complete the pretest of the writing task, following training on the use of the software.
Training. Subjects were told to memorize visages of the actors' to be shown later in the video clips on cards that also contained the characters' names, and were tested on their recall of that information. This test involved showing each actor's visage, and asking subjects to recall the character's name. This test was written, to enable spelling checks. The criterion for passage of this test was 100%. Subjects then underwent a short (10-15 minute) training session on the use of the GW. Detailed training materials are presented in Appendix IV: Instructions. Additionally, subjects were allowed to write a practice essay (entitled, *Advantages and Disadvantages of Summer Life in Blacksburg*). A test was given following the training period, and subjects who did not pass underwent training again. This test was an oral examination, with the experimenter asking subjects to invoke each different feature of the GW and to explain its use. Passing criterion for this test was 100% for invoking the features and reasonable understanding of their use (as determined by the experimenter).

Practice. Once the actors had been memorized and the software was understood, dyads underwent one practice task session. This practice session began with the first scene of the first difficulty category of video clips. Five questions were asked for each practice session, and these were answered with a one paragraph news article, using the no GW, face-to-face control condition environment. More practice sessions were to be given as necessary until the dyad had achieved stabilization on the task. Stabilization was achieved when participants were observed using the full capability of the software, and when their solution appeared reasonable (i.e. no more than one unique error was found with the computer grammar and spelling checker). Provision for practice sessions with more than the first scene was made, but all dyads achieved the stabilization requirement after the first practice session. Following the practice sessions, the experimental trials were conducted.
Sequence. A dyad performed as many trials as desired per sitting; this was allowed due to the difficulty of providing for the schedules of both members of a subject team. Each task sequence was split into two parts: the information gathering portion, and the document generation portion. In the first portion, the subjects viewed the video clip for the trial and were allowed to take notes. No discussion was permitted during this phase. The particular video to be shown was assigned based on the session in order to eliminate confounding effects of scene with treatment (i.e. scene five of the second difficulty category was used for the first trial, scene one of the third category of difficulty for the second trial, etc.). Scenes were specifically confounded with the order of the trials to enable testing of the scene/ordering effect with a supplementary Latin square statistical analysis. This ensured that such effects were separated from treatment effects.

The second portion of the task sequence was the document generation activity. The following scenario was used to provide the correct emphasis for generation of the task documents: Subjects were told that they were a pair of journalists reporting on scenes of interest in some newsworthy situation. One of the subjects was told that he or she was based “on-scene” and the other was told that they were based at the home office. The “home office” person (the “lead typist”) was selected randomly. The subjects were further told that their editor had also seen the scenes and had prepared questions that needed to be answered for short articles on each clip (one paragraph per room), and that all questions must be answered (even if a guess was necessary). These articles were to be summary/descriptions of observations of the video clips. Notes taken during the observation of the clips could be used to aid in the writing process. Further, the requested format required events (questions) to be reported/answered in proper time sequence as they actually were observed. Subjects were told that grammatical and written quality was the key priority in produced articles, and speed was the second priority (maximum time deadline of 1 hour). This was explained by noting that last-minute editorial changes to
the documents were costly, even more than costs arising from the time taken to produce the articles. Another newspaper employee would take care of formatting issues and headlines. For each different clip, participants were told that a different communications scheme had been made available (due to the vagaries of corporate equipment availability), but in all cases the document must be written on the corporate computer system (Aspects™). In some cases, the corporation could not afford to have the on-scene member of the team co-located with the home office person (i.e. the non-face-to-face conditions). Detailed instructions are presented in Appendix IV: Instructions.

Payment of subjects for this study was not prorated; payment for maximum time taken was always given (to avoid providing a confounding negative time incentive). Thus, all subjects received equal pay for time spent in experimental trials.

Exit Questionnaire. Upon completion of the experiment, subjects were asked to fill out an exit survey regarding their opinions of the GW and the various media environments. This questionnaire is reproduced in Appendix VI. The questionnaire contains statements to be answered with a five point Likert scale of agreement as well as open ended questions.

A graphical depiction of the experimental procedure is shown in Figure 1.
Figure 1. Procedure for the Study.

Apparatus

Video Viewing Area. Each task session began with a viewing of the scene for the trial in one of the office areas. A VHS tape recording of the scene was played on a VCR and TV.

2 Offices. Two offices were used for this study, one with two Macintosh™ computers and the other with one. These computers were on a local area network, running Aspects™ GW. This software is a standard functioning word processor, with limited formatting and other utilities. It allows simultaneous typing by multiple users, although users cannot type synchronously on the same paragraph. A bar to the right of the document window shows who has ownership of a particular paragraph or group of paragraphs. Users have the option of “linking” their individual document views, that is, the software can be set to display the same portion of the document on each monitor. In this mode, the architecture follows the WYSIWIS principle advocated by Stefik et al. (1987). Users also have the option of using a built in telepointing facility, wherein each user can see the other’s pointer. Lastly, there is a built in “chat” facility, where users can
send each other text messages and alert each other with a message incoming alert bell. Subjects were not encouraged to use this facility (although they were not discouraged as detailed earlier for reasons of generalizability to real world situations). In all cases an audio channel was available.

Several other communication media were available in the office and were used as previously described in the different treatment conditions. A standard closed-circuit telephone with no ringer was implemented in the two offices. Monochrome video (standard television bandwidth) was available of both participants' visages. The cameras were set up so that eye contact could be made comfortably (Smith et al., 1991). This system was set up to the left of both users, to provide the illusion that users were sitting across a table from each other. A graphic showing the experimental workstation is shown in Figure 2.

![Figure 2. Experimental Apparatus.](image)

**Monitoring.** A video editing system recorded a split view of the two participants' visages, to enable future "data snoop" concerning eye contacts between the members of the team. A mixed audio signal from the two participants was also recorded, again for
purposes of future data collection on specific speech patterns. A timer device recorded
time data superimposed on the video tape. Lastly, the final written product resulting from
each session was saved to computer disk.

Statistical Analyses.

A Greenhouse-Geisser correction was used on the data to ensure no loss of
conservatism due to heterogeneity of covariance, shown to exist by an intercorrelation
matrix. A Latin square was used to balance the dyads with the conditions, necessitating
the aforementioned use of six dyads. The use of a Latin square for balancing allowed a
supplementary statistical analysis. This supplementary data analysis allowed testing of
the effect of presentation order using a single observation statistical analysis, which was
confounded with the scene used for each trial. This supplementary analysis enabled
testing of this combined effect of presentation order and scene. It was unlikely that a
practice effect would be indistinguishable from an effect of scene due to the probabilities
of particular scenes generating task times in such a way as to steadily decrease, as they
presumably would under the influence of a practice effect. In addition, this second
statistical analysis allowed testing of the team effect; that is, the between-dyad
differences on the dependent measures. Post-hoc tests were done on all significant effects
of the main analysis: the Least Significant Difference test was used for pairwise
comparisons, as this was an exploratory study, and any differences that possibly may
exist were to be identified. The experimental design is shown in Figure 3, and
presentation orders are shown in Table 4.
<table>
<thead>
<tr>
<th></th>
<th>Face-to-Face</th>
<th>Audio Only</th>
<th>Audio &amp; Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>No GW</td>
<td>n=6 dyads</td>
<td>n=6 dyads</td>
<td>n=6 dyads</td>
</tr>
<tr>
<td>GW</td>
<td>n=6 dyads</td>
<td>n=6 dyads</td>
<td>n=6 dyads</td>
</tr>
</tbody>
</table>

*Figure 3.* Experimental Design.
**Table 4.** Trial Presentation Order.

<table>
<thead>
<tr>
<th>Team</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>F</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>6</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Letter Code</th>
<th>Groupware</th>
<th>Communication Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>No</td>
<td>Audio</td>
</tr>
<tr>
<td>B</td>
<td>No</td>
<td>Audio + Video</td>
</tr>
<tr>
<td>C</td>
<td>No</td>
<td>Face-to-Face</td>
</tr>
<tr>
<td>D</td>
<td>Yes</td>
<td>Audio</td>
</tr>
<tr>
<td>E</td>
<td>Yes</td>
<td>Audio + Video</td>
</tr>
<tr>
<td>F</td>
<td>Yes</td>
<td>Face-to-Face</td>
</tr>
</tbody>
</table>
CHAPTER 3 — RESULTS

Dependent Measures

In this study, time was the key dependent measure. Care was taken that quality did not degenerate due to time pressure (as dictated in instructions to the subjects), since a constant level of quality provided an accurate time measure. Team Differences in time taken did not affect the main effects of the GW and Media Type factors due to the within-subject design. Presentation Order effects did not affect these factors due to balancing.

Various measures of grammatical quality of the documents were taken (using the CorrecText™ computer grammar checker supplied with Microsoft Word 5.0™; Language Systems, 1990) to insure that subjects had followed instructions and that the communications environment had not affected the writing ability of the participants. More than three different grammatical and spelling errors in a given document would have indicated a grammatical quality decrement. A grammatical quality increment could not be measured due to the ceiling effect of documents that had no grammatical errors. For these reasons, grammatical quality could not be used as a measure that would truly show differential effects between conditions, only as a measure that could show conditions that were of lower grammatical quality than other conditions. No observations showed more than three unique grammatical errors.

Two other quality measures of the written documents were taken, in regard to the quality of the content of the written articles. Both the number of correct answers and the number of answers provided in the correct order (both out of 15 questions per document) were measured. These quality measures, due to the lack of ceiling effects, could be used to show differential effects between conditions. No differential effects on these measures were found in the data except for differences found between scene/trial combinations. ANOVAs of these effects and supplementary Latin square analyses of these data are shown in Appendix VII.
The lack of grammatical decrements in the data, combined with the lack of significant differences among the content measures for the main effects of interest in the study, support the assumption that quality of the written documents produced in the study was constant. The affect of this assumption is to increase the content validity of the differences shown in the time data on the main factors of interest.

Results from the exit questionnaire were important in evaluating the perceived benefits and problems with both the software and the software-media environment combinations. Questions using a Likert scale of agreement were analyzed with within-subject one factor ANOVAs and Least Significant Difference tests when necessary to find significant differences in ratings of various features of the software and media environments. A set of general questions was evaluated using a nonparametric distributional fit test. Open ended questions provided insights on how to improve the GW.

Analyses of Time Data

Answers to many of the questions stated earlier in the purpose section of this document were obtained through analysis of the time data. Several analyses were concluded on the time data including: an F-Max test to determine if heterogeneity of variance was present; an intercorrelation matrix to check for violations of sphericity which would necessitate a correction for heterogeneity of covariance; a two factor within-subject ANOVA; a single observation Latin square ANOVA; and a series of F-ratio tests to compare variability. The F-Max test showed significant heterogeneity of variance in the data, but this was discounted due to the robustness of an ANOVA with equal sample sizes. The intercorrelation matrix showed that pairwise correlations between conditions were not close to equal, necessitating the use of a Greenhouse-Geisser correction factor. Significant differences were found for Media Type, Presentation Order, and Team Differences. There were also significant differences
between variances of certain treatment conditions, determined by comparisons of their ratios of F-Values with the F-Max critical value.

An F-Max test was performed on the data initially to check for heterogeneity of covariance. With five degrees of freedom and six comparisons at the .05 level, the critical value for this test was 18.7. This value was well exceeded by the largest ratio between F-Values in the time data, which was 32.58.

The intercorrelation matrix (shown below in Table 5) shows a range of pairwise correlations from .05 to .97; this indicates a clear violation of sphericity requirements. Thus, it was deemed that a Greenhouse-Geisser correction was necessary to ensure that the ANOVAs were appropriately strict.

<table>
<thead>
<tr>
<th>Table 5. Intercorrelation Matrix.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercorrelation Matrix: Pearson's Correlation Coefficients (r)</td>
</tr>
<tr>
<td>Audio</td>
</tr>
<tr>
<td>Audio</td>
</tr>
<tr>
<td>Audio+Video</td>
</tr>
<tr>
<td>Face-to-Face</td>
</tr>
<tr>
<td>Audio/GW</td>
</tr>
<tr>
<td>Audio+Video/GW</td>
</tr>
<tr>
<td>Face-to-Face/GW</td>
</tr>
</tbody>
</table>

With an F-Value of 6.33 and a P-value of .02 (corrected), the communication media type effect was significant (Table 6). Mean times (in minutes) to complete the task for the levels of this factor were: audio only, 31.57; audio plus video, 31.50; and face-to-face, 25.54. Least Significant Difference post-hoc contrast results showed that the face-to-face condition resulted in significantly less time to complete the documents than either the audio alone or audio plus video conditions. However, the audio alone and audio plus video conditions did not differ from each other (Table 7).
Table 6. Main Analysis of Time Data.

<table>
<thead>
<tr>
<th>ANOVA Summary Table</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>G-G*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team</td>
<td>5</td>
<td>1974.95</td>
<td>394.99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groupware</td>
<td>1</td>
<td>241.28</td>
<td>241.28</td>
<td>1.03</td>
<td>.3573</td>
<td>0.4</td>
</tr>
<tr>
<td>Groupware * Team</td>
<td>5</td>
<td>1174.56</td>
<td>234.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media Types</td>
<td>2</td>
<td>287.87</td>
<td>143.94</td>
<td>6.33</td>
<td>.0168</td>
<td>.02</td>
</tr>
<tr>
<td>Media Types * Team</td>
<td>10</td>
<td>227.52</td>
<td>22.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groupware * Media Types</td>
<td>2</td>
<td>175.92</td>
<td>87.96</td>
<td>.73</td>
<td>.5077</td>
<td>.505</td>
</tr>
<tr>
<td>Groupware * Media Types * Team</td>
<td>10</td>
<td>1211.58</td>
<td>121.16</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: Task Time

Table of Epsilon Factors for Groupware
<table>
<thead>
<tr>
<th>df</th>
<th>Adjustment</th>
<th>Dependent: Task Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Groupware</td>
<td>Media Types</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Groupware * Media Types</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.98</td>
</tr>
</tbody>
</table>

*G-G: Greenhouse-Geisser corrected P-Value; Greenhouse-Geisser

Table 7. LSD tests for Communication Media Effect.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio</td>
<td>A</td>
</tr>
<tr>
<td>Audio + Video</td>
<td>A</td>
</tr>
<tr>
<td>Face-to-Face</td>
<td>B</td>
</tr>
</tbody>
</table>

The variability comparisons presented in Table 8 show that there was more than one ratio of F-Values that exceeded the criterion for heterogeneity of variance for the within-subject ANOVA. Supplemental comparisons of variability show that the audio and audio plus video conditions that did not have groupware usage had significantly larger variability than the audio plus video condition that did use groupware. An interaction graph showing the relation of means to standard errors of the means that illustrates the effects of this heterogeneity of variance is presented in Figure 4.
Table 8. Comparisons of Variance for Treatment Conditions.

F_{max}(5,6,.05): 18.7

<table>
<thead>
<tr>
<th>Trial</th>
<th>AV/NoGW</th>
<th>FTF/NoGW</th>
<th>A/GW</th>
<th>AV/GW</th>
<th>FTF/GW</th>
</tr>
</thead>
<tbody>
<tr>
<td>A/NoGW</td>
<td>1.49</td>
<td>5.62</td>
<td>8.92</td>
<td>21.88</td>
<td>3.79</td>
</tr>
<tr>
<td>AV/NoGW</td>
<td>X</td>
<td>8.37</td>
<td>13.28</td>
<td>32.58</td>
<td>5.64</td>
</tr>
<tr>
<td>FTF/NoGW</td>
<td>X</td>
<td>1.59</td>
<td>3.89</td>
<td>1.48</td>
<td></td>
</tr>
<tr>
<td>A/GW</td>
<td>X</td>
<td></td>
<td>2.45</td>
<td>2.35</td>
<td></td>
</tr>
<tr>
<td>AV/GW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5.78</td>
</tr>
</tbody>
</table>

Interaction Plot
Effect: Groupware * Media Types
Dependent: Task Time (minutes)
With Standard Error error bars.

Figure 4. Interaction Graph for Task Time.
The supplementary Latin square analysis of the time data showed an effect of Trial (a combination of scene and temporal effects), as well as an effect of Team Differences. An ANOVA summary table for these effects is presented in Table 9. Times to complete the task decreased steadily with each trial, showing significant differences in the Trial effect (P-value of .003). This effect is shown graphically in Figure 5. Note that due to the use of a balanced Latin square design, this effect did not affect the results of the previous within-subject ANOVA. Significant differences were also found for the effect of Team, with a P-value of .0006. This effect did not affect the previous ANOVA due to the use of within-subject design.

Table 9. Latin Square ANOVA.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial</td>
<td>5</td>
<td>1486.01</td>
<td>297.20</td>
<td>5.27</td>
<td>.0030</td>
</tr>
<tr>
<td>Team</td>
<td>5</td>
<td>1974.95</td>
<td>394.99</td>
<td>7.01</td>
<td>.0006</td>
</tr>
<tr>
<td>Condition</td>
<td>5</td>
<td>705.08</td>
<td>141.02</td>
<td>2.50</td>
<td>.0648</td>
</tr>
<tr>
<td>Residual</td>
<td>20</td>
<td>1127.64</td>
<td>56.38</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: Task Time
Analysis of Questionnaire Data

Significant differences in response frequencies from the exit questionnaires are presented in several charts, as well as with some supporting statements and comments. One other chart of questionnaire data that did not show significant differences is shown in Appendix VII, along with several ANOVA analyses of questionnaire data that showed no significant differences. Several ANOVAs were performed on the Likert scale data and rank data (admittedly, these data may not be considered interval, but the exploratory nature of this study may justify these analyses to suggest avenues for future research; also, the ANOVA is a notably robust test, which can be of use disregarding violations of assumptions). An applicable non-parametric test, the Friedman two-way analysis of
variance by ranks, was not used due to its less powerful nature than its parametric counterpart; in an exploratory study such as this, all possible significant effects are hoped to be found. Note that only those statistical tests that showed significant differences are shown in tables. Several charts showed significant trends in the data which were supported by subsequent ANOVAs, and in one case a nonparametric distributional fit test.

Figure 6 suggests that users had positive opinions relating to Aspects™, and had negative opinions of work conditions without the software. This suggestion is supported by significant differences found in the corresponding ANOVA presented in Table 10.

![Aspects Preferences Chart](image)

**Figure 6.** Aspects Preferences Chart.

**Table 10.** Aspects Opinions ANOVA.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>11</td>
<td>1.12</td>
<td>.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GW</td>
<td>1</td>
<td>15.04</td>
<td>15.04</td>
<td>67.31</td>
<td>.0001</td>
</tr>
<tr>
<td>GW * Subject</td>
<td>11</td>
<td>2.46</td>
<td>22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: Aspects Opinions
Figures 7, 8, and 9 (data derived from questions four through seven, parts a, b, and c) shows that opinions were varied with respect to features of the software, with a possible exception of the “Chat box”. This feature was found to be significantly different in regards to essentialness to work, shown in the ANOVA in Table 11 and the Least Significant Difference contrasts in Table 12. All three other features were perceived as being significantly more essential to work than the “Chat Box”. The “Telepointer” was found to be significantly more essential to work than the “Linked View”.

**Essential to Work**

![Essential to Work Chart]

*Figure 7. Essential to Work Feature Opinions Chart.*
Easy to Use

Figure 8. Easy to Use Feature Opinions Chart.

Detrimental - Should be Removed

Figure 9. Detrimental Feature Opinions Chart.
Table 11. Essential to Work Feature Opinions ANOVA.

ANOVA Summary Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>G-G*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>11</td>
<td>22.00</td>
<td>2.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>3</td>
<td>20.67</td>
<td>6.89</td>
<td>8.32</td>
<td>.0003</td>
<td>.001</td>
</tr>
<tr>
<td>Features * Subject</td>
<td>33</td>
<td>27.33</td>
<td>0.83</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: Essential Feature?

Table of Epsilon Factors for df Adjustment

Dependent: Essential Feature?
G-G Epsilon
Features

\.75

*G-G: Greenhouse-Geisser corrected P-Value

Table 12. Essential to Work Feature Opinion LSD Tests.

<table>
<thead>
<tr>
<th>Chat Box</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linked View</td>
<td>B</td>
</tr>
<tr>
<td>Ownership Bar</td>
<td>B C</td>
</tr>
<tr>
<td>Telepointer</td>
<td>C</td>
</tr>
</tbody>
</table>

Figure 10 (data derived from questions one, two, three, and eight) may indicate several points on which users agreed: that the software does increase productivity and is not distracting; that face-to-face communication is preferable; and that there is little added benefit to video communication. The nonparametric test shown in Table 13 supports these conclusions: for all statements save that concerning the video, there was a definite difference between the distributions of the obtained responses and a uniform distribution, with differences leading to the above conclusions. For the statement concerning video, there was not a significant difference from a uniform distribution, indicating that there were no consensus opinions about it; this also leads to the
conclusion that users did not see a clear benefit or detriment from the added video communication channel. However, two respondents did indicate that video communication was distracting (from question 9b).

**Figure 10.** Aspects™ General Questions Chart.
Table 13. Kolmogorov-Smirnov Analysis on Several Questions.

Relative Cumulative Frequency Data for Various Questions.

<table>
<thead>
<tr>
<th>Uniform Distribution</th>
<th>Aspects Increases Productivity</th>
<th>Video Preferred</th>
<th>Face-to-Face preferred</th>
<th>Aspects is Distracting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Agree</td>
<td>0.200</td>
<td>0.667</td>
<td>0.083</td>
<td>0.333</td>
</tr>
<tr>
<td>Agree</td>
<td>0.400</td>
<td>1.000</td>
<td>0.167</td>
<td>0.833</td>
</tr>
<tr>
<td>No Opinion</td>
<td>0.600</td>
<td>1.000</td>
<td>0.333</td>
<td>0.833</td>
</tr>
<tr>
<td>Disagree</td>
<td>0.800</td>
<td>1.000</td>
<td>0.667</td>
<td>1.000</td>
</tr>
<tr>
<td>Strongly Disagree</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Kolmogorov-Smirnov test for distributional fit

Differences between uniform distribution and question data distributions:

- Differences between uniform distribution: 0.467, 0.117, 0.133, 0.210
- Differences between question data: 0.600, 0.233, 0.400, 0.317
- Differences between distribution: 0.400, 0.267, 0.233, 0.517

Maximum Differences: 0.600, 0.267, 0.433, 0.517

Critical Difference = 0.375 @ alpha=.05

Some written comments supported the above preferences, and added additional points for consideration (these comments resulted from questions 12 and 13). Several respondents indicated their preference and support for the collaboration possibilities provided by the software. “It helps to... have... software to overcome the hassles of collaborative writing between two partners not in [the] same room/ state/ country, etc.” “The worst condition... was when my partner was in another room and had no typing allowed.” “The software... saved time.” “It makes writing articles much easier when two people cooperate.” “Great concept... divides the time of writing in half or better... and being able to proof each other’s work... is great.” These types of responses may have been due in part to a Hawthorne effect, wherein subjects exaggerated their positive feelings about the software because the package was the emphasis of the experiment and/or because of the newness of the package. Several respondents, however, had
negative comments about the software. “Hard to go into partner’s text and edit.” “If someone points out something with their pointer it’s in the way of... trying to correct the mistake.” “...slow...” “Biggest problem [was] not being able to use the arrow keys... having to use the mouse... was difficult...” “[A problem was] that we both couldn’t work on the same paragraph...” Several users mentioned these last three points. One respondent suggested the addition of a “text locking” feature to lock out changes by other participants.
CHAPTER 4 — DISCUSSION

Certain results obtained from this study are in agreement with previous research, while other results provide contrasting information to previous data. One main conclusion that can be gained from this study is that it is practical to perform objective research on groupware issues; also, it is practical to use writing tasks as instruments of this research. Particularly, the “Clue” writing task may be useful for further groupware studies, certainly for further group writing and co-authoring studies. Clearly, the use of a within-subject design was critical in dealing with the variability levels associated with writing and with group work, as predicted by many of the cited CSCW studies (Smith et al., 1991; Neale et al., 1992).

As there has been little previous research in the area of groupware, and specifically co-authoring, few data exist with which to compare the results of this study. The non-significant differences in task time means of the GW effect (significance hypothesized to have been subverted by high variability levels) suggest that real improvements in efficiency may result from the inclusion of groupware in team work paradigms. Significant differences may have been shown with additional statistical power. Differences would have indicated that in conditions where teamwork is necessary and cannot occur face-to-face, GW may alleviate losses in efficiency resulting from standard audio or audio and video methods of collaboration. Furthermore, the comparison of variability rates along with the difference in mean times for the groupware and non-groupware conditions in the non-face-to-face regimes may indicate that groupware can subvert communications barriers resulting from distance (a very important avenue of research for disabled persons interested in surrogate travel). Certain with-GW variability rates are significantly lower than these without-GW variabilities as shown in Table 8.
There are several possible explanations for the reduction in variability of the time data associated with the use of the groupware. Firstly, it is possible that groupware forces team members into a more structured mode of work. It could be that the specter of a group member’s eavesdropping on another group member’s writing process forces that member to proceed at a more measured pace. Another possible explanation for the increased similarities in time to complete the articles for trials where teams had the use of groupware relates to individual differences in writing productivity. It is possible that the groupware created disjoint group conditions (see Vaske and Grantham, 1990) and that performance was determined by the best member of the group. In this case, when the groupware was used, the best member of the group would complete a larger share of the work, giving a similar time to complete in various trials. When the groupware was not available, the lead typist would complete the larger share of work, and the lead typist, chosen randomly at the start of the experiment, may or may not have been the best member of the group.

All of the measures of quality used in this study resulted in support for an assumption of similar quality levels of all documents produced in the study. The lack of a grammatical/style decrement throughout this experiment provides reassurance that extensive training in group work strategies is not necessary for the introduction of groupware of this type to the work place. With the minimum of training provided for in this experiment, subjects were able to complete their fairly complex writing tasks without altering their grammatical writing levels. The lack of significant differences between conditions in number of answers reported correct and number of answers in chronological order also support the assumption of similar content quality throughout the articles. The significant differences in the Trial effect (these analyses are shown in Appendix VII) are interpreted to mean little in regards to the factors of interest. It is possible that these
differences relate to the material in the particular scenes shown; however, this would
contradict the results from the pretests of the question sets.

The clear result for the factor of communication media both supports and refutes
previous research in this area. Although Gale et al. (1991) reported that video was useful
to a degree in supporting cooperative work, the objective results from this study indicate
no differences between audio alone and audio plus video across various conditions. This
result does agree with some of the early communications research (in particular
Champness and Reid, 1970; Weeks and Chapanis, 1976). The difference shown between
these remote conditions and the face-to-face conditions does disagree with some of this
earlier research, however (Chapanis et al., 1977; Short et al., 1976). There are two
possible explanations for this difference in result of face-to-face communication from
previous research. One possible explanation relates to the fact that groups had exposure
to groupware in practice prior to experimental trials; it is possible this practice may have
given the groups opportunity to create strategies for themselves that decreased task times
in communication rich conditions. Another possible explanation concerns the
experimental task. It could be argued that the types of discussion involved in the
experimental task increase task time significantly in non-face-to-face conditions. The
experimental task environment may also be the cause of differences between results
concerning video in this study and previous studies.

One very interesting result from this study concerns the relative benefit of GW as
opposed to video communication. Even though the GW did not show significant time
improvements, the reduction in variability shown due to the GW was a significant effect,
whereas the video communication showed no effects. Given the hypothesis that less
variability throughout the study would have shown time improvements with the GW, it is
possible that GW may be a useful substitute for face-to-face communication when video
has not shown itself to be such a substitute, given the larger differences between means for various non-face-to-face experimental conditions.

The clear decline in the times reported for various presentation orders indicates a practice effect; it is unlikely that particular scenes degenerated in difficulty in that way by chance. Thus the use of a balanced Latin square or some other method of balancing seems critical in within-subject studies of this type.

The very low P-value associated with Team Differences shows both the high levels of variability involved in these types of studies, as well as a fertile area of research in CSCW. One possible explanation can be offered for these large between-team differences. Even though quality levels were comparable throughout the experiment, it could not be verified that they were exactly the same for all groups, particularly in the area of writing style issues. It could be argued that some groups were much more meticulous in their document generation, whereas others were more careless. These quality differences could easily go unnoticed by computerized grammar checkers, which are notably limited in their scope, and could surely have caused large, consistent time differences. It is also important to note that even with all the controls involved in this study, there was no way to control these large between-team differences; further studies must take into account the fact that there will always be these types of differences and large variabilities in co-authoring studies.

Results from the exit questionnaire tended to support the objective data recorded from the study. Positive opinions about the software in general support the differences between means shown by the interaction graph from Figure 4 (although these differences were not significant). The results from the general questions about various experimental conditions also support the significant result of quicker task completion in the face-to-face regime and the lack of benefit resulting from added video communication.
The Likert scale results from the "Easy to Use", "Essential to Work" and "Detrimental to Work" feature opinion questions show that the features were deemed easy to use and not detrimental in general, although there was no clear requirement for any specific feature. The difference in "Essential" feature opinions for the "Chat Box" may have been due to the fact that there was always an audio channel available for communication; textual communication was then superfluous. The comments section did bring out three clear problems with the software: sluggishness (this can be traced to the use of computers with slow processors; in fact one respondent suggested the restriction of use of the software to machines that employ the M68030 processor or better); lack of support for keyboard cursor movement; and the one editor per paragraph restriction.

**Conclusions**

The simple fact that objective research data were collected and were in some cases significant in this study is a landmark result for CSCW, specifically groupware and co-authoring research. The use of a within-subject design and many procedural techniques for control of variability enabled significant discussion of the effects of various media combinations on group work. Although no significant results were found for the groupware factor per se, the interaction plot shown in Figure 4 tantalizes with the suggestion of real productivity improvements created by the use of groupware, and certainly significant results were found for groupware's reduction of variability in certain conditions. Perhaps with additional statistical power (a larger sample size would achieve this) truly significant differences between the GW and no-GW conditions would have resulted.

As mentioned previously in the discussion section, the results showed non-significant differences between mean times of groupware and non-groupware variants of audio only and audio plus video conditions. This study and previous research have
shown that video communication is not a substitute in most cases for face-to-face communication; it is hoped that GW could become this surrogate, but more research will have to be done to confirm this. The significantly reduced variability levels shown by certain groupware conditions were an indicator of more similar group performance in GW conditions than conditions were only one person documents the workings of a group. This alone is an interesting result that should be further investigated.

The results from the questionnaires provide support that users have positive opinions about the software and believe that it increases productivity. Perhaps with some improvement and evolution, this software or similar packages may become accepted by the user community.

This author hopes that other research groups can use the "Clue" task developed, along with increased computer and financial resources, to collect objective data with which to begin to pick apart the unruly questions associated with group work and groupware. This task could be used to compare various co-authoring packages (should they become available), as well as training techniques and other communication media variants. The relative ease of use and inexpensiveness of the "Clue" task enable researchers to answer complex questions relating to group writing and writing in general that may have been previously unapproachable. Additionally, other tasks similar to this task could be developed to improve the variability reductions and perhaps begin to assess individual differences in group writing.

In this new area of research, co-authoring through groupware, in conjunction with a topic that has always been anathema to objective research, writing, the addition of an experimental task and some small preliminary objective data can go a long way towards stimulating new results to help us understand was has been up to now a difficult area.
REFERENCES


APPENDICES

Appendix I. Task Pretest Fact Sheets.

Fact Sheet 1:

Aspects™ developed by Group Technologies of Arlington, VA.

Works with any Macintosh™ personal computers that are linked together.

Multiple persons can type and edit a text document simultaneously.

Only one person at a time may manipulate a given paragraph.

Has a text communication facility.

Is relatively inexpensive.

Can be slow.
Fact Sheet 2:

Aspects™ developed by Group Technologies of Arlington, VA.

Allows multiple persons to manipulate a drawing.

Can work over a computer network or by modem.

Has limited features.

Can handle multiple conferences and network zones.

Allows enforcement of a given view for subsets of conferees.

Can work by itself.
Fact Sheet 3:

Aspects™ developed by Group Technologies of Arlington, VA.

Allows multiple persons to manipulate a painting.

Can work over a modem or over a network.

Can work with files in popular formats.

Allows “telepointing”; participants can see each other point at portions of a document.

Has been used in educational, scientific, and business environments.

May sometimes have trouble keeping different versions of a document held by different conferees the same.
Appendix II. Questions For the Video Clips.

Practice Question Set (1)

_Hall_

What weapon is on the wall?

Who has the candlestick?

Why is Col. Mustard in Paris?

Who has the poison?

Who has the gun initially?
Question Set 1

Dining Room:
What weapon does Mr. Green have?
Who drugged the tea?
How much poison must Inspector Pry eat?
Who tries to clean off their clothing?
Who talks about an airplane?

Billiard Room:
What does Madame Rose leave on the table?
Who is hiding under the table?
What is suggested Inspector Pry do with the money?
Who has the knife?
Who asks Inspector Pry to accept a bribe?

Hall:
Who takes the money out of Inspector Pry’s mouth?
Who says “Shoot”?
Who claims to be Inspector Pry?
Who leans on a shoulder?
Who suggests the assailant is a busboy?
Question Set 2

_Hotel Room:_

Who has the knife?

What did Mr. Green do to cars?

Who is an electrical genius?

What is the name of the train the characters took?

What does Sgt. Gray throw on the floor?

Who uncovers a small table?

What kind of problem does Mrs. Peacock think the phones have?

Who suggests the characters need a telephone repairman?

_Billiard Room:_

What is the extra benefit of Plum's invention?

What is Prof. Plum developing?

Who has the knife?

What kind of fabric softener is used?

What is running short?

Who has the poison?

Who has the gun?
Question Set 3

Lounge:

Who helps Mrs. White order the food?
What is Miss Scarlet doing at the bar?
What book is Mrs. Peacock reading?
How many orders of fried rice are asked for?
What weapon does M. Brunette have?
Where does Miss Scarlet say her home country is?
What is the name of the Chinese restaurant employee?

Billiard Room:

What packing material is used in the box?
How long has Prof. Plum worked on his invention?
Who has a drink?
What weapon is Sgt. Gray given?
What weapon does Col. Mustard take from Sgt. Gray?
What is an advantage of Polyester fuses?
What weapon does Miss Peach have?
Where must the microfilm be taken?
Question Set 4

Hotel Room:

Who is hungry?
Who kills a fly?
Who drinks the liquor?
Who takes the rope?
By what cliché is the liquor referred to?

Lounge

Who takes something from a jacket pocket?
What weapon does Mrs. Peacock have?
Why is the microfilm hidden?
Who eavesdrops on Miss Scarlet?
Who takes the candlestick?

Hotel Room:

Whose hat is near Col. Mustard?
Why does Sgt. Gray approach Col. Mustard?
Who does Miss Peach sneak past in the lounge?

Hall:

Why does Miss Peach take the tapestry?
What does Mrs. White apologize for?
Question Set 5

_Hall:_

Who has the gun?

Where is Miss Scarlet taken?

Who takes the microfilm from Miss Peach?

Who puts something in the plant?

_Dining Room:_

Who has the poison?

Who is lying on the floor?

Why was the food order not correct?

Who could eat a horse?

Who puts the candlestick on the tray?

Who takes something from Madam Rose’s purse?

What does Miss Scarlet call Mr. Green?

What weapon does Mr. Green have?

Who has the knife?

What does Mrs. Peacock suggest the food might be?

Who takes a wig from someone?
Question Set 6

*Dining Room:*

What weapon was on the floor?
Who gorged themselves?
What did Mrs. White throw over her shoulder?
Who grabbed the candlestick?
Who hurt their finger?
Who started the fire?

*Lounge:*

Who’s drinking a beer at the bar?
Who stirred their drink?
What weapon was under the rope?
Who wants to open a gas station?
Who is unsure about surrendering their weapon?
Who was opposed to violence?
Who says, "Someone’s got in the building!"

*Hall:*

Who was wanted for extortion and murder?
What is the amount of the highest award on the wanted poster?
Appendix III. Participant’s Informed Consent.

PARTICIPANT’S INFORMED CONSENT FORM

This form constitutes informed consent by you to participate in this study. Please read it in its entirety and then sign on the next sheet.

Thank you for participating in this research. This study is being conducted by the Human-Computer Interaction Laboratory of the Human Factors Engineering Center at Virginia Tech. It addresses the need to determine how special software termed “groupware” affects writing performance under differing conditions.

If you choose to participate in this research you will be asked to observe several short video clips and write a series of short articles with the provided software on a personal computer. Additionally, you will be asked to complete several standardized tests. No physical activities other than handwriting and/or typing will be required. You will be required to collaborate during the experiment with another subject. You will be paid $5 per hour of participation in the study.

The experiment will be conducted in approximately seven 1-1.5 hour periods. After completing and signing this consent form and background questionnaire, you will be asked to complete several tests and trained in the use of the software. Then, you will be asked to view some printed experimental materials and watch some short video clips. Over the next several days, you will be asked to complete several trials of the experiment. You will be given detailed explanations of the tasks required. Throughout the experimental trials your voice and picture will be recorded on video tape. If at any time during the experiment you feel that you cannot participate further or that you need help understanding any portion of the task or the equipment, please inform the experimenter.

Data collected during the study will be confidential and will be treated with anonymity.

There are no known sources of discomfort in this study.

As a participant in the study, you are entitled to certain rights:

1) You may withdraw from the experiment at any time for any reason without forfeiting pay for time spent up until withdrawal (your pay will be prorated).

2) Any of the research team members will answer any questions that you may have, and you should not sign this consent form until you understand fully all of the terms involved. The members of the research team are:
Charles A. Green, Graduate Student (ph. 951 - 7414)
Dr. Robert C. Williges, Virginia Tech Faculty Member (ph. 231 - 6270)

Additional questions regarding your rights as a subject should be addressed to Dr. E.R. Stout, Chairman of the Institutional Review Board, 301 Burruss Hall (ph. 231-5281).

Your signature below indicates that you have read this document in its entirety, that your questions have been answered, and that you consent to participate in the study described.

Signature:___________________________ Date:_______
Printed name:_______________________
Address:____________________________
___________________________________
___________________________________
Social Security #:______________________
Appendix IV. Instructions.

Instructions

Training

Actors and weapons:

Take a few minutes to memorize the names and faces of each of these actors.
You will be required to correctly remember the name for each visage, including correct spelling.

[Written Exam]

Software:

Use the scroll bars to navigate the work area. You can either drag the slider with your mouse, click on the bar above or below the slider, or click on the arrows at the top or bottom of the scroll bar.

Use the mouse to select text. This text can be replaced or deleted by typing new text or pressing the delete key, respectively.

You may type text as with any typewriter. The delete key deletes the character to the left of the cursor, except when text is selected. Text will always insert itself where the cursor is. Press the return key to start a new paragraph.

Several editing commands are available from the Edit menu. The Undo command undoes the last action taken by you or your partner. The Cut command places selected text into a temporary storage area, while the Copy command copies selected text to the temporary storage area. By selecting “Show Clipboard” from the Windows menu you can view the contents of this temporary storage area. The Paste command places the contents of the temporary storage area at the location of the cursor or in place of selected text.

When the two of you are able to type simultaneously, you will not both be able to manipulate the same paragraph(s) of text. In conditions where you and your partner are
able to type simultaneously, a bar at the left side of your screen shows which of you is working on a particular paragraph or group of paragraphs. Black indicates what you are working on, gray indicates what your partner is working on. If your partner is working on a paragraph(s), you cannot type in or edit text of that paragraph(s). The software will "beep" to tell you when this is occurring.

In conditions where you and your partner are able to type simultaneously, you may use the "view with" options to force your partner’s view to correspond exactly to your view. You may use the "unlink view" option to turn this feature off. There is a small icon in the upper right hand corner of the screen that shows whether your view is linked or not. If your view is linked, this icon appears as an open eye. If your view is not linked, the icon appears as a closed eye.

When the two of you are able to type simultaneously, you may use the "chat box" to communicate by text message with your partner. To show the chat box window (to read or send messages) select "chat box" from the windows menu. To send a message, click on the chat box, type a message and press return. If you would like to alert your partner to an incoming message, click on the icon that has the shape of a bell in the chat box. You may move, close, and re-size the chat box window.

When the two of you are able to type simultaneously, this software has a facility whereby you can allow your partner to see where your mouse pointer is. At the bottom of the screen is a picture of a pointing hand. If you do not have this window, select "conference tools" from the windows menu. If you click on the pointing hand, you will be able to manipulate the text of your document or type, and your partner will be able to see your pointer. If you click on the pointing hand again, your partner will no longer be able to see your pointer.
Now I would like the two of you to collaborate on a two page essay entitled “The Advantages and Disadvantages of Summer Life in Blacksburg”. This is a practice writing task to help you become familiar with the software.

[Oral Exam]

Task Pretest

The two of you will be writing three short newspaper articles (one paragraph each) that reflect information listed in fact sheets that will be provided. The articles are about the software you will be using in this experiment. Please do not elaborate on the facts given; simply organize them to produce a short article. Only one of you will have access to a computer to type the articles on, but the two of you should cooperate in the creation of each article.

I would like you to write the articles in the journalistic “hard news” writing style you have learned about in your class. Try to write the best articles you can; the quality of these articles is of utmost importance. Also, try to complete them as quick as you can; you will be timed, and you must finish each article within at most one half hour. Do not give a title or format the articles. You can think of this as a situation wherein you are working for a large newspaper. Your editor tells you that it is very expensive for him to make changes to your articles. In addition, you have other work to do and need to finish this task as quickly as you can. Another employee will title and format the articles.

Trials (administered during practice)

Today, and for several days to come, you and your partner will be participating in an experiment. Please do not speak to your partner or any of the other participants about this experiment at any time until the full experiment has been completed. Each time you arrive here, you will watch a short video clip that contains the characters you have already become familiar with. You may take notes during this time, but you may not speak with your partner. After watching the video clip, the two of you will write a short
article about the clip that contains embedded in it the answers to questions that will be given to you after watching the video clip. You must put the information required in sequence in the article as you observed it in the video clip. You must also put the information relating to each room in a separate paragraph. If you do not know the answer to a question, you must guess. You may use notes taken during observation of the video clips to help you answer the questions. Each trial will have a different type of setup. In some trials, both of you will have computers, and you will both be able to type and edit the article at the same time. In other trials, only one of you will have access to a computer. This person will be the same for each such trial, and will be identified by the experimenter. In some trials, the two of you will be located in the same room, while in others, you will be in separate rooms and will communicate by way of a phone and/or closed circuit video. Note that the phones do not ring; so do not hang up the phone.

I would like you to write the articles in the journalistic “hard news” writing style you have learned about in your class. Try to write the best articles you can; the quality of these articles is of utmost importance. Also, try to complete them as quick as you can; you must finish each article within one hour. Do not give a title or format the articles. You can think of this as a situation wherein you are working for a large newspaper. Your editor tells you that it is very expensive for him to make changes to your articles. In addition, you have other work to do and need to finish this task as quickly as you can. Another employee will title and format the articles.

Your editor has also seen the video clips, and has provided the questions to you to enforce his decision as to what information needs to be in the articles. As for the different setups you will experience, you can imagine that one of you is an “on-scene” reporter, while the other of you is based at the home office. The home office person is the person who has access to a computer in those trials where only one computer is available. The newspaper you work for can only afford to bring the “on-site” person back
to the home office on some assignments. The availability of closed circuit video communication is also dependent on the vagaries of corporate management.

Note that there are only five weapons you will see in the video clips. Please write these in your notes. These weapons are: the rope, the poison, the gun, the knife, and the candlestick.
Appendix V. IRB Protocol.

An Empirical Study on the Effects of a Collaboration-Aware Computer System and Several Communication Media Alternatives on Product Quality and Time to Complete in a Co-authoring Environment

Introduction

The purpose of this study is to determine the effect of co-authoring software on the time to complete and grammatical quality of team-produced documents, in conjunction with the effects of face-to-face versus audio versus audio and video communication. In order to do this, a task was developed that would promote sufficiently complex document production to evaluate style and quality issues.

Subjects

Twelve subjects will be involved in this study in teams of two. Subjects will be selected from volunteers who have recently completed a journalistic writing course, in order that they can write in the constrained journalistic "hard news" style. These subjects will be pretested on two counts: writing ability as measured by the subtests from the Woodcock-Johnson Psycho-Educational Battery (1989-1990) and dominance as measured from a subtest of the Personality Research Form (Jackson, 1987) in order to match them to teams. Note that subjects will not be told their scores on these pretests to prevent any psychological effects resulting from that knowledge. Subjects will be trained in the use of a commercial groupware package (Aspects™ from Group Technologies of Arlington, VA) and in the task procedure. Subjects will then watch a series of video clips from the board game "Video Clue II"™ by Parker Brothers, and will write short newspaper articles about the clips. Two teams of subjects will write three additional articles about the groupware package used from fact sheets provided for task pretesting purposes. Subjects will be asked to sign an informed consent form, and will be notified that they will receive five dollars per hour for their time. 

82
Procedure

The procedure for this study following the pretests and training is straightforward, and is estimated from pilot testing to take approximately one hour. A dyad arrives at the lab (having agreed to participate, signed informed consent documents, trained, and pretested and matched into teams as discussed above) and is greeted by an experimenter. The subjects view a short video clip and then spend up to an hour creating an appropriate newspaper article that answers questions about the clip provided by the experimenter. Because these clips are adapted from the Parker Brothers’ Video Clue II™ table-top game which has been available for purchase by the general public, the material is assumed to be non-threatening and inoffensive, if not an example of exemplary film production. Detailed instructions given during training for the task and the software and for task pretesting are attached.

Treatments

There are to be six treatment conditions for this 2x3 within-subject study. These conditions relate to terms under which the written articles created by the subject(s) are produced. One factor is that of groupware and computer use; in some conditions, both subjects may type and edit their document simultaneously; in other conditions only one subject may have access to the document. The second factor is a media condition factor; the three levels are face-to-face, audio communication only, and audio plus video communication.

A different video clip and question set is provided for each trial and assignment given to subjects for each treatment condition. They are to write a newspaper article that answers 15 provided questions about the video clip. These answers are to be embedded in the document, and are to be addressed in temporal sequence as the events, objects, locations, and personalities they refer to were shown in the video clip. Note that the subject(s) will be told to guess answers to questions they do not know (in order to aid in
scoring of sequence). Subjects are told they may call in an experimenter at any time to help them understand the questions or work with the software; they are also told to alert the experimenter when they have completed their article.

Dependent Measures

Several forms of data will be collected from the study. Both video and audio records will be made of the subjects’ performances during the study, in addition to their written articles. Two dependent measures will be collected from this data. An aggregate time to complete the report will be measured; this is the key measure for the study. A computerized grammatical test will be run on the articles, including calculations providing a Flesch-Kincaid and a grade level index. It should be noted that possible further data analysis will be done on both the video and audio data, and the written reports.
Appendix VI. Exit Questionnaire.

Circle the "+" to indicate your agreement with the following statements:

+------------------------------------------+
| Strongly Agree  Agree  No Opinion  Disagree  Strongly Disagree |
+------------------------------------------+

1. The conditions where we could both type were more productive than the conditions where only one of us could type.

2. It really helped to be able to see my partner when we were not working in the same office.

3. We worked much better when we were working in the same office.

4. The feature that forced both of us to view the same portion of the document (linked view) was:
   a. Essential to my work.
   b. Easy to use.
   c. Detrimental; it should be removed.

5. The chat box was:
   a. Essential to my work.
   b. Easy to use.
   c. Detrimental; it should be removed.

6. The feature that enabled me to point out portions of the document (telepointer) was:
   a. Essential to my work.
   b. Easy to use.
   c. Detrimental; it should be removed.

7. The bar on the left of the screen that showed where each of us was working (ownership bar) was:
   a. Essential to my work.
   b. Easy to use.
   c. Detrimental; it should be removed.
8. It was distracting for both of us to be typing at the same time.

9a. Under which conditions (in the same office, talking on the phone, talking on the phone and using the closed-circuit video) was it easiest to work? Why?

9b. Under which conditions (in the same office, talking on the phone, talking on the phone and using the closed-circuit video) was it hardest to work? Why?

10a. In general, how did you feel about the software/computer system you were using when you could both type?

10b. In general, how did you feel about the software/computer system you were using when only one of you could type?

11. Rank each of the following features of the software in order of usefulness, and explain how you used them.
12. What were the biggest problems with the software, and why?

13. What other comments do you have about the software or the communication media?
Appendix VII. Additional Charts and Analyses.

**ANOVA Summary Table**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>G-G*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>5</td>
<td>14.81</td>
<td>2.96</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groupware</td>
<td>1</td>
<td>.25</td>
<td>.25</td>
<td>.08</td>
<td>.7905</td>
<td>.7905</td>
</tr>
<tr>
<td>Groupware * Subject</td>
<td>5</td>
<td>15.92</td>
<td>3.18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media Types</td>
<td>2</td>
<td>6.06</td>
<td>3.03</td>
<td>.43</td>
<td>.6640</td>
<td>.6522</td>
</tr>
<tr>
<td>Media Types * Subject</td>
<td>10</td>
<td>70.94</td>
<td>7.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groupware * Media Types</td>
<td>2</td>
<td>10.17</td>
<td>5.08</td>
<td>1.94</td>
<td>.1937</td>
<td>.1998</td>
</tr>
<tr>
<td>Groupware * Media Types * Subject</td>
<td>10</td>
<td>26.17</td>
<td>2.62</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: # Correct

**Table of Epsilon Factors for df Adjustment**

<table>
<thead>
<tr>
<th>Dependent: # Correct</th>
<th>G-G Epsilon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groupware</td>
<td>1.00</td>
</tr>
<tr>
<td>Media Types</td>
<td>.94</td>
</tr>
<tr>
<td>Groupware * Media Types</td>
<td>.90</td>
</tr>
</tbody>
</table>

G-G*: Greenhouse-Geisser corrected P-Value

# Correct ANOVA

**ANOVA Summary Table**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team</td>
<td>5</td>
<td>14.81</td>
<td>2.96</td>
<td>1.78</td>
<td>.1623</td>
</tr>
<tr>
<td>Trial</td>
<td>5</td>
<td>79.81</td>
<td>15.96</td>
<td>9.61</td>
<td>.0001</td>
</tr>
<tr>
<td>Condition</td>
<td>5</td>
<td>16.47</td>
<td>3.29</td>
<td>1.98</td>
<td>.1252</td>
</tr>
<tr>
<td>Residual</td>
<td>20</td>
<td>33.22</td>
<td>1.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: # Correct

# Correct L.S. ANOVA
Interaction Plot
Effect: Trial
Dependent: # Correct
With Standard Error error bars.

Graph of Trial Means for # Correct
ANOVA Summary Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>5</td>
<td>16.56</td>
<td>3.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groupware</td>
<td>1</td>
<td>1.78</td>
<td>1.78</td>
<td>.25</td>
<td>.6352</td>
</tr>
<tr>
<td>Groupware * Subject</td>
<td>5</td>
<td>34.89</td>
<td>6.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Media Types</td>
<td>2</td>
<td>2.39</td>
<td>1.19</td>
<td>.51</td>
<td>.6136</td>
</tr>
<tr>
<td>Media Types * Subject</td>
<td>10</td>
<td>23.28</td>
<td>2.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groupware * Media Types</td>
<td>2</td>
<td>.72</td>
<td>.36</td>
<td>.06</td>
<td>.9375</td>
</tr>
<tr>
<td>Groupware * Media Types * Subject</td>
<td>10</td>
<td>55.61</td>
<td>5.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: # in Order

Table of Epsilon Factors for df Adjustment

<table>
<thead>
<tr>
<th>Dependent: # in Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>G-G Epsilon</td>
</tr>
<tr>
<td>Groupware</td>
</tr>
<tr>
<td>Media Types</td>
</tr>
<tr>
<td>Groupware * Media Types</td>
</tr>
</tbody>
</table>

G-G*: Greenhouse-Geisser corrected P-Value

# in Order ANOVA

ANOVA Summary Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team</td>
<td>5</td>
<td>16.56</td>
<td>3.31</td>
<td>1.53</td>
<td>.2247</td>
</tr>
<tr>
<td>Trial</td>
<td>5</td>
<td>70.56</td>
<td>14.11</td>
<td>6.53</td>
<td>.0009</td>
</tr>
<tr>
<td>Condition</td>
<td>5</td>
<td>4.89</td>
<td>.98</td>
<td>.45</td>
<td>.8065</td>
</tr>
<tr>
<td>Residual</td>
<td>20</td>
<td>43.22</td>
<td>2.16</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: # in Order

# in Order L.S. ANOVA
Interaction Plot
Effect: Trial
Dependent: # in Order
With Standard Error error bars.

Graph of Trial Means for # in Order
Feature Rank Chart.

ANOVA Summary Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>G-G*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>11</td>
<td>3.47E-18</td>
<td>3.15E-19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>3</td>
<td>10.17</td>
<td>3.39</td>
<td>2.24</td>
<td>.015</td>
<td>.178</td>
</tr>
<tr>
<td>Features * Subject</td>
<td>33</td>
<td>49.83</td>
<td>1.51</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: Rankings

Table of Epsilon Factors for df Adjustment

Dependent: Rankings

G-G Epsilon

Features .80

*G-G: Greenhouse-Geisser corrected P-Value

ANOVA of Feature Rankings.
ANOVA Summary Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>G-G*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>11</td>
<td>25.25</td>
<td>2.30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>3</td>
<td>2.92</td>
<td>.97</td>
<td>2.45</td>
<td>.0807</td>
<td>.1087</td>
</tr>
<tr>
<td>Features * Subject</td>
<td>33</td>
<td>13.08</td>
<td>.40</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: Detrimental

Table of Epsilon Factors for df Adjustment
Dependent: Detrimental

<table>
<thead>
<tr>
<th>G-G Epsilon</th>
</tr>
</thead>
<tbody>
<tr>
<td>.67</td>
</tr>
</tbody>
</table>

*G-G: Greenhouse-Geisser corrected P-Value

ANOVA of “Detrimental” Feature Opinions.

ANOVA Summary Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Value</th>
<th>P-Value</th>
<th>G-G*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject</td>
<td>11</td>
<td>20.67</td>
<td>1.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Features</td>
<td>3</td>
<td>2.83</td>
<td>.94</td>
<td>1.63</td>
<td>.2021</td>
<td>.2128</td>
</tr>
<tr>
<td>Features * Subject</td>
<td>33</td>
<td>19.17</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Dependent: Easy to Use?

Table of Epsilon Factors for df Adjustment
Dependent: Easy to Use?

<table>
<thead>
<tr>
<th>G-G Epsilon</th>
</tr>
</thead>
<tbody>
<tr>
<td>.80</td>
</tr>
</tbody>
</table>

*G-G: Greenhouse-Geisser corrected P-Value

ANOVA of “Easy to Use?” Feature Opinions.
VITA

Charles A. Green is a masters student in Industrial Engineering at Virginia Polytechnic Institute and State University. He received a B.S. in Mechanical Engineering from the University of Illinois at Urbana-Champaign in 1990. His research interests include consumer product design, human-computer interaction, and groupware.

Charles A. Green