INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road. Ann Arbor MI 48106-1346 USA
313/761-4700  800/521-0600
The software inspection: The task and mechanisms for group support

Tyran, Craig Kenneth, Ph.D.
The University of Arizona, 1993

Copyright ©1993 by Tyran, Craig Kenneth. All rights reserved.
THE SOFTWARE INSPECTION:
THE TASK AND MECHANISMS FOR GROUP SUPPORT

by

Craig Kenneth Tyran

Copyright © Craig Kenneth Tyran 1993

A Dissertation Submitted to the Faculty of the

COMMITTEE ON BUSINESS ADMINISTRATION

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 9 3
As members of the Final Examination Committee, we certify that we have read the document prepared by Craig Kenneth Tyran entitled The Software Inspection: The Task and Mechanisms for Group Support and recommend that it be accepted as fulfilling the requirements for the Degree of Doctor of Philosophy.

Dr. Joey George  
Date 4/28/93

Dr. Jay Nunamaker  
Date 4/28/93

Dr. Doug Vogel  
Date 4/28/93

Final approval and acceptance of this document is contingent upon the candidate's submission of the final copy of the document to the Graduate College.

I hereby certify that I have read this document prepared under my direction and recommend that it be accepted as fulfilling the requirement.

Dr. Joey George  
Date 4/28/93
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

Signed: [Signature]
ACKNOWLEDGEMENTS

I would like to acknowledge the financial support of IBM Corporation and the American Assembly of Collegiate Schools of Business, which helped to support my doctoral studies at the University of Arizona.

Additionally, I acknowledge the following individuals:

Dr. Joey George, my committee chair, who has been an excellent advisor and a wonderful friend. It has been a true pleasure to be associated with him.

Dr. Jay Nunamaker, former head of the MIS department and current director of the Center for Management Information, who has always supported my endeavors at Arizona. Some of my fondest memories of Arizona include our high stakes golf matches on the golf courses of Tucson.

Dr. Ram Nidumolu and Dr. Doug Vogel, who have offered support and encouragement.


The administrative staff in the MIS department who have been very helpful: Betty Albert, Lanna Greffet, and Joan Seidel.

My sister, Dr. Cindy Tyran, and my brother, Garry Tyran, who have each helped to "break the trail" for me by setting a standard of excellence in character and achievement.

And lastly, my parents, Ben and Jeanne Tyran, who have always supported me in every way imaginable. I am very lucky to have parents whom I love so much.
DEDICATION

This dissertation is dedicated to my mother,

Jeanne Tyran,

who has been with me every step of the way.
# TABLE OF CONTENTS

**LIST OF ILLUSTRATIONS** ........................................ 14

**LIST OF TABLES** ................................................... 15

**ABSTRACT** .......................................................... 17

**CHAPTER ONE: INTRODUCTION** .................................. 19

1.1 Software reviews: Inspections and walkthroughs ............. 21
1.2 Software reviews: A historical perspective .................. 22
1.3 Software Reviews in Industry .................................. 23
1.4 Group support for software review teams .................... 26
1.5 Research objectives and method ................................ 27
1.6 Overview of dissertation ....................................... 28

**CHAPTER 2: LITERATURE REVIEW: SOFTWARE REVIEWS** ....... 29

2.1 Research studies: The effectiveness and benefits of software reviews ........................................ 30
2.1.1 Field studies: Design and code inspections ............. 30
2.1.2 Field study: User requirements specification inspections ... 33
2.1.3 Semi-controlled studies: User requirements specification inspections .......................................................... 34
2.1.4 Educational benefits of the software review process ......... 35
2.1.5 Semi-controlled study: The educational benefit of programming reviews .................................................. 36
2.1.6 Summary .................................................................................................................................................. 37

2.2 Research studies: Investigation of variations to the software review process ................................................................. 38
2.2.1 Field study: Comparison of inspection and walkthrough techniques ........................................................................ 39
2.2.2 Semi-controlled study: Team review versus individual computer-based testing .................................................. 39
2.2.3 Semi-controlled study: Two-person inspection technique ...... 41
2.2.4 Semi-controlled studies: N-fold inspection technique ....... 43
2.2.5 Field study: Relationship between inspection performance and procedures .................................................. 45
2.2.6 Summary .................................................................................................................................................. 47

2.3 Future directions for software inspection research .................. 48

CHAPTER 3: LITERATURE REVIEW: GROUP RESEARCH ........................................... 52
3.1 Group productivity ........................................................................................................................................... 52
3.2 Task characteristics ........................................ 54
  3.2.1 Disjunctive tasks ...................................... 55
  3.2.2 Complementary tasks .................................. 57
3.3 Process ..................................................... 57
  3.3.1 Potential sources of process loss ....................... 59
  3.3.2 Potential sources of process gain ....................... 60
3.4 Support for groups .......................................... 61

CHAPTER 4: RESEARCH QUESTIONS AND HYPOTHESES ............... 64
4.1 Theoretical group support mechanisms .......................... 64
  4.1.1 Process support ........................................ 65
  4.1.2 Task support ........................................... 68
  4.1.3 Process structure ....................................... 69
  4.1.4 Task structure .......................................... 69
4.2 Variables and measures ...................................... 70
  4.2.1 Independent variables .................................. 70
  4.2.2 Dependent variables ................................... 73
4.3 Research questions and hypotheses ............................ 76
  4.3.1 Experiment 1 .......................................... 77
  4.3.2 Experiment 2 .......................................... 80
4.4 Summary .................................................... 86
CHAPTER 5: EXPERIMENTAL STUDY 1: METHOD AND RESULTS . . . 87

5.1 Method ............................................. 87
  5.1.1 Task ........................................... 87
  5.1.2 Subjects ...................................... 89
  5.1.3 Procedure .................................. 90
  5.1.4 Dependent variables ......................... 96

5.2 Results ............................................. 97
  5.2.1 Group resources .............................. 97
  5.2.2 The characteristics of the task ............... 101
  5.2.3 Regression analysis .......................... 104
  5.2.4 Group process ................................ 106
  5.2.5 Perceptions of the group process .............. 114
  5.2.6 Intangible benefits .......................... 115

CHAPTER 6: EXPERIMENTAL STUDY 2: METHOD AND RESULTS .... 118

6.1 Method ............................................. 118
  6.1.1 Task ........................................... 118
  6.1.2 Subjects ...................................... 119
  6.1.3 Procedure .................................. 121
  6.1.4 Dependent variables ......................... 132
6.2 Results ....................................... 133

6.2.1 Defect identification session: Performance and perceptions . 133

6.2.1.1 Confound checks ..................... 133
6.2.1.2 Group performance and perceptions .......... 139

6.2.2 Comparison of performance: Individual versus interactive
defect identification ........................................ 147

6.2.3 Defect evaluation session: Performance and perceptions . 148

6.2.3.1 Confound checks ..................... 148
6.2.3.2 Group performance and perceptions .......... 152

6.3 Summary ..................................... 158

CHAPTER 7: DISCUSSION AND CONCLUSIONS ....................... 159

7.1 Discussion ..................................... 159

7.1.1 Experiment 1 .................................. 159

7.1.1.1 Nature of the task ..................... 159
7.1.1.2 Variation in individual and group resources . . 161
7.1.1.3 Correlation of group resources and process to

performance ................................................. 162
7.1.1.4 Impacts of task support ................. 162
7.1.1.5 Perceptions of satisfaction and acceptance .... 164
7.1.1.6 Perceptions of understanding and confidence .... 164
7.1.2 Experiment 2 ........................................... 165
  7.1.2.1 Impacts of process support .................. 165
  7.1.2.2 Impacts of task support .................... 171
7.2 Implications ........................................... 175
  7.2.1 Implications for practice .................... 175
  7.2.2 Implications for research ..................... 177
7.3 Limitations ........................................... 181
7.4 Concluding remarks ................................... 183

APPENDIX A: EXPERIMENTAL INSPECTION TASK: SYSTEM SPECIFICATION ........................................... 185

APPENDIX B: EXPERIMENTAL INSPECTION TASK: LIST OF DEFECTS ........................................... 193

APPENDIX C: OVERVIEW HANDOUT FOR INSPECTION EXERCISE ........................................... 198

APPENDIX D: EXPERIMENT 1: DIRECTIONS FOR BASELINE TREATMENT ........................................... 205
LIST OF ILLUSTRATIONS

Figure 2.1: Fishbone diagram of contributors to inspection quality ........... 49
LIST OF TABLES

Table 4.1: Group support mechanisms and impacts on process loss and gain . . . 66
Table 5.1: Group resources ........................................ 98
Table 5.2: Variations in group resources ................................ 98
Table 5.3: Group acceptance rate of valid defects ................................. 102
Table 5.4: Hierarchical regression of group resources and group process on group performance .................... 105
Table 5.5: Ratios of group potential performance, loss, and gain to actual performance . ........................................ 108
Table 5.6: Perceptions of potential sources of group process loss .................. 112
Table 5.7: Perceptions of the group process ........................................ 115
Table 5.8: Perceived intangible impacts of the group process ........................ 117
Table 6.1: Pre-exercise perceptions ........................................ 134
Table 6.2: Perceptions of the identification session: Confound checks ........ 136
Table 6.3: Number of defects detected ........................................ 140
Table 6.4: Perceptions of potential sources of group process loss and gain for the identification session .................. 141
Table 6.5: Perceptions of the group process for the identification session .......... 146
Table 6.6: Perceptions of the evaluation session: Confound checks ................ 150
Table 6.7: Ratios of number of post-identification session defects, loss, and gain to number of post-evaluation session defects .... 154

Table 6.8: Perceptions of potential sources of group process loss and gain for the evaluation session .... 156

Table 6.9: Perceptions of the group process for the evaluation session .... 157

Table 7.1: Experiment 1: Summary of findings for each research question .... 160

Table 7.2: Experiment 1: Summary of findings for each hypothesis .... 163

Table 7.3: Experiment 2: Summary of findings for each hypothesis .... 166
ABSTRACT

Group software reviews play an important role in the software development process and may potentially benefit from group support. The purpose of this experimental study was to learn more about the nature of the software review task and to assess the impacts of selected group support mechanisms on review team performance. Two experiments were conducted. For each experiment, a review task involving the inspection of an analysis specification document (i.e., narrative, data flow diagrams, data dictionary) was performed by teams of undergraduate college student subjects. A review procedure based on Fagan's (1976) inspection technique was followed. To assess the impacts of group support on different aspects of the review process, the review task was split into two subtasks: defect identification and defect evaluation. Each subtask was investigated separately. The independent variable for the study was type of group support. The dependent variables were group performance and perceptions of the group process (e.g., sources of loss and gain). Group support was manipulated across treatments using non-electronic and electronic approaches. Electronic group support was provided by using an electronic meeting system. The experimental task was found to be a complementary task comprised of Eureka-type disjunctive subtasks. For the defect identification subtask, review teams receiving electronic process support (consisting of parallel communication and group memory) found significantly more defects than unsupported teams. The better performance of the electronically-supported teams was attributed to decreased process
loss associated with domination and side-tracking. Satisfaction and acceptance of findings were rated higher for the unsupported teams compared to the electronically-supported teams. For the defect evaluation subtask, review teams receiving task support (consisting of a display of findings from the identification subtask) experienced significantly less process loss than unsupported teams. This outcome was attributed to the fact that the task support mechanism ensured that all findings from the identification session were raised to the review teams for evaluation. There were no differences with respect to satisfaction and acceptance of the findings. The implications of the study for researchers and practitioners are discussed.
CHAPTER ONE: INTRODUCTION

Software plays a major role in modern organizations, as it is used to run the computer-based machines which collect, store, transform, and organize the data used to conduct business. Unfortunately, the development of software is a major headache for many modern organizations. Frequently software is delivered late and does not meet user requirements due to defects. Problems with software are exacerbated by the fact that software is a large -- and growing -- component of operations and costs for organizations. For example, it has been estimated that software costs accounted for 10 percent of the United States defense budget in 1990 (Dunn, 1990). Since the cost of data processing in organizations has been quadrupling every ten years, it is likely that software problems will continue to grow unless measures are taken to improve the process of software development.

Many software problems may be attributed to the development of low quality software that is characterized by numerous defects. As suggested by Schulmeyer (1990), the way to address these types of software problems is to improve software quality through quality assurance methods, since such improvements will reduce defects and thus speed up development time (due to less rework required prior to release) and decrease the costs.

---

1 A software "defect" may be defined as anything which may cause or lead to the failure to satisfy a software requirement (Fagan, 1986).
and personnel demands associated with program maintenance (due to fewer defects in the released software).

Software defects may be reduced through two types of quality assurance methods: defect prevention or defect detection. Examples of preventative methods include formal approaches such as proof of correctness (Dijkstra, 1976), statistical process control (Schulmeyer, 1990), and cleanroom engineering (Mills, et al., 1987). While detection methods are desirable, such approaches typically need to be supplemented by detection methods (Fagan, 1986). At one time, defects were primarily detected through software testing at the end of the development life cycle and by the user following software release (Myers, 1979). However, in recent years software review methods such as inspections and walkthroughs have been used to detect defects earlier in the software development process. Reports from industry indicate that software review methods are a popular and effective way to improve software quality (Fagan, 1976).

This dissertation is concerned with learning more about how to improve the performance of software review teams. The first section of the chapter describes the most common software review techniques. The second and third sections discuss software reviews with regard to their historical context and their role in industry. The fourth section addresses the notion of group support for software review teams. The fifth section summarizes the research objectives and method of this dissertation. Finally, an overview of the
1.1 Software reviews: Inspections and walkthroughs

For the purposes of this dissertation, a "software review" is defined as a group meeting which is conducted to uncover defects in a software work product (e.g., requirements specification, design document, code, test plan).² Two of the most common review techniques are the software "inspection" and the "structured walkthrough" (Freedman and Weinberg, 1982). The software inspection approach is a formally defined process involving a series of well defined inspection steps and roles, a checklist to aid error detection, the formal collection of process and product data, and exit criteria for work products (Fagan, 1976). By contrast, the structured walkthrough approach is a less formal and less comprehensive technique which does not adhere as rigidly to a standard format (Yourdon, 1989).

Despite the differences between these techniques they share the same specific objective for the group review session: error detection. To maintain a focus on error detection, groups using either technique are instructed to avoid error correction. When potential errors are brought to the attention of the group they are discussed up to the point that the

² The definition of a "software review" used for this dissertation follows from the IEEE standard terminology (IEEE, 1989).
group determines that an error has been identified. At the end of the group session, a list of the defects that have been identified is provided to the author of the work product (who is responsible for correcting the defects). The inspection and walkthrough techniques also share other common aspects. Each technique prescribes a preparation step in which review participants become familiar with the work product under consideration on their own prior to the review. Also, proponents of both types of review techniques recommend that review teams should consist of three to six people and that review sessions should take no more than two hours (Fagan, 1976; Yourdon, 1989).

1.2 Software reviews: A historical perspective

Weinberg and Freedman (1984) have suggested that the concept of software reviews has been around as long as software itself, as early software developers were known to share their work informally with others to help improve the quality of their work. When software projects became more complex in the 1960's, the need for more sophisticated review procedures became apparent and more formalized review procedures evolved. In the 1970's, articles focusing on team software reviews first began to appear in the literature (e.g., Fagan, 1976; Myers, 1978). In the past fifteen years, software reviews have become a popular software development activity. Recent industry surveys indicate that reviews have gained wide acceptance as a development tactic and are performed by
a majority of organizations (Carey and McLeod, 1988; Necco, et al., 1987).\(^3\) In addition to being used on a stand-alone basis, software reviews are also an important component of systems development methodologies such as Joint Application Development (JAD) (e.g., August, 1991).

### 1.3 Software Reviews in Industry

Software reviews are popular in industry because they have been found to be an effective way to uncover defects existing in a variety of software work products including: requirements specifications (Doolan, 1992; Martin and Tsai, 1990), design documents (Ackerman, et al., 1984), user interface designs (Bias, 1991), program code (Fagan, 1986; Fowler, 1986), and test plans (Graden, et al., 1986). The detection rate for a review varies depending on the type of work product being inspected (e.g., requirements specification, design document, code) and the type of review method used. However, studies have found that 30 to 90 percent of the defects in a work product may be uncovered through reviews (Fagan, 1986; Myers, 1978; Schneider, et al., 1992). The favorable attitude that many in the software development field have toward the review process is epitomized by a statement made by Barry Boehm (a well known expert in the field of systems development), who has written that "the structured walkthrough has been

\(^3\) In the Necco, et al. (1987) survey, 90 percent of the respondents indicated that their organization performed software reviews, while 65 percent of the respondents in the Carey and McLeod (1988) survey reported that their organization performed reviews.
the most cost effective technique to date for eliminating software errors." (p. 85, Boehm, 1987).

While it appears that most software development organizations perform software reviews, it is not clear what specific types of software review procedures are being used in the field. As discussed above, there are strong distinctions in the literature regarding the differences between the inspection and walkthrough techniques. However, an in-depth interview survey by Zelkowitz, et al. (1984) found that the ways that software reviews are conducted in the field vary greatly. For example, depending on the organization, the terms "inspection" and "structured walkthrough" may take on meanings which are different that those adopted in the literature. Given this situation, it is quite likely that organizations are not following the review procedures recommended by the literature and may thus be using a procedure that does not utilize the full potential of the software review concept.

Reviews can take up a significant portion of a project's time and budget if performed on a consistent basis throughout the life of a project. According to one industry estimate, project teams performing Fagan's inspection technique may devote four to fifteen percent of project time to the review process (Ackerman, et al., 1989). Assuming that inspections are performed regularly throughout the development cycle, it has been estimated that the costs for inspections typically amount to fifteen percent of project costs
While these costs may seem high, they are justified by proponents of inspections based on improved worker productivity and software quality, as well as intangible benefits.

Productivity may be improved through reviews since the correction of a defect early in the software development cycle can be 10 to 100 times less expensive than rework performed at the system testing stage (Doolan, 1992; Fagan, 1976; Pressman, 1992). Given this multiplier effect it makes sense to conduct reviews prior to system testing. The quality of the released software has also been found to improve due to reviews. For example, after the introduction of inspections at IBM in 1974, the systems developers reported that the number of defects found in their released code for their System/370 software products was reduced by two-thirds (Fagan, 1986). In another case, inspections of requirements specifications were estimated to greatly reduce the number of "fault reports" that would have been logged by users of the system (Doolan, 1992). Lastly, several authors have suggested that there are intangible benefits associated with reviews including education, development of backup knowledge, and team building (e.g., Boehm, 1987; Fagan, 1986; Yourdon, 1989).

As noted earlier, computer-based testing has historically been the primary means of identifying software defects. However, recent industry reports suggest that inspections are a cheaper way to detect and fix software defects (Ackerman, et al., 1989). While
such findings might suggest that software testing should be replaced by reviews, findings from a laboratory study indicate that testing may be more appropriate than reviews for finding certain types of defects (and vice versa) (Myers, 1978). Hence, the prevailing wisdom is that reviews and testing are complementary and that both techniques should be used to promote optimal software quality (Myers, 1979).

1.4 Group support for software review teams

Despite the fact that a large body of research exists in the field of small group behavior, no rigorous studies have examined the software review (or any type of group review task) from a group perspective. Since group researchers who have examined other types of tasks have found that unsupported groups typically do not reach their productive potential, it is likely that the traditional (i.e., unsupported) approach to software reviews may not be optimal, as the lessons learned by group researchers have yet to be applied to software reviews.

One of the newer approaches to group support which may have the potential to support software review teams involves the use of electronic meeting systems (EMS) technology. An EMS is a computer-supported environment designed to make group meetings more productive (Dennis, et al., 1988). In recent years, researchers have examined the use of EMS technology for a variety of group applications including systems development
(e.g., Carmel, et al., 1992; Wanniger and Dickson, 1992). However, the applicability of EMS -- as well as non-electronic support techniques -- to the software review task has not yet been explored.

1.5 Research objectives and method

Given the important role of the software review to software development, it is worth considering how the productivity of review teams may be improved through support techniques. In an effort to improve the practice of software reviews, this dissertation had two primary objectives: 1) Learn more about the characteristics of the software review task so that the opportunities for group support could be better assessed, and 2) Investigate the impact of selected group support mechanisms on the performance and perceptions of review teams. An experimental research method involving the inspection of a user requirements document by undergraduate student subjects was used to address the research objectives. An inspection technique based on Fagan’s (1976) guidelines was employed. The key outcome measures for the study were individual and team inspection performance (i.e., the number of defects detected), perceptual outcomes (e.g., satisfaction), and perceptions of group process issues (e.g., blocking, domination, evaluation apprehension).
1.6 Overview of dissertation

This dissertation is composed of seven chapters. Chapters Two and Three review the literature relating to software reviews and group research, respectively. Chapter Four introduces the research questions and hypotheses addressed by this study. Chapter Five presents the method and results for the first experimental study that was undertaken as part of the dissertation. The first study was exploratory in nature and was conducted to learn more about the inspection task and to assess the impact of a simple task support mechanism on inspection team productivity. Chapter Six presents the method and results for a second experimental study that was performed. The objective of the second study was to investigate the impact of process and task support mechanisms on inspection team productivity. Chapter Seven discusses the study results and the implications of the study.
CHAPTER 2: LITERATURE REVIEW: SOFTWARE REVIEWS

Although software reviews (i.e., inspections and walkthroughs) are considered a particularly important activity for software development (Bohem, 1987) and are widely implemented in industry (Necco, et al., 1987), there has been relatively little research conducted in this area. The research that does exist is fragmented and is varied with regard to the type of review technique studied, the research objective, and the research setting and subjects. The purpose of this chapter is threefold: survey the existing research on software reviews, summarize what has been learned from the existing studies, and identify promising directions for software review research.

The formal inspection approach described by Fagan (1976) is the software review technique that has been investigated most frequently. However, researchers have also performed studies which have utilized a more informal review process (e.g., the walkthrough). Research studies concerned with software reviews can be classified broadly into two categories: 1) Research reporting on the effectiveness and benefits of software reviews, and 2) Research aimed at assessing the effectiveness of variations to Fagan's (1976) inspection approach and procedures. The degree of researcher control for most inspection research has been limited, as most of the published research on inspections has been conducted in field settings using professional subjects or in relatively uncontrolled experimental-type conditions. While research in this area has historically
been performed by investigators from industry (e.g., Fagan, 1976; Myers, 1978), in recent years a few studies have been reported by researchers from academia (e.g., Bisant and Lyle, 1989; Schneider, et al., 1992). A review of the research on software reviews is presented below.

2.1 Research studies: The effectiveness and benefits of software reviews

Several studies have investigated the effectiveness and benefits of software reviews. These studies have focused on reviews involving a variety of software work products: programming code, design specifications, and user requirement specifications. The dependent variables examined in these studies included error detection efficiency (i.e., portion of errors uncovered through the review process), systems development productivity, and educational benefits.

2.1.1 Field studies: Design and code inspections

Fagan's study (1976) of the impact of inspections on coding productivity has been widely cited as a landmark study demonstrating the value of inspections (e.g., see Ackerman, et al., 1989; Bisant and Lyle, 1989). This study was conducted in a field setting and compared the productivity of a design and programming group using the inspection technique to a group that did not use any inspections. The work product that was
reviewed by the inspection group was part of an actual operating system software project and was considered by Fagan "to be of sufficient size and nature to be representative for study purposes" (p. 187). Since the study was conducted by Fagan, it is assumed (although not explicitly stated) that Fagan's (1976) methodology regarding the inspection technique was followed (e.g., inspection team sizes of three to five people, adherence to the inspection procedure).

Fagan (1976) found that the inspection team experienced a 23% increase in productivity as a result of conducting inspections after the design and code steps (i.e., the total personnel time required to build the subcomponent -- inspection, coding and testing time -- was 23% less than that of the no-inspection case). The improvement in productivity was attributed to the fact that the cost to fix an error increases the longer it goes undetected. Fagan also conducted a comparative analysis to ascertain whether the productivity improvement of the inspection group may have been due to the "Hawthorne Effect." Based on this analysis, no significant Hawthorne Effect was found.

Fagan (1976) also reports the results from two field studies examining the "error

---

1 It should be noted that the comparability between the work products used for Fagan's inspection and no-inspection cases is somewhat clouded, as Fagan did not provide details about the no-inspection case used in his comparison. Hence, it is not clear whether the work product of the no-inspection group was the same as that of the inspection group or a completely different work product altogether.

2 The Hawthorne Effect is a phenomenon that may confound the results of human productivity studies due to the subjects' knowledge that they are participating in a study (Cook and Campbell, 1979).
detection efficiency" of the inspection process. Error detection efficiency was calculated by dividing the number of defects identified during the inspection process by the total number of defects existing in the work product before inspection (Fagan, 1976). The number of defects in the work product before inspection is approximated by adding up the number of defects found in the work product through the inspection, testing, and early usage phases. The results suggest that design and code inspections can be used to detect a majority of the defects which exist in a work product. In a study involving an operating systems program developed at IBM, the error detection efficiency was found to be 66 percent (i.e., inspections detected 66 percent of the defects while 34 percent of the defects found in the work product were found outside of the inspection process during testing and usage). In a second study involving an application program developed at Aetna Life and Casualty, the error detection efficiency was 82 percent.

A more recent paper by Fagan (1986) cites studies which also found high error detection efficiencies associated with inspections. Data collected from a project developed at IBM found an efficiency of 93 percent due to design and code inspections. In two other cases involving systems developed at American Express and the Standard Bank of South America, Fagan found detection efficiencies of "over 50 percent" (p. 745: Fagan, 1986). In these latter two cases the detection efficiencies were high despite the fact that a trained inspection moderator was not used and the Bank case conducted only code inspections.
2.1.2 Field study: User requirements specification inspections

In addition to examining the effectiveness of inspections of work products generated in the later stages of systems development (i.e., design and code), research has also investigated inspections involving the major software work product generated during the early phase of systems development: the user requirements specification. Doolan’s (1992) report on the experiences of a software development group at Shell Research indicates that inspections of requirements specifications can have a very favorable payback for practitioners. The inspection teams in Doolan’s study consisted of four to six people and followed Fagan’s inspection technique. Estimates of the "return on investment" for the inspections was based on data collected from inspections of eleven separate requirements specifications and Shell’s historical data regarding the costs to fix defects in released software. The types of defects that were detected were classified according to severity. For example, a "major" defect was defined as an incorrect specification or oversight that would be expected to prompt a "fault report" (i.e., a user request for software repair, modification, or enhancement). The cost associated with handling a fault report was not trivial as an average fault report required one person-week of labor and cost $3500.

In all eleven cases in Doolan’s study, the requirements specification inspection uncovered many major defects (range: 19 to 125 defects per specification). Furthermore, at least
one "super-major" fault (i.e., a fault which would have rendered the software virtually useless) was found in six of the specifications. Overall, Doolan concluded that for every hour invested in inspection an average of nearly 30 hours of future rework time was avoided. The favorable findings from this study are particularly interesting since they contradict the position held by one well known researcher in the area that inspections are ineffective in detecting "high-level" design errors, such as those made in the requirements analysis process (Myers, 1979).

2.1.3 Semi-controlled studies: User requirements specification inspections

The error detection efficiency for teams inspecting a user requirements document using Fagan's technique has been investigated in two related studies: a pilot study performed by Martin and Tsai (1990) and a replication study by Schneider, et al. (1992). The findings from the replication study are discussed here. The work product reviewed for the study was a comprehensive specification for real-time track control for railway companies prepared by the researchers (the size of the document was 22 pages). The number of defects in the document was established by previous inspection to be 99 defects. The subjects in the replication study included 27 computer science graduate students and the size of the ad hoc teams was three people. On average, the error detection efficiency for each team was found to be 35 percent. In their discussion, the authors of this study noted that the error detection efficiencies reported in the literature
are generally higher than this value (e.g., detection efficiencies for the studies cited earlier are above 50 percent). The authors attribute the lower detection efficiency to the fact that the work product used for their study was a specifications document rather than a design document or programming code.³

2.1.4 Educational benefits of the software review process

In addition to providing a mechanism for early defect detection, various researchers have also reported that the inspection process can offer educational benefits for the inspection participants. The observed educational benefits include:

1) Education in software development skills: The feedback provided by the inspection process allowed developers to learn about the mistakes they (or others) made and thus helped to educate the participants about the pitfalls associated with development activities (Bisant and Lyle, 1989; Fagan, 1976; Myers, 1978) as well as the standard organizational development practices (Doolan, 1992).

2) Education in project and technical knowledge: The steps of the inspection process (overview, preparation, and inspection) were viewed as an effective way to introduce

³ While not mentioned by Schneider, et al. (1992), it is also possible that the variation in detection efficiency across the studies may be due to differences in subjects (i.e., professional workers versus students), research settings, and the realism of the inspection work product.
participants to a new project or technical subject in a short amount of time (Doolan, 1992; Fagan, 1976).

3) Preparation for replacement personnel: The inspection was considered to be a useful means for technology transfer and helped to prepare replacement development personnel in the event that someone was removed from a project (Doolan, 1992).

2.1.5 Semi-controlled study: The educational benefit of programming reviews

The observed educational value of programming reviews has been supported to some extent by a study by Lemos (1979) which explicitly investigated the relationship between inspections and learning. This study was conducted to find out whether undergraduate college students can improve their programming skills by participating in a review. The dependent variables for the study were Cobol program writing, reading and grammar skills. A total of 215 subjects participated in the study. Subjects in the experimental group participated in classroom review activities which involved meeting in ad hoc teams of three or four students. The review technique used by these subjects was informal and did not strictly follow Fagan’s technique (e.g., the review involved no preparation step, no checklists, and subjects corrected as well as detected defects). Subjects in the control group did not participate in any classroom review activities. The skills of the students were measured using post-test exams (pre-tests were also conducted to collect data to use
as covariates).

The results of Lemos' (1979) study indicated that subjects who had participated in the reviews performed significantly better than the control group with respect to writing Cobol programs on the post-test. There was no significant differences with respect to Cobol grammar skills. Due to statistical problems, no analysis was reported for Cobol reading skills. The findings thus indicated that the educational value of informal reviews depends on the type of proficiency measure being tested. Lemos concluded that reviews can be an effective pedagogical tool for an introductory programming course.

2.1.6 Summary

Based on the studies described above, group software reviews appear to be an effective means of detecting software defects, improving productivity, and promoting educational benefits. The studies investigated the full spectrum of software work products ranging user requirements specifications (Doolan, 1992; Martin and Tsai, 1990; Schneider, et al., 1992) to design specifications and programming code (Fagan, 1976; Fagan, 1986; Lemos, 1979). Although none of these studies were conducted under highly controlled research conditions, they do seem to offer reasonable support for the notion that software reviews are a worthwhile practice for software developers. In particular, the field data collected by Fagan (1976; 1986) and Doolan (1992) is valuable since it has helped to
establish the value of software inspections as an important technique for software developers.

2.2 Research studies: Investigation of variations to the software review process

In addition to substantiating the value of the software review process, it is also important to investigate alternate ways of performing software reviews so that the state of the practice may be enhanced. There are relatively few published studies which have investigated variations to the software review process. Research in this area has included an eclectic assortment of studies: a comparison of Fagan’s (1976) formal inspection procedure to an informal walkthrough, a comparison of a team review approach to individual computer-based testing, studies which consider variations in the number of participants in the inspection process, and a study which considers the effect of the preparation and inspection rates on inspection quality. Based on the dearth of research in this area, it does not appear that there has been a focussed stream of research aimed at improving the quality of software inspections. The primary dependent variables measured for these studies have been inspection effectiveness (measured as the number of defects detected) and post-review labor efficiency.
2.2.1 Field study: Comparison of inspection and walkthrough techniques

There is only one reported study comparing the relative effectiveness of Fagan's formal inspection technique to the more informal walkthrough technique. Fagan's (1976) field study compared the defect detection performance of two different teams of programmers working on separate subcomponents of an operating systems project (according to Fagan the two subcomponents used for the study were "fully comparable"). The dependent variable in the study was the number of undetected defects remaining in each work product following the review process. The number of defects remaining in the work products were tallied during seven months of testing that was performed beyond the review stage. Fagan concluded that the inspection process was superior to the informal walkthrough process since the work product reviewed by the inspection technique contained 38 percent fewer defects than the walkthrough work product. This finding is interesting since it suggests that the type of review process employed can influence team performance. However, it is not possible to determine what aspects of the review process impacted team performance since Fagan does not elaborate on the reasons underlying the differences in the performance of the inspection and walkthrough groups.

2.2.2 Semi-controlled study: Team review versus individual computer-based testing

Myers' (1978) research is the only study which has directly compared the detection
efficiency of the team review approach to a detection approach which does not involve a review (i.e., computer-based program testing). This study -- which is referred to by its author as a "controlled experiment" -- considered three different approaches to the detection of defects in a section of program code: 1) An unspecified team review technique (team size of three), 2) A one-person computer-based testing with tester access to the program’s specification only, and 3) A one-person computer-based testing with tester access to the program’s specification and listing.

The subjects in Myers’ study included 59 "highly experienced" programming professionals who were asked to find program defects existing in a PL/1 program consisting of 63 statements and 15 known defects. While a time limit of 90 minutes was imposed on the walkthrough teams, there was no time limit for those performing individual testing. On average, the error detection efficiency of the subjects was rather low as subjects (i.e., testers working alone or in three-person teams) found 34 percent of the defects in the program. No significant differences across the treatment groups were found with respect to the number of defects detected. However, the labor cost per error detected was significantly higher for the review teams. While the individual testing approaches were found to be relatively more cost effective than a

---

4 It should be noted that computer-based testing is an option for a work product consisting of program code, but typically is not an option for other work products generated during the software development process (e.g., requirements specification, design specifications).

5 The type of review technique used by subjects participating in the review treatment was not controlled by the experimenter, as each review team was permitted to decide upon the review technique used.
review, Myers cautioned against concluding that reviews were not cost effective since field experience had found otherwise.

2.2.3 Semi-controlled study: Two-person inspection technique

As described earlier, Fagan's inspection approach calls for three to five participants (including a moderator). However, researchers have suggested that there may be situations in which fewer -- or more -- people could be involved in the process. For example, in development environments with very limited personnel resources it may not be possible for more than two reviewers to participate in an inspection. Alternatively, in situations where the error detection efficiency of individual teams is found to be low it may be desirable to involve multiple teams of reviewers in the inspection process (and thereby increase the chances of detecting defects). Three studies have focussed on variants to Fagan's approach which involve fewer or more participants in the inspection process.

Research by Bisant and Lyle (1989) investigated the effectiveness of an inspection method that involved only two people and did not use a moderator. A pretest-posttest control group design was used for this study. The dependent variable was time required to complete a programming task. The subjects were 32 undergraduate college students who were given the task of completing two programming assignments (there was no
control over programming language used as students were allowed to code their assignments using any high-level language of their choice). All subjects prepared self-report logs of the amount of time they spent on each of the assignments. Subjects in both groups performed the first assignment without using inspections (i.e., the pretest). While the control group also completed their second assignment without using inspections, those in the experimental group performed two separate in-class inspections (one for the program design and one for the program code). The inspection procedure called for subjects to pair up and inspect each other’s work. Twenty minutes was spent on the inspection of each work product (plus five minutes of preparation time to look over the work product before the inspection). While no moderator was used, the logic of each work product was read aloud by one of the reviewers.

Bisant and Lyle (1989) found that subjects in the experimental group improved significantly with respect to programming speed as a result of the two-person inspection. Slower programmers generally exhibited the most improvement. The results were attributed to the early detection of program defects and to problem knowledge imparted as a by-product of the inspection process. The authors suggested that their findings would likely generalize to professional programmers (although not to the same degree) and that the two-person inspection approach would be a useful approach for organizations with personnel limitations. Also, they indicated that this approach might serve as a useful way to introduce the inspection concept to personnel who are unfamiliar with the
inspection method.

2.2.4 Semi-controlled studies: N-fold inspection technique

While Bisant and Lyle (1989) considered a team inspection technique which reduces the labor required for an inspection, another pair of studies has focused on an extension of Fagan's approach which is much more labor intensive. A pilot study by Martin and Tsai (1990) along with a follow-up study by Schneider, et al. (1992) examined the merits of an "N-Fold" inspection technique which calls for the replication of an inspection by several independent inspection teams. The authors hypothesized that different inspection teams would find different errors in a work product (i.e., limited overlap) and thus the aggregation of findings from several inspection teams would improve overall error detection efficiency.

As described in a previous section, each of the N-Fold studies involved the inspection of a user requirements document. The dependent variable was number of defects detected by a team. Subjects for the pilot study included 40 undergraduate and graduate college students organized into ten teams (four people per team), while the follow-up study included 27 graduate students organized into nine teams (three people per team). The design of the follow-up study was more tightly controlled than that of the pilot study, as a special team selection procedure was used to equalize ability differences across
teams and all subjects were given identical lectures on topics relevant to the inspection exercise (e.g., requirements analysis, fault types and inspection procedures). Although the researchers made efforts to control aspects of the N-fold studies, there were some important lapses of control. For example, the inspection sessions were not monitored, so it is not certain whether a common inspection approach was used by the groups in the study. Also, there were no controls with regard to the time that teams spent on preparing for the inspections or conducting the inspections.

In each of the N-fold studies the researchers found that the error detection efficiency of individual teams was relatively low: 27 percent for the pilot study and 35 percent for the follow-up study. However, it was found that the teams in the study tended to detect different errors and that the aggregation of findings across teams served to greatly improve overall error detection efficiency. The aggregate error detection efficiency was 84 percent for the pilot study (N=ten teams) and 78 percent for the follow-up study (N= nine teams). Based on their findings that the N-Fold approach can substantially improve error detection efficiency over that of independent inspection teams, the authors conclude that the N-Fold inspection approach may have application in industry for certain types of large-scale real-time systems where reliability and robustness are of critical importance and error detection efficiency during inspections needs to be maximized.

An interesting finding arising from the N-Fold studies was the wide variation in team
inspection performance (Schneider, et al., 1992). In the pilot study the ratio of performance (i.e., number of errors detected) of the best team to worst team was 4.7 to 1. In the follow-up study -- in which special measures were taken to assemble teams of equal ability -- the performance ratio was 2.3 to 1. While such variations may seem large, it is not uncommon for experimental studies involving software-related tasks to reveal large differences between individuals (e.g., see Curtis, 1980; Moher and Schneider, 1982). For example, an early study comparing programmer performance under batch and time sharing environments reported that the maximum differences of programming performance of student subjects was 28 to 1 (Sackman, et al., 1968). Myers' (1978) study of program testing and walkthroughs using highly experienced programmers as subjects also uncovered large variations in individual performance. These findings regarding group and individual performance differences suggest that personnel selection and training may play a key role in improving team inspection performance.

2.2.5 Field study: Relationship between inspection performance and procedures

Lastly, there has been one study which has examined specific aspects of the inspection process which may influence error detection. As discussed in Bisant and Lyle (1989), Buck (1981) investigated the effect of the following procedural variables on inspection performance: the rate of coverage during the inspection, the amount of preparation on
inspection performance, and inspection team size. While the details provided by Bisant and Lyle regarding this study are limited, it appears that Fagan's (1976) inspection procedures were followed and that no explicit experimental manipulations were used; instead, descriptive data regarding the performance of groups was collected and then analyzed. Buck examined the results of 106 inspections of a constant software module that was used for training purposes. Group sizes ranged from three to five people and the inspection involved a review of programming code.

A key finding from Buck's (1981) study was that error detection was dependent on the rate of coverage of lines of code during the inspection. Groups which inspected code below a threshold rate of 125 lines per hour (not including commented lines) detected 43 percent more major errors than groups who proceeded at a faster rate. While increases in preparation time improved the likelihood that groups would proceed at the desired rate of coverage, extra preparation did not appear to help the performance of groups which had an excessive rate of coverage during the inspection. Group size was not found to significantly impact error detection. While the apparent lack of control regarding the independent variables in this study may have resulted in confounded findings (e.g., it is difficult to isolate the effects due to variations in group size, preparation time, and coverage rate), the study does suggest that procedural aspects associated with the inspection process can influence the performance of inspection teams.
2.2.6 Summary

Two major observations can be drawn based on the review of research presented in Section 2.2. First, it is evident that very little research has focussed on ways to improve the performance of software review teams. Among the studies reviewed in Section 2.2, only the N-fold studies (Martin and Tsai, 1990; Schneider, et al., 1992) and the study by Buck (1981) were aimed at improving the practice of software error detection. It thus appears that since the time Fagan (1976) published his classic article on software inspections seventeen years ago, little progress has been made in improving the practice of software reviews. Second, the research studies in this area have generally involved research designs with a limited degree of control. For example, none of the semi-controlled studies described in Section 2.2 (Bisant and Lyle, 1989; Martin and Tsai, 1990; Myers, 1978; Schneider, et al., 1992) implemented research controls to ensure that the experimental teams were following a standard set of procedures before or during the group review sessions. Due to the lack of controls it is hard to know whether variations across treatment groups were due to the treatments or due to procedural differences. The problems associated with the lack of control highlights the need for future research which incorporates a research design with more control.
2.3 Future directions for software inspection research

Given the widespread use of software reviews in industry, it is surprising to find that relatively few research efforts have investigated ways to improve the software review process. Even modest improvements in the conduct of the inspection process are likely to provide tangible payoffs since the cost of fixing a software defect rises substantially the longer the defect goes undetected (Fagan, 1976; Pressman, 1992). While a few studies have investigated ways to improve inspection quality, the existing research is fragmented and there does not presently appear to be an active research agenda concerning software inspections.

To provide direction for future research on software inspections it is necessary to identify the factors that may influence the quality of reviews. A useful starting point for researchers in this area is the fishbone cause/effect diagram introduced by Fagan (1986). The diagram identifies several key contributors to software inspection quality (see Figure 2.1). Fagan developed the diagram based on ten years of studies, evaluations, and observations of inspections at IBM. According to Fagan, the four main contributors to inspection quality are: the inspection process, product inspectability, managers, and
Figure 2.1: Fishbone diagram of contributors to inspection quality (from Fagan, 1986)
programmers. The main contributors to quality are represented as the major "branches" on the diagram. A number of subcontributors which influence the main contributors are also identified. The subcontributors are represented as "twigs" to the appropriate branches. For example, as indicated in Figure 2.1, a subcontributor to "inspection process" is "process execution." "Process execution" is in turn influenced by "team synergy" and "rate of inspection." Each of these subcontributors is in turn influenced by other factors (shown as twigs in the diagram).

The studies described in section 2.2 of this chapter which focus on the variations to the software review process appear to fit into the "inspection process" branch of Fagan's diagram. For example, Buck's (1981) study clearly fits into Fagan's framework as part of the "rate of inspection" twig. Also, Fagan's (1976) field comparison of the inspection and walkthrough approaches is a study that fits under the "well defined inspection process" twig. While the studies by Bisant and Lyle (1989), Martin and Tsai (1990), and Schneider, et al. (1992) do not fit as neatly into Fagan's diagram, they each involve an investigation of alternate approaches to the traditional inspection process and thus could be considered as studies falling under the "inspection process" branch of Fagan's diagram. Based on the interesting findings arising from these studies, researchers may

---

6 Fagan (1986) uses the term "programmers" to describe one of the main branches of his diagram. However, it appears that this term could be replaced with the term "developers" without changing the spirit of Fagan's diagram. The term "developers" is suggested since it would be a more general term to use and would be appropriate for inspections of systems development work products from all stages of the systems life cycle.
be motivated to perform more in-depth research into the "inspection process" with the intent of improving the quality of software inspections.

One particularly fruitful area for additional research concerns the group process aspects of the inspection task. Findings from the domain of group research suggest that the effectiveness of groups depends strongly on issues relating to the group process (e.g., see McGrath, 1984; Shaw, 1981). For example, the productivity of interactive groups often can be influenced by process losses (e.g., domination of a meeting by one or more participants) and process gains (e.g., stimulation due to the exchange of ideas among participants) (Lamm and Trommsdorff, 1973). An awareness of the group process issues which are relevant to the software inspection task would be valuable in helping to identify and investigate ways to improve the productivity of software inspection teams. The next chapter provides a summary of relevant literature from the area of group research.
CHAPTER 3: LITERATURE REVIEW: GROUP RESEARCH

Since the software review is a group activity, any study of software reviews need to take the body of group research into account. This chapter begins with a review of Steiner's work relating to group productivity. Second, the issue of task characteristics is explored. Next comes a discussion of the potential sources of group process loss and gain. Lastly, a brief review of approaches for group support is provided.

3.1 Group productivity

Group productivity lies at the heart of any investigation into support for review teams. Steiner (1966) has suggested that group productivity is a function of three major factors:

1) The **group resources** includes all of the knowledge, abilities, and skills that may be relevant to performing a task. For group tasks, the distribution of resources across individuals is an important issue to consider since group productivity may depend on the heterogeneity of the group's resource base.

2) The **nature of the task** defines the requirements necessary for performing the task. These requirements include the types and amounts of resources needed to perform a task, as well as the prescribed way in which the resources must be combined.
and utilized by a group.

3) The **group process** describes the actions actually taken by groups to transform resources into a group product. Elements of group process may include interpersonal communication, the structure of interaction, decision rules, and unproductive actions.

Steiner (1972) theorized that actual group productivity can be defined as the difference between a group’s "potential productivity" and the losses arising from a faulty group process. Potential productivity is determined by group resources while losses are a function of the efficacy of the group process. Steiner’s formula for actual productivity presumes that gains arising from the group process are not possible. However, it has been shown that group interaction may result in findings that would not occur if individuals were working alone (e.g., Miner, 1984; Shaw and Ashton, 1976). Hence, Steiner’s formula for actual group productivity may be modified to include both gains and losses associated with the group process:

\[
\text{Actual Productivity} = \text{Potential Productivity} - \text{Process Loss} + \text{Process Gain}
\]

(Equation 1)

Assuming that the contributions associated with each of the subcomponents of actual
productivity can be assessed, Steiner’s modified formula may be applied to gain insight into the group process. To estimate a group’s potential productivity, Steiner (1966) has proposed several models which are appropriate for different types of group tasks. Process losses and gains can be estimated by comparing potential productivity with actual productivity. Specifically, process losses represent the portion of potential productivity that is not employed by the group and process gains represent the portion of actual productivity that arises as a result of group interaction (i.e., not predicted by potential productivity).

3.2 Task characteristics

Group research has been conducted for a variety of tasks (e.g., creativity, planning, negotiation). However, no research has focussed specifically on the review task (i.e., a task in which a group meets to identify defects in a work product). To determine which areas of group research may be relevant to the software review task, it is first necessary to determine the generic task type for the software review. McGrath (1984) has developed a "circumplex model" of group task types which he uses to classify group tasks into eight different types. Based on McGrath’s definition, the software review appears to be an "intellective" task.

Intellective tasks are tasks which have a correct answer (McGrath, 1984). Since there
is generally a correct way and an incorrect way to prepare a software work product (e.g., program code can be developed with or without "bugs"), the software review is a task in which the review team members should be able to arrive at the "correct answer" (i.e., the team properly diagnoses a defect existing in the work product). Hence, the software review may be viewed as an intellective task. An intellective task may in turn be further described based on the way that the group performing the task combines its resources. Steiner's productivity models for disjunctive and complementary tasks appear to be relevant to our discussion of the software review task (Steiner, 1966).

3.2.1 Disjunctive tasks

The task of detecting a specific defect in a document may be viewed as a "disjunctive" subtask (Steiner, 1972). For each of these subtasks, the potential productivity of the group may be inferred based on whether there is someone in the group who has the knowledge or insight to detect the defect. Actual group productivity with respect to each subtask depends on whether an individual's resources are made available to the group and whether the contribution is accepted by the group.

One aspect of the disjunctive task that has been found to influence group productivity is

---

1 There may be some review situations where there is uncertainty regarding the correctness of a defect diagnosis. For example, if the user requirements for a system are not clearly defined, then there may be uncertainty regarding the correctness of the analysis system specification document.
the degree to which the correct answer is apparent to members in the group (Steiner, 1972). Tasks for which the answer is immediately evident and accepted are called "Eureka" tasks. If a task has a strong Eureka nature then process loss associated with the failure to accept valid findings is likely to be minimal. Hence, group productivity for a Eureka task hinges on making sure that each individual's resources are made known to the rest of the group. An experimental study involving a strongly Eureka-type task found that group members utilized their knowledge and resources using a "social decision scheme" (i.e., a combination rule) referred to as "truth wins," in which the group accepted a correct answer even if only one member of the group found the answer (Laughlin, et al., 1976).

However, not all disjunctive tasks are Eureka-type tasks. Researchers have found that for many of the disjunctive tasks that have been studied, the correct answers are not always obvious to a group. For these disjunctive tasks, group members typically use a social decision scheme referred to as "truth supported wins," in which at least two members of the group need to have found the answer before it is accepted by the group (e.g., Laughlin, 1980). Group productivity for non-Eureka disjunctive tasks thus does not depend simply on whether or not at least one person presents a correct answer to the group, but also depends on whether there are other members of the group who also recognize the correct answer and are willing to support it.
3.2.2 Complementary tasks

The use of groups to perform a review is typically justified on the basis that individual members of a group will pool their knowledge and identify more defects than a single person working alone (e.g., Yourdon, 1989). Restated in Steiner's terms, the implicit assumption underlying the practice of review teams is that the overall task of finding all of the defects in a work product is a "complementary" task. Complementary group tasks are those in which resources are distributed among different members of a group and more than one individual is responsible for a group's output. Group productivity for complementary tasks depends on how well the group uses the disparate resources of its members. For these types of tasks there is a strong potential for a group to outperform an individual working alone assuming that: 1) the different members of the group can provide unique resources to the group (i.e., nonoverlapping perspectives and abilities), and 2) the group can effectively utilize both the overlapping and the unique resources of the group members (Steiner, 1966).

3.3 Process loss and gain

It has been suggested that "a synergistic effect" can be created during a software review such that the interactive review group may perform better than the sum of its parts (Fagan (1976): p. 190). While the notion of process gain resulting from an interactive
group review has an intuitive appeal, group researchers who have studied other group
tasks have found that it is difficult (if not impossible) to achieve net process gain. In
fact, researchers have typically found that interactive groups perform worse than
comparable "nominal" groups (e.g., see Hill, 1982; Lamm and Trommsdorff, 1973;
Steiner, 1972).² For example, group research in the area of idea generation has
repeatedly shown that interactive "brainstorming" groups are less effective than nominal
groups (e.g., see the review by Diehl and Stroebe, 1987).

As suggested by Steiner’s modified equation for actual group productivity (see equation
1 in section 3.1), the poor performance of interactive groups relative to nominal groups
may be attributed to a negative balance of process loss and gain. Researchers have
identified a number of potential sources of process loss and gain which may operate
during group meetings. The identification of the specific sources of group process loss
and gain has been valuable for group researchers, since such information may be used
to help develop techniques that may improve the performance of groups. Following is
a summary of several of the potential sources of process loss and gain that have been
identified.

² A "nominal" group may be defined as a group of individuals who work independently and then pool
their findings (Shaw, 1981).
3.3.1 Potential sources of process loss

Cognitive inertia takes place when a group discussion follows a common thread of thought (Hararai and Graham, 1975; Lamm and Trommsdorff, 1973). "Side-tracking" away from the topic of interest is a particularly unproductive form of cognitive inertia.

Domination arises when one or more members of the group monopolizes the group’s meeting time or influences the group in an unproductive manner (Jablin and Seibold, 1978).

Evaluation apprehension occurs when participants are apprehensive about contributing to the group since they think that their comments may be received in a negative way (Diehl and Stroebe, 1987; Harari and Graham, 1975; Lamm and Trommsdorff, 1973).

Failure to remember to the contributions of others may cause information to be forgotten or may cause group members to waste meeting time by repeating contributions made by other members (Nunamaker, et al., 1993).

Free riding takes place when group members count on others to work on the task at hand (Diehl and Stroebe, 1987).
Information overload arises when information is presented to group members faster than it can be processed (Hiltz and Turoff, 1985).

Production blocking results when group members need to wait before they can make their contributions to their group. Group members may not use their time effectively while waiting to contribute a comment since they are trying to remember their comment or are listening to the comments of others. Additionally, members may forget or suppress contributions while they wait for others to finish speaking (Diehl and Stroebe, 1987; Lamm and Trommsdorff, 1973).

3.3.2 Potential sources of process gain

Learning arises when group members watch others perform the same task and improve their abilities through observation and group participation (Hill, 1982; Lamm and Trommsdorff, 1973).

More information may be made available to the group by combining the resources of the group members (Lamm and Trommsdorff, 1973; Steiner, 1972).

More objective evaluation may occur since members who do not contribute a comment to the group may be more likely to evaluate the comment in an unbiased way (Hill,
Synergy happens when information provided by one member of the group is used by others in a new and productive way due to the application of different knowledge and skills (Lamm and Trommsdorff, 1973; Osborn, 1957)

3.4 Support for groups

Group researchers have proposed a number of approaches aimed at supporting interactive groups (e.g., Van Gundy, 1981). Although not all of the approaches have lived up to the claims made by their proponents, research studies suggest that several of the proposed methods may help to improve group performance. For example, research indicates that the Delphi method and the Nominal Group Technique may support groups performing the generation and decision making tasks (Delbecq, et al., 1975), while the use of a discussion leader may aid groups performing an intellective task (Maier and Solem, 1952). Although techniques such as these appear to be useful, there still appears to be potential to improve group performance through additional research.

In the ongoing quest to improve the performance of groups, researchers have not limited themselves to "traditional" approaches to group support. Since the early 1980s, there has been a growing interest in the use of electronic information technology to support groups.
Researchers have built and investigated a wide assortment of information technology-based systems that may support group work. The systems have gone by a variety of names including group decision support systems (GDSS) (e.g., DeSanctis and Gallupe, 1987), group support systems (GSS) (e.g., Jessup and Valacich, 1993), computer-supported cooperative work (e.g., Kraemer and King, 1988), groupware (e.g., Johansen, 1988), and electronic meeting systems (EMS) (e.g., Dennis, et al., 1988).

For this study, the term electronic meeting system (EMS) will be used to refer to information technology that is used to support group meetings. As defined by Dennis, et al. (1988), an EMS is:

"An information technology-based environment that supports group meetings, which may be distributed geographically and temporally. The IT environment includes, but is not limited to, distributed facilities, computer hardware and software, audio and video technology, procedures, methodologies, facilitation, and applicable group data. Group tasks include, but are not limited to, communication, planning, idea generation, problem solving, issue discussion, negotiation, conflict resolution, systems analysis and design, and collaborative group activities such as document preparation and sharing" (p. 593).

Currently, EMS technology has evolved to the point where it is routinely being used in industry settings. For instance, the EMS developed at the University of Arizona, GroupSystems³, is being sold commercially by Ventana Corporation of Tucson, Arizona, and is being used in hundreds of organizational sites around the world. Based on the

³ GroupSystems is a registered trademark of Ventana Corporation.
findings from the early EMS research studies, it appears that EMS technology holds promise for improving the performance of groups by addressing a number of the potential sources of group process loss and gain.

In what ways may an EMS -- or a non-electronic approach -- address group process loss and gain? The next chapter offers an answer to this question by introducing the four general types of group support mechanisms that may be used to support groups (Nunamaker, et al., 1991). The discussion of the support mechanisms also includes references to relevant EMS studies. The support mechanisms are introduced in the next chapter since they provide a convenient bridge between the group literature and the research questions and hypotheses explored in this dissertation.
The foregoing review of the group research literature suggests that the way in which group meetings are conducted can significantly influence group performance. Since the software review task has received little attention from group-oriented researchers in the past, this research area is relatively "untapped," as potential ways to enhance the performance of software review teams through group support techniques have not yet been explored. When approaching the design of a research study (such as this one) which is aimed at improving the practice of team software reviews, several important questions may be raised: What are the generic types of group support techniques that we should consider? How might these support techniques impact the group process? What are the key independent and dependent variables? To help address these questions -- and thereby guide this research study -- this chapter builds on the previous chapter by introducing mechanisms that may theoretically be used to support software review teams. Next, the specific independent and dependent variables that are germane to this study are introduced. Finally, the research questions and hypotheses are presented.

4.1 Theoretical group support mechanisms

Nunamaker, et al. (1991) have proposed that there are at least four theoretical group support mechanisms which may be used to aid groups by influencing the balance of
group process loss and gain: process support, task support, process structure, and task structure.\(^1\) These mechanisms are described below. Additionally, the potential impacts of the support mechanisms on process loss and gain are discussed and summarized in Table 4.1 (see Nunamaker, et al., 1993 for a more detailed discussion). Studies which have examined the impacts of the support mechanisms on groups are also cited.

### 4.1.1 Process support

Process support relates to the communication infrastructure that may be used to aid communication among group members, where the communication infrastructure includes the various communication media and channels used by a group (e.g., a blackboard, an electronic communication channel). Process support may be delivered to a group through a group memory, parallel communication, and anonymity.

**Group memory** provides a group with a record of the group inputs during a meeting and offers the group participants the chance to decouple from the group discussion to ponder the task at hand without missing contributions by others. This type of process support may reduce process loss associated by failure to remember group inputs and incomplete use of information. Additionally, a group memory which helps groups to filter

\(^1\) Although these theoretical mechanisms have been identified by EMS researchers, in many cases the mechanisms may be implemented through non-electronic -- as well as electronic -- means.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Process Support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Group memory</td>
<td>o Failure to remember</td>
<td>-</td>
<td>o More information</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>o Info overload</td>
<td>-</td>
<td>o Synergy</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>o Incomplete use of information</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Parallel communication</td>
<td>o Blocking</td>
<td>-</td>
<td>o More information</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>o Domination</td>
<td>-</td>
<td>o Synergy</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>o Free riding</td>
<td>-</td>
<td>o Learning</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>o Info overload</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>o Anonymity</td>
<td>o Evaluation apprehension</td>
<td>-</td>
<td>o More objective evaluation</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>o Free riding</td>
<td>+</td>
<td>o Learning</td>
<td>+</td>
</tr>
<tr>
<td><strong>Task Support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>o Incomplete use of information</td>
<td>-</td>
<td>o More information</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>o Incomplete task analysis</td>
<td>-</td>
<td>o Synergy</td>
<td>+</td>
</tr>
<tr>
<td><strong>Process Structure</strong></td>
<td>o Effects depend upon technique</td>
<td>-</td>
<td>o Effect depend upon technique</td>
<td>+</td>
</tr>
<tr>
<td><strong>Task Structure</strong></td>
<td>o Incomplete task analysis</td>
<td>-</td>
<td>o More information</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>o More objective evaluation</td>
<td>+</td>
</tr>
</tbody>
</table>

Key to symbols: Increase in process loss/gain: +
Decrease in process loss/gain: -

Table 4.1: Group support mechanisms and impacts on process loss and gain
(from Nunamaker, et al., 1993)
information may reduce information overload. Group memory may promote process gain by making more information available to the group and by increasing synergy. While research studies have not generally focussed on the impacts of group memory, EMS researchers have observed that participants in EMS meetings in both experimental and business settings tend to utilize group memory in the ways described above (e.g., see Gallupe, et al., 1992; Tyran, et al., 1992).

Parallel communication allows all group members to communicate at the same time. Typically this type of process support is implemented using an EMS which has a parallel communication channel. Since meeting participants do not need to wait on others to make contributions, the parallel communication may reduce process loss associated with domination, production blocking, and free riding. However, since parallel communication offers the potential to increase group input it may increase loss related to information overload. Since more group input may be generated by a group using this type of support mechanism, parallel communication may encourage process gain related to making more information available to the group, synergy, and learning. A number of EMS studies indicate that parallel communication can improve the performance of groups involved with idea generation by reducing production blocking (Dennis, Valacich, and Nunamaker, 1990; Gallupe, et al., 1991; Gallupe, et al., 1992). Findings from a study of business groups using EMS technology also indicated that parallel communication reduced loss due to domination and production blocking (Tyran, et al., 1992).
Anonymity allows group members to make contributions to their group without being identified. Anonymity is typically delivered by means of an EMS which has an anonymous communication channel. Anonymity may reduce process loss associated with evaluation apprehension, but may increase loss due to free riding. Since anonymity may offer a more open, low-threat environment, it may also encourage gain due to more objective evaluation of group findings and learning. Experimental studies have found that groups working under anonymity are more critical than nonanonymous groups, but have not detected differences in performance due to anonymity (Jessup, et al., 1990; Jessup and Tansik, 1991; Valacich, et al., 1992).

4.1.2 Task support

Task support relates to the information and computation infrastructure that may be used to address task-related activities. Examples of task support mechanisms are databases and calculators. Task support may reduce process loss associated with incomplete use of information and incomplete task analysis. Task support may encourage process gain by making more information available to the group and by increasing synergy. Not much research has focussed on the effects of task support. EMS studies involving business groups indicate that task support involving access to databases of previous meeting sessions and organizational data are useful to groups (Nunamaker, et al., 1989; Dennis, Heminger, Nunamaker, and Vogel, 1990).
4.1.3 Process structure

Process structure relates to techniques and rules that may influence communication patterns and content among group members. Examples of process structure include agenda setting, facilitation, and group process techniques (e.g., Nominal Group Technique). The impact of process structure on loss and gain likely depends on the type of technique used. Studies involving non-electronic (e.g., Hackman and Kaplan, 1974) and EMS-supported groups (e.g., Easton, et al., 1990) have found that process structure can improve group performance.

4.1.4 Task structure

Task structure relates to techniques, decision frameworks, and models that may be used to analyze task-related information. Examples of task structure mechanisms are computer simulation models, "what if" spreadsheet analysis, and the SWOT business analysis technique (i.e., Strengths, Weaknesses, Opportunities, and Threats). Task structure may reduce loss due to incomplete task analysis and may increase gain by making more information available to the group and by encouraging a more objective evaluation of group findings.
4.2 Variables and measures

4.2.1 Independent variables

The independent construct of interest for this study was the group technique used to aid groups performing the software review task. The corresponding independent variable that was manipulated was the type of group support employed by software review teams. With regard to this independent variable, a key research design question focussed on deciding which of the four generic types of support mechanisms should be manipulated in this study. To help decide what mechanisms to investigate, it was necessary to consider the elements of the software review task.

There are essentially two subtasks embodied in all software review techniques: 1) Defect identification, and 2) Defect evaluation. For some techniques, the defect identification and evaluation subtasks may be performed concurrently (e.g., see the inspection procedure described by Fagan (1976)), while for other techniques it is suggested that defect evaluation should be a separate subtask that follows defect identification (e.g., see the inspection procedure recommended by the IEEE standard (1989)). Since the identification and evaluation subtasks appear to be different (and may thus require different types of group support), this research study was designed to explicitly separate the group activities associated with each subtask to isolate the effects of group support.
Having identified the two subtasks of the software review task, we now return to our original question: Which types of support mechanisms should be manipulated in this study? To answer this question, one must take into account the characteristics of each subtask. The reasoning underlying the choices of group support mechanisms to be manipulated for each subtask are discussed below.

**Defect identification**

The defect identification subtask is a divergent task in which members of the review team are required to identify potential defects in a work product and make their findings available to the rest of the team. For this subtask it is sufficient for team members to get their findings "out on the table"; they do not need to deliberate on their findings and decide on whether a finding is a valid defect (i.e., no need for defect evaluation). As prior research suggests that "process support" is a useful type of mechanism for other types of divergent tasks (e.g., brainstorming), this type of mechanism may therefore also be appropriate for the defect identification subtask. Thus, the independent variable that was manipulated for the identification subtask was process support.

As noted earlier, process support may be provided in at least three ways: group memory,
parallel communication, and anonymity. While group memory could be provided by both non-electronic (e.g., a blackboard) and electronic means (e.g., an EMS), parallel communication and anonymity are generally provided using an EMS. For this study, two ways of providing process support during the defect identification subtask were used: group memory and parallel communication. To help simplify the interpretation of the findings, anonymity was controlled across groups (i.e., anonymity was not available for any groups).

Defect evaluation

While the identification of defects is a divergent task, the evaluation of defects is a convergent task, since team members need to reach consensus on the validity of defects identified earlier. The type of group support that is appropriate for this subtask depends on the degree of task equivocality (Nunamaker, et al., 1993). If this subtask was highly equivocal in nature, then effective group utilization of group resources could be problematic. In this case, group performance would be contingent on the status, assertiveness, and confidence of the person(s) who presents a valid finding to the group, as well as the relative proportion of group members who hold contrary positions (Johnson and Torcivia, 1967; Maier and Solem, 1952; Thomas and Fink, 1961). As group support for this task situation would be required to address these types of concerns, it would likely need to be rather comprehensive and might include a combination of all four
types of support mechanisms.

On the other hand, if the defect evaluation subtask proved to be low in equivocality (i.e., a Eureka task), then a relatively simple type of support mechanism may suffice. For example, if the defect evaluation subtask was a Eureka-type task, then one way to support this subtask would be to provide a "task support" mechanism to ensure that all individual findings generated during the identification subtask were made available to the group. This type of mechanism could potentially reduce process loss due to the incomplete use of information. Task support of this type could be provided by both non-electronic (e.g., a written list of individual findings) or electronic means (e.g., electronic access to files of individual findings via an EMS).

For this study, it was expected (based on the author's personal experiences with software reviews) that the defect evaluation subtask would prove to be a Eureka task. Hence, the independent variable that was manipulated for the evaluation subtask portion of the software review was task support.

4.2.2 Dependent variables

The dependent constructs of interest for this study are: 1) Group performance, 2) Process loss and gain, and 3) General participant perceptions (satisfaction, acceptance,
effectiveness). The dependent variables and measures associated with these constructs are described below.

**Group performance**

Group performance was defined to be the quantity of defects identified by the review team. While it would also be possible to define a second group performance variable based on the quality of the group review (e.g., by using "expert" judges to assign importance scores to each defect existing in the work product), this type of measure was not used for the study since it was decided that a quality measure would be inherently subjective and would not add much to the interpretation of the findings.

**Process loss and gain**

As described earlier, Nunamaker, et al. (1993) have proposed specific relationships that may exist between the theoretical group support mechanisms and the sources of process loss and gain (e.g., process support in the form of parallel communication may reduce loss due to air time fragmentation). In this study, the relationships between the process and task support mechanisms and the related process loss and gain constructs (see Table 4.1) were investigated to gain insight into the performance of software review teams. Since it is difficult to measure process loss and gain directly, perceptual measures were
used to assess the various process loss and gain constructs. Unfortunately, few researchers have investigated the process loss and gain constructs discussed earlier. Hence, relatively few validated measures for these constructs exist. While validated measures were used for this study whenever possible, in many cases new measures had to be developed. Following is a summary of the measures used for process loss and gain.

**Process loss:** Validated multi-item scales from the literature were found to assess production blocking and evaluation apprehension. Production blocking and evaluation apprehension were each measured using three-item scales developed by Gallupe, et al. (1992). No validated measures from the literature were available to assess domination, side-tracking, failure to remember, information overload, or free riding. Measures for these constructs were developed by the researcher.

**Process gain:** No measures were available in the literature to assess any of the process gain constructs. Measures of synergy, more information, and learning were developed by the researcher.

---

2 In general, there is a dearth of validated measurement instruments in the area of EMS research. As noted by Zigurs (1993), little has been done to promote greater validity and reuse of EMS instruments, as EMS researchers rarely provide much detail regarding the measures used for their studies.
General participant perceptions

Since the satisfaction and acceptance exhibited by group members may be important to the implementation of a group technique and the longevity of a group (McGrath, 1984), each of these constructs was measured. For each construct, a two-item scale developed by Tjosvold and Field (1983) was used. It was also of interest to gain a sense of each participant's perception of the effectiveness of their group, so that a comparison between a team's actual performance and perceived performance could be made. A simple single-item question was developed to capture each subject's perception of their team's effectiveness.

4.3 Research questions and hypotheses

Since relatively little is known about the nature of the software review task, this study was aimed at addressing exploratory research questions about the software review task, in addition to more formal hypotheses concerning the impact of group support mechanisms on software review performance and process loss and gain. The experimental research method was selected for this study to provide a controlled setting for investigating the research questions and hypotheses. Since it was not possible to design one experiment to address all of the issues of interest, two laboratory experiments were designed. This section presents the research questions and hypotheses that were
investigated by each experiment.

4.3.1 Experiment 1

Experiment 1 was designed to address research questions and hypotheses about the nature of the review task, the intangible benefits of the group review, the correlation of group resources and group process to performance, and the impact of task support on the performance and perceptions of teams performing the evaluation subtask. A discussion of the research questions and hypotheses investigated in Experiment 1 follows.

Before attempting to devise an approach to support the group review task it is necessary to understand more about the factors influencing the performance of review teams. Since the performance of a review team depends on the nature of the task and the knowledge base of its members (i.e., its resources), the research questions focussed on these factors:

**Research Question 1:** What are the characteristics of the review task?

a: Is the review task a Eureka task?

b: Can the review task be described as a complementary task comprised of disjunctive subtasks?
**Research Question 2:** What is the degree of variation in defect detection ability (i.e., resources) across groups and across individuals?

Steiner's theory predicts that group performance is a function of the task, group resources, and the group process. Assuming that the task is held constant, the group process is manipulated, and group resources and performance can be measured, it is possible to investigate Steiner's theory by using regression analysis. In this study, Steiner's relationship was examined for teams performing the defect evaluation subtask. It was expected that both group resources and group process (i.e., the task support mechanism) would correlate with group performance:

**Hypothesis 1:** Group performance for the defect evaluation subtask will be significantly correlated with group resources and group process.

a: There will be a positive correlation between performance and group resources.

b: There will be a positive correlation between performance and a group process involving task support.

As discussed earlier, team performance for the evaluation subtask is likely to depend on the degree to which group members make their findings available to the group. A simple task support mechanism (a visual display of individual findings) was devised to ensure
that the findings of all individuals would be made available to the entire group. The task support was manipulated across three experimental treatments (one treatment received no task support, while the other two treatments received essentially the same type of task support). In the event that the experimental task proved to be a Eureka task, it was predicted that the presence of task support would improve group performance by reducing process loss and increasing process gain. As little is known about the impact of task support on group member satisfaction, acceptance and perceived effectiveness, it was expected that there would be no differences across treatments with respect to these measures. (Note: For the purposes of this study, the treatment groups receiving no group support will be referred to as the "Baseline" groups; the groups receiving the non-electronic group support will be referred to as the "Manual" groups; and the groups receiving electronic group support will be referred to as the "EMS" groups.) The hypotheses testing the impacts of task support are listed below:

**Hypothesis 2:** Groups receiving task support (i.e., EMS and Manual groups) for the defect evaluation subtask will perform better than the unsupported groups (i.e., Baseline groups).

a: Groups receiving task support will experience less process loss than the unsupported groups.

b: Groups receiving task support will experience more process gain than the unsupported groups.
Hypothesis 3: There will be no differences across task support treatment groups performing the defect evaluation subtask with regard to members' satisfaction with the group, group outcome acceptance, and perceived effectiveness.

It has been suggested that software reviews may promote staff training since the group interaction allows group participants to improve their understanding of the defects in a work product by learning from the insights of others (Yourdon, 1989). The review process may also have a favorable impact by making team members more confident of their review skills and the quality of their work product. The fourth hypothesis focussed on the validity of these notions:

Hypothesis 4: With regard to the review task, the subjects will perceive higher levels of understanding and confidence following the group evaluation review (as compared to perceptions following the individual identification review).

4.3.2 Experiment 2

Experiment 2 was designed as a follow-up experiment to address hypotheses related to the impact of process support on the performance and perceptions of teams performing
the identification subtask, and the impact of task support on the performance and perceptions of teams performing the evaluation subtask. Compared to the first experiment, Experiment 2 involved a more comprehensive set of perceptual measures regarding the sources of process loss and gain. While both experiments were designed to examine the impacts of task support, the procedure of Experiment 2 differed from Experiment 1 in that the identification subtask was done individually in Experiment 1 and interactively in Experiment 2. A discussion of the hypotheses investigated in Experiment 2 follows.

Process support has the potential to enhance the balance of process loss and gain for review teams working interactively on the defect identification subtask. In this study, three types of process support for the identification subtask were considered: no process support, non-electronic process support (group memory delivered via a scribe and whiteboard), and electronic process support (group memory and parallel communication channel delivered via an EMS). The hypotheses regarding impact of process support on group performance and perceptions of process loss and gain follow from the propositions set forth by Nunamaker, et al. (1993). As past EMS research findings regarding the general perceptions of group members (e.g., satisfaction) who have received EMS process support are mixed (Dennis, et al., 1990-1991), no differences were hypothesized. The hypotheses for the effects of process support on the defect identification subtask are:
Hypothesis 5: There will be a difference in group performance between groups using different types of process support mechanisms for the defect identification subtask.

a: The EMS groups will perform better than the Manual and Baseline groups.

b: The Manual groups will perform better than the Baseline groups.

Hypothesis 6: There will be a difference in perceptions regarding sources of process loss between groups using different types of process support mechanisms for the defect identification subtask.

a: The EMS groups will perceive less process loss than the Manual and Baseline groups with respect to: domination, production blocking, free riding, and side-tracking. The EMS groups will perceive more process loss than the Manual and Baseline groups with respect to information overload.

b: The EMS groups will perceive less process loss than the Baseline groups with respect to "failure to remember."

c: The Manual groups will perceive less process loss than the Baseline groups with respect to: "failure to remember" and information overload.
Hypothesis 7: There will be a difference in perceptions regarding sources of process gain between groups using different types of process support mechanisms for the defect identification subtask.

a: The EMS groups will perceive more process gain than the Manual and Baseline groups with respect to: synergy, more information, and learning.

b: The Manual groups will perceive more process gain than the Baseline groups with respect to: synergy, more information, and learning.

Hypothesis 8: There will be no differences across process support treatment groups performing the defect identification subtask with regard to members' satisfaction with the group, group outcome acceptance, and perceived effectiveness.

Group researchers investigating the brainstorming task have repeatedly found that nominal groups of people working individually perform better than interactive groups (Diehl and Stroebe, 1987; McGrath, 1984). However, recent work by EMS researchers suggests that interactive EMS-supported groups using process support consisting of group memory, parallel communication, and anonymity can outperform nominal groups for the brainstorming task (Valacich, et al., in press). Since the brainstorming and defect identification task are both divergent tasks, it is possible that process support delivered
via an EMS may be also be used to provide similar results for interactive groups performing defect identification. The performance of nominal and interactive groups performing defect identification can be compared by comparing the findings from Experiment 1 and Experiment 2. This leads to the next hypothesis to be tested:

**Hypothesis 9**: Interactive groups receiving electronic process support will perform better than nominal groups on the defect identification subtask.

The propositions set forth by Nunamaker, et al. (1993) were used to guide the hypotheses with regard to the impacts of task support on the perceived sources of process loss and gain. With regard to the subjects’ perceptions relating to the set of process loss items described earlier, Nunamaker, et al. (1993) have indicated that task support will have no effect: hence no differences were expected. However, differences were expected with respect to the sources of process gain per Nunamaker, et al. (1993). There is little published research to guide expectations regarding the impact of task support on satisfaction, acceptance, and perceived effectiveness. Hence, no differences were hypothesized. The hypotheses for the effects of task support on the defect evaluation subtask are:
Hypothesis 10: Groups receiving task support (i.e., EMS and Manual groups) for the defect evaluation subtask will perform better than the unsupported groups (i.e., Baseline groups).

a: Groups receiving task support will experience less process loss than the unsupported groups.

b: Groups receiving task support will experience more process gain than the unsupported groups.

Hypothesis 11: There will be no difference in perceptions regarding sources of process loss between groups receiving task support (i.e., EMS and Manual groups) and the unsupported groups (i.e., Baseline groups) for the defect evaluation subtask with respect to: domination, failure to remember, production blocking, side-tracking, and information overload.

Hypothesis 12: There will be a difference in perceptions regarding sources of process gain between groups receiving task support (i.e., EMS and Manual groups) and the unsupported groups (i.e., Baseline groups) for the defect evaluation subtask.

a: Groups receiving task support will perceive more process gain than unsupported groups with respect to: synergy and learning.
Hypothesis 13: There will be no differences across task support treatment groups performing the defect evaluation subtask with regard to members' satisfaction with the group, group outcome acceptance, and perceived effectiveness.

4.4 Summary

This chapter has introduced the key concepts, variables, and research questions and hypotheses that were investigated by this dissertation. Two experimental studies were designed and conducted to address the research questions and hypotheses. The first experiment was exploratory in nature and was aimed at addressing the research questions and gaining insight into the effects of task support on performance for the defect evaluation subtask. The second experiment was aimed at addressing the hypotheses related to impact of group support mechanisms on performance and perceptions for each of the subtasks of the group review process (i.e., defect identification and evaluation). The method and results for each experiment are presented in the next two chapters.
CHAPTER 5: EXPERIMENTAL STUDY 1: METHOD AND RESULTS

This chapter describes the research method and the results of the first experimental study that was performed. This study was exploratory in nature and had the following objectives: 1) Learn more about the characteristics of the experimental inspection task so that the opportunities for support for inspection teams could be assessed, 2) Evaluate the impact of a simple task support mechanism to gain insight into some of the reasons underlying performance loss and gain experienced by inspection teams performing the evaluation subtask.

5.1 Method

5.1.1 Task

The task required the subjects to identify defects existing within an "analysis specification document." The specification document described the user requirements for a subsystem of a fictional real estate company and included three main parts typical of a specification document (Yourdon, 1989): a descriptive narrative section, a data dictionary, and three data flow diagrams (DFDs). Subjects were told to assume that the narrative section was a correct description of the user requirements. The specification document used for the study is presented in Appendix A.
The specification document used for the experimental task was developed specifically for this study. Attempts were made to develop a realistic and unambiguous specification which included the types of defects that may occur in specifications generated in the workplace (e.g., see Teague and Pidgeon, 1985). The specification was refined and optimized during a pilot laboratory study involving twelve project teams from an upper-division management information systems (MIS) course. The task was designed so that it could be adequately reviewed within the time constraints imposed by the laboratory exercise (45 minutes for the individual defect identification review and 45 minutes for the group defect evaluation review).

As suggested by Teague and Pidgeon (1985), there are four major criteria for evaluating a system specification: completeness, consistency, communicability, and correctness. Several defects representing each of these types were planted in the specification. Examples of defects included incomplete data dictionary entries, unbalanced DFDs, poorly labeled data flows, and incorrect logical design of processes and data flows. While some defects could be identified by mechanically applying the standard rules associated with developing an analysis specification (e.g., the consistency of data flows and data labels across diagrams), most of the defects were more subtle and required human judgment for detection (e.g., the correctness of the logic represented by DFDs, the communicability of the DFD labeling).
A "master list" of defects for the specification was prepared to assess the performance of the inspection teams. The master list was prepared by cataloguing: 1) All defects that were originally planted in the specification by the experimenter, and 2) Any additional valid defects that were uncovered during the Pilot Study. The master list was generated based on the commonly accepted standards for analysis specification documentation that were discussed in the course lectures (e.g., see (Eliason, 1990)). The master list appears to be complete as no new valid defects were detected by inspection teams following the Pilot Study. The master list of defects for the specification document includes twenty nine defects and is presented in Appendix B.

5.1.2 Subjects

The participants were 57 upper-division MIS students from two sections of a semester-long course in Systems Analysis and Design. All students from each section participated in the study. Lectures and experimental procedures for students in each section were the same. Students from this course were selected since the course work provided an appropriate background for addressing the experimental task (and the subjects had demonstrated their competence of the necessary skills on course assignments and exams). The mean age of the subjects was 24 years old (std=4.6). Thirty five were male (61%) and 22 were female (39%). A large majority of the participants (79%) reported work experience (full-time or part-time).
The subjects were randomly assigned into fifteen groups (twelve four-person teams and three three-person teams). Team sizes of three or four people were used based on literature recommendations regarding the optimal size of software review teams (Fagan, 1976). Prior to the experiment, each group had spent six to seven weeks working together on an extensive group term project involving the analysis of a business system. The laboratory study was conducted several weeks after the teams had been assigned to ensure that each team had an opportunity to "develop as a team" and to thus reduce the potential for confounds associated with experimentally created groups which do not have experience working together (DeSanctis, 1989).

5.1.3 Procedure

To address the research questions and hypotheses, this experiment was designed to allow measurement (or estimation) of the following: 1) The individual resources of each review team member prior to the team review; 2) The potential performance of each review team; 3) The actual performance of each review team; and 4) Individual perceptions regarding the inspection process. The experimental procedure used to capture this information consisted of two major portions: 1) An individual defect identification session, and 2) A team defect evaluation session to discuss and evaluate the individual findings.
The specific review procedures followed for this study were based on Fagan's inspection approach (Fagan, 1976). Due to experimental design considerations there were three ways in which the experimental review procedure differed from Fagan's approach: 1) The individual review portion of the experiment required subjects to explicitly detect defects in the work product (Fagan's approach requires reviewers to become familiar with the work product prior to the review, but does not require defect detection); 2) Fagan's "Reader" roleplayer was not included in the group reviews (however, Fagan's "Moderator" roleplayer -- considered by Fagan to be "the key person in a successful inspection" (Fagan, 1976: p. 190) -- was included); and 3) Two of the three experimental treatments involved the use of task support (Fagan's approach does not include task support).

5.1.3.1 Task support manipulations

Task support for the group defect evaluation process was manipulated across three experimental treatments. The "Baseline" treatment did not provide any task support, as group members were required to speak up and introduce their findings to the group. For this treatment no visual aids were used and there was no guarantee that all individual findings would be introduced to the group. The other two treatments did provide a simple task support mechanism for the defect evaluation session to ensure that all individual findings would be available to the group. The "Manual" treatment provided
task support by means of handwritten lists on transparencies (generated during the individual review) projected onto an overhead screen. The "EMS" (i.e., electronic meeting system) treatment provided support through access to electronic textual files of individual findings projected onto an overhead screen using an EMS and a BARCO projector. The EMS used in this study was the University of Arizona's GroupSystems. The "Group Outliner" tool was used to provide task support (for a description of GroupSystems and Group Outliner see (Nunamaker, et al., 1991)). For this simple application of task support, it was expected that the Manual and EMS approaches would serve equally well to support the defect evaluation process.

5.1.3.2 Pre-experimental procedure

Each group was randomly assigned to one of the three treatment groups described above (there were four four-person teams and one three-person team per treatment). Upon reporting to the experimental site, each team was provided with a brief overview relating to review procedures (all students had previously received a class lecture on this subject). (See Appendix C for the overview handout that was reviewed at the beginning of the experiment.) It was emphasized that the purpose of a review was to detect defects in the specification and not to correct defects. Next, each team received written and verbal instructions describing the procedures to be followed. (See Appendices D, E, and F for the written instructions for the Baseline, Manual, and EMS treatment groups,
respectively.) Subjects were told that their efforts would be evaluated to determine course credit for the exercise. Any questions regarding the procedure were answered.

5.1.3.3 Individual defect identification session

The first portion of the experiment was devoted to a 45 minute individual review of the specification document. Subjects were instructed to spend the first 15 minutes to become familiar with the specification document and the last 30 minutes to detect as many defects with the document as possible. As is recommended for software inspections, each subject was provided with a checklist of generic types of analysis specification defects to aid in defect detection (Fagan, 1976). The checklist used for the experiment was prepared based on the checklists by Teague and Pidgeon (1985) and is presented in Appendix G. The format for recording individual findings differed slightly across treatments: Baseline groups recorded their findings using pencil and paper, Manual groups used marker pens and transparencies (several defects could be recorded on a single transparency), and EMS groups typed their responses into private window screens using the Group Outliner interface. As anonymity was not an independent variable of interest for this exploratory study, the transparencies and electronic files were coded such that it would be clear to all participants during the group review who was responsible for each finding. At the end of the individual review session, each participant completed a questionnaire (see Appendix H) and then was given a five minute break.
5.1.3.4 Group defect evaluation session

Following the break, each group performed the group evaluation review portion of the experiment. The purpose of the evaluation review was to create an aggregate group list of defects based on the findings from the individual review. Each team received written and verbal instructions describing the procedure to be followed (see Appendices D, E, and F). For each potential defect that was raised by an individual, the team was told to decide whether the defect was valid and belonged on the group list. Each team was told that they were not limited to the findings that were identified during the individual review; they were encouraged to identify new defects during the group review. Teams were informed that their performance would be based on the number of defects identified on the group list, with deductions for any incorrectly diagnosed defects. Subjects received classroom credit for the exercise based on their team’s performance. The author served as the moderator and scribe for all experimental sessions. The moderator followed the accepted practice of frequently soliciting contributions and input from all review participants (Fagan, 1976). All groups were able to complete the group review within the allotted time of 45 minutes. All group review sessions were audio tape recorded. Following the group review session, the subjects completed a second questionnaire (see Appendix I) and were released. Specific procedures for each of the treatments follow.
**Baseline treatment:** Different members of the group took turns verbally introducing and describing one of their findings to the group for approval (no specific pattern for member contributions was prescribed). If needed, the member who introduced a finding clarified the nature of the defect for the others. The group then decided whether the defect belonged on the group list. In the event that the finding was accepted, the moderator recorded the defect on the team's group list. This procedure was continued until the team had no more defects to discuss. To help group members keep track of their group's findings, the moderator read aloud the most recent findings at regular intervals (every 5-10 minutes).

**Manual treatment:** Prior to the group review, the overhead transparencies with the findings from the individual review were collected by the moderator. The moderator interleaved the transparencies such that the contributions from all members were positioned evenly throughout the stack of transparencies. The moderator began the group review by projecting the first transparency on the screen. Members of the group discussed each finding on the transparency and decided whether the defect belonged on the group list. The disposition of each finding (i.e., defect, repeat, or no defect) was noted by the moderator directly on the transparency. This procedure was continued until the team had no more defects to discuss.

**EMS treatment:** Prior to the group review, the electronic files from Group Outliner with
the findings from the individual review were collected by the moderator. The moderator interleaved the files into an aggregated Group Outliner file such that the contributions from all members were positioned evenly throughout the aggregate file. The review procedure for the EMS session was the same as that of the Manual treatment. A "chauffeured style" of EMS support was provided (i.e., the moderator made all keyboard entries into the EMS; see (Nunamaker, et al., 1991)).

5.1.4 Dependent variables

The dependent variables for Experiment 1 were individual and group performance and selected perceptions. Individual and group performance was measured by counting the number of distinct defects identified in the analysis specification document. The master list of defects for the specification document was used as the "answer key." The amount of process loss and gain occurring during the evaluation session was measured for each group by comparing the list of defects that had been found following the identification session with the list of defects that were reported on each team's "Action List" (i.e., the list of defects that each team turned in following the evaluation session as its final product for the exercise). Perceptions regarding the inspection exercise were measured using the post-identification and post-evaluation session questionnaires. The questionnaire items and scales were aimed at assessing the subjects' perceptions of the group defect evaluation process (e.g., satisfaction, sources of gain and loss) and the
intangible benefits of the process (e.g., understanding of the defects existing in the document).

5.2 Results

The results from Experiment 1 are organized into five major sections: 1) Group resources, 2) Characteristics of the task, 3) Regression analysis, 4) Group process, 5) Perceptions of the group process, and 6) Intangible benefits.

5.2.1 Group resources

The resource base for each group was determined by evaluating the performance of group members during the individual review. "Overlapping" group resources were defined as the defects that were identified by more than one individual in a group. For example, if the same defect was found by two team members it was counted as one overlapping defect. "Unique" group resources were defined as defects that were identified by only one individual. "Total resources" were defined to be the sum of the nonredundant overlapping and unique defects. Table 5.1 summarizes the means of overlapping, unique, and total resources for each treatment group. One-way analysis of variance (ANOVA) tests indicated no significant differences across treatment groups with respect to overlapping, unique, or total resources. Unique findings generally comprised
<table>
<thead>
<tr>
<th>Resources</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overlapping</td>
<td>Mean</td>
<td>7.8</td>
<td>11.0</td>
<td>9.4</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>3.3</td>
<td>3.3</td>
<td>2.4</td>
<td>0.29</td>
</tr>
<tr>
<td>Unique</td>
<td>Mean</td>
<td>7.6</td>
<td>5.2</td>
<td>6.8</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>2.9</td>
<td>1.9</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Total Resources =</td>
<td>Mean</td>
<td>15.4</td>
<td>16.2</td>
<td>16.2</td>
<td>0.20</td>
</tr>
<tr>
<td>Potential Performance</td>
<td>Std</td>
<td>1.7</td>
<td>1.6</td>
<td>3.3</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) Five groups per treatment cell
2) One-way ANOVA with (2,12) degrees of freedom

Table 5.1: Group resources

<table>
<thead>
<tr>
<th></th>
<th>Group Resources (No. of nonredundant defects detected)</th>
<th>Fraction of total defects existing within document</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>13</td>
<td>0.45</td>
</tr>
<tr>
<td>Mean (Std)</td>
<td>15.9 (2.2)</td>
<td>0.55</td>
</tr>
<tr>
<td>Maximum</td>
<td>20</td>
<td>0.69</td>
</tr>
<tr>
<td>Ratio of Maximum to Minimum</td>
<td>1.54</td>
<td>---</td>
</tr>
</tbody>
</table>

Notes: 1) N = 15 groups
2) Number of defects existing in document = 29

Table 5.2: Variations in group resources
a sizeable portion of group resources for teams in each treatment group as the mean portion of total resources that were unique ranged from 32 to 49 percent.

Although the subjects came from two different class sections, there did not appear to be any difference between sections with regard to group resources. ANOVA tests showed no significant differences between sections with regard to overlapping (F(1,13)=1.76, p=0.21), unique (F(1,13)=0.98, p=0.34), or total resources (F(1,13)=0.68, p=0.42). Additionally, the small variations in group sizes (i.e., twelve groups had four subjects, three groups had three subjects) did not appear to contribute to differences in group resources across groups since there was no measured correlation between group size and group resources (r squared=0.18, p=0.52).

How successful were the subjects in identifying the defects existing in the specification? As indicated in Table 5.2, the average amount of group resources for all fifteen teams was 15.9 defects, which corresponds to 55 percent of the defects existing in the specification. Thus, there did not appear to be a "ceiling effect" for the experimental inspection exercise since even the even the most effective team had group resources (20 defects) corresponding to only 69 percent of the defects in the specification. While a review of the group findings show that some defects were more likely to be identified than others, all of the defects in the specification appeared to be identifiable by some (if not all) of the subjects, as each defect was identified at least once across the pool of
The data in Table 5.2 also shows that there was a noticeable variation in detection performance with respect to both groups and individuals. With respect to group resources, the members of the best (i.e., most prolific) group found 54 percent more defects than the worst (i.e., least prolific) group. Although the spread in resources between the best and worst groups in this experiment may seem large, the difference is lower than that found for other published studies of inspections where the "defect detection" ratio comparing the best to worst teams were 4.7:1 (Martin and Tsai, 1990), 3:1 (Myers, 1978), and 2.3:1 (Schneider, et al., 1992). From an experimental perspective, a smaller variation in group resources across groups is desirable to reduce the error term due to inherent group differences. The lower spread in this study can probably be attributed to the relatively strict experimental conditions of this study which ensured that all subjects spent the same amount of time inspecting the document. Additionally, the random assignment of subjects to groups also may have reduced the chances for a large spread in resource differences across teams. The three studies mentioned above did not enforce controls for inspection time and did not randomly assign subjects to groups.

Differences in group resources across groups is a function of the variation in defect detection ability across individuals. In this study, the best subject (among the 57 subjects. 
subjects) found 14 defects during the individual review and the worst subject found 4 defects, resulting in a performance ratio of 3.5:1. This finding shows that large variations in demonstrated ability existed for the subjects in this study despite the fact that they had received similar training and were subjected to the same experimental conditions. This finding is consistent with findings by other researchers investigating the software engineering domain who have found large individual differences for subjects performing programming tasks (e.g., Sackman, et al., 1968; Shneiderman, et al., 1977) and inspection of programming code (Myers, 1978).

5.2.2 The characteristics of the task

To gain a sense of whether a task is a Eureka task, one may examine the frequency with which a group accepts a correct finding. If the frequency of acceptance is high then there is evidence that the task is a Eureka task. Data was collected regarding the acceptance rates for defects that were properly diagnosed by individuals and introduced during the group review. Table 5.3 shows the acceptance rates for defects that were properly diagnosed by individuals and introduced during the group evaluation review. The fractional acceptance rate of overlapping defects across treatments ranged from 96 to 100 percent and the fractional acceptance rate of unique defects ranged from 85 to 94 percent. These findings offer strong evidence that the experimental task may be viewed as a Eureka task, since the groups were very likely to accept a properly diagnosed defect
even if only one member had identified the defect.

A calculation of the overall acceptance rate of misdiagnosed defects suggests that the groups did a fairly good job of discriminating between properly diagnosed and misdiagnosed defects. Thirty misdiagnosed defects were introduced during review sessions and only six of them (20%) were accepted by the groups. Based on this low rate of acceptance, it appears that review groups did not simply provide "rubber stamp" approval to all defects that were introduced.

<table>
<thead>
<tr>
<th>Type of Resource</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overlapping</td>
<td>Mean</td>
<td>1.0</td>
<td>0.98</td>
<td>1.0</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.0</td>
<td>0.06</td>
<td>0.0</td>
<td>0.40</td>
</tr>
<tr>
<td>Unique</td>
<td>Mean</td>
<td>0.85</td>
<td>0.94</td>
<td>0.90</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.12</td>
<td>0.09</td>
<td>0.10</td>
<td>0.43</td>
</tr>
</tbody>
</table>

Notes: 1) Five groups per treatment cell
2) One-way ANOVA with (2,12) degrees of freedom

Table 5.3: Group acceptance rate of valid defects

Based on the laboratory data, it appears that the lower acceptance rate for unique defects may be attributed to the lack of "backup support." A review of the group session recordings indicates that the individuals who failed to gain acceptance for their unique defects did not adequately articulate the nature of the defect to their group and thus those
defects were not recognized by the group and were dismissed. Properly diagnosed overlapping defects, on the other hand, had the benefit of "backup support" and thus were almost always accepted (in this study only one properly diagnosed overlapping defect out of 142 was not accepted). Hence, while an overlap in findings did not increase total group resources, it did seem to increase the probability that a finding would be accepted by the group.

It was expected that the group review task could be described as a complementary task comprised of disjunctive subtasks. The empirical data supports this expectation. The complementary model is appropriate for addressing the overall group review task since it was evident that unique findings were generally spread about the team members of each group. On average, a single team member typically did not account for a large majority of the team’s unique findings, as the portion of a treatment group’s unique findings found by the most prolific team member ranged from 38 percent (EMS treatment groups) to 52 percent (Manual groups). For most of the teams (60 percent), all team members had the potential to contribute at least one unique finding to their group (i.e., they had found at least one unique finding during the individual identification session). Additionally, the subtasks involving the detection and acceptance of each proposed defect were seen to be disjunctive (as defined by Steiner (1972)) because optimal group performance was contingent on the group’s acceptance of the finding of the individual (or set of individuals) that had identified a valid defect.
5.2.3 Regression analysis

Steiner's theory predicts that actual group performance is a function of the task, group resources, and the group process. For this study the type of group task was held constant (i.e., the evaluation of individual findings), while the group resources varied randomly across groups, and the group process was manipulated across the three treatments by varying task support. The hypothesized relationships of resources and group process with actual performance (i.e., the dependent variable) were tested using hierarchical regression analysis. In this analysis, two blocks of variables representing the independent variables (i.e., group resources and group process) were entered sequentially into the regression model. The "group resources" block was represented by variables expressing the amount of unique and overlapping resources for each group (determined based on the individual review). The "group process" block was defined using dummy variables to represent the treatment followed by each group (i.e., Baseline, Manual, and EMS). The contribution of each block of variables was assessed by the significance of the F statistic associated with the change in the value of R squared, where R is the coefficient of multiple correlation. The change in R squared values reflect the amount of additional variance explained in the dependent variable as each block was entered into the regression.

The results of the analysis are presented in Table 5.4. The overall R squared for the
regression was very high (0.93), indicating that a large portion of the variance in actual group performance was explained by the regression equation. As hypothesized, group resources were found to be significantly related to actual performance, as the change in R squared value indicates that a substantial portion of the variance is explained by the "group resources" block of variables. Additionally, the beta weights for both unique and overlapping resources were significant. No support was found for the hypothesis that task support (i.e., group process) would be significantly correlated to group performance, as a relatively small portion of the variance (0.007) was explained by the "group process" block. Neither of the beta weights for the treatment variables were significant.

<table>
<thead>
<tr>
<th>Block</th>
<th>Independent Variables</th>
<th>Beta</th>
<th>p</th>
<th>Beta</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Group Resources</td>
<td>1. Overlapping</td>
<td>1.13</td>
<td>0.000</td>
<td>1.11</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>2. Unique</td>
<td>0.87</td>
<td>0.000</td>
<td>0.87</td>
<td>0.000</td>
</tr>
<tr>
<td>2. Group Process</td>
<td>1. Manual</td>
<td>0.52</td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. EMS</td>
<td>0.34</td>
<td>0.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R squared</td>
<td></td>
<td>0.920</td>
<td></td>
<td>0.927</td>
<td></td>
</tr>
<tr>
<td>Change R squared</td>
<td>0.920 *</td>
<td>0.007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>69.3</td>
<td>31.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d.f.</td>
<td>2,12</td>
<td>4,10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) N = fifteen groups  
2) *: change in R squared is significant (p < .001)

Table 5.4: Hierarchical regression of group resources and group process on group performance
Based on these findings, it appears that group resources is a critical determinant of group performance for the Eureka-type experimental task since the teams which had members who entered the evaluation session with the most findings from the individual session were the teams that performed the best. Although the results indicate that "group process" was not significantly related to performance, group process for the team evaluation task should not be necessarily be dismissed as an inert factor. As will be discussed below, although there were no significant differences with respect to the treatment block, the analysis of process loss and gain showed that the relative amount of process loss and gain did vary across treatments as loss and gain were low for groups receiving task support (i.e., the Manual and EMS groups), while loss and gain were higher for the Baseline group. Since the loss and gain tended to cancel out for the Baseline groups, the underlying impact of task support on the treatment groups was masked. Also, it should be noted that only one major element influencing group process was manipulated in this study: task support for the defect evaluation session. It may be possible that other types of group process mechanisms could result in a greater variation in group performance (e.g., see results related to the impact of process support discussed in Chapter 6).

5.2.4 Group process

Group process was investigated by comparing actual group performance to predicted
performance using Steiner's modified formula for group productivity (see equation 1 in Chapter 3).\footnote{For this discussion, Steiner’s term “productivity” is replaced by the term “performance,” since “performance” is viewed as being a more appropriate term for the context of this study.} The potential group performance for each team was estimated based on Steiner’s productivity models for complementary and disjunctive tasks (Steiner, 1972). Specifically, potential performance was calculated to be the total group resources (i.e., overlapping plus unique defects). Process losses were defined to be the number of valid findings uncovered during the individual reviews that were not utilized during the group review. Process gains were defined as valid defects that were newly identified during the group sessions and accepted by the group.

The relative performance of the treatment groups was evaluated by comparing the treatment means for potential performance, loss, and gain. To aid interpretation of the findings across treatments, the three components of Steiner’s modified model -- potential performance, loss, and gain -- are expressed as a fraction of actual performance (see Table 5.5). For all treatments, the mean potential performance was somewhat higher than that of the mean actual performance, indicating that there was a net process loss across all treatments. The mean ratio of potential to actual performance was similar across treatments (range: 1.02 to 1.09). Contrary to expectations, no significant difference between the Baseline groups and the groups receiving task support was found.
<table>
<thead>
<tr>
<th>Ratio</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Potential Performance</strong></td>
<td>Mean</td>
<td>1.09</td>
<td>1.02</td>
<td>1.03</td>
<td>1.36</td>
</tr>
<tr>
<td>Actual Performance</td>
<td>Std</td>
<td>0.11</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td><strong>Process Loss</strong></td>
<td>Mean</td>
<td>0.15</td>
<td>0.04</td>
<td>0.05</td>
<td>6.81</td>
</tr>
<tr>
<td>Actual Performance</td>
<td>Std</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Overlapping Findings:</td>
<td>Mean</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>Process Loss</td>
<td>Std</td>
<td>0</td>
<td>0.03</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Unique Findings:</td>
<td>Mean</td>
<td>0.15</td>
<td>0.03</td>
<td>0.05</td>
<td>8.27</td>
</tr>
<tr>
<td>Process Loss</td>
<td>Std</td>
<td>0.07</td>
<td>0.04</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Process Gain</td>
<td>Mean</td>
<td>0.06</td>
<td>0.02</td>
<td>0.01</td>
<td>1.30</td>
</tr>
<tr>
<td>Actual Performance</td>
<td>Std</td>
<td>0.06</td>
<td>0.03</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) Five groups per treatment cell
2) One-way ANOVA with (2,12) degrees of freedom

Table 5.5: Ratios of group potential performance, loss, and gain to actual performance

While the similar values for the ratio might suggest that the experimental manipulations did not influence the group review process, a closer examination of the data indicates that there was a significant difference with respect to process loss across treatments (see Table 5.5). The findings with respect to process loss and gain are explored below.
5.2.4.1 Process loss

As indicated in Table 5.5, there was a significant difference across treatments with regard to the fractional loss ratio, as the mean process loss for Baseline groups (15 percent of actual performance) was at least three times as much as that of the other two treatments (four and five percent). An explanation for the relatively high process loss for the Baseline groups may be found by examining the differences in process loss associated with the overlapping and unique findings (see Table 5.5). The process loss associated with overlapping findings was very low (range: zero to one percent of actual performance). Almost all of the process loss observed in the study can be traced to the failure of groups to fully utilize their unique findings.

Why was process loss typically associated with unique findings and why was this type of loss greater for the Baseline groups? Further examination of the data reveals that process loss involving unique findings only occurred during two scenarios. In the first scenario, a unique defect was introduced to the group for evaluation but was not accepted since the nature of the defect was poorly articulated by the group member (refer to the earlier discussion regarding the Eureka nature of the task). This scenario explains all of the process loss for the Manual and EMS groups and a portion of the process loss for the Baseline groups. This type of process loss did not appear to be influenced by the level of task support, as there were no major differences across treatment groups. Overall,
the mean process loss associated with this scenario (range: three percent for Manual to six percent for Baseline) was likely minimized due to the Eureka nature of the task.

In the second scenario, an individual had a unique finding but did not introduce the finding to the group. This type of scenario only occurred for the Baseline groups (because these groups had no task support mechanism to ensure that all resources would be made available to the group). Baseline groups experienced an average of nine percent additional fractional process loss arising from this scenario.

Why did some group members fail to introduce all of their findings to their group? It is doubtful that lack of motivation was a problem since the participants were aware that course credit was at stake and observations of the subjects indicated that all subjects were fully engaged in the review process. Also, it is unlikely that unfamiliarity with the group was an issue since group teams had spent much time working together on their course project. Problems associated with lack of team cohesion probably did not occur, as mean scores for the cohesion scale (measured on the post-individual review questionnaire) indicated that the teams were generally cohesive and liked their team members. Lack of opportunity for "air time" can also be ruled out as a factor since the members of all the Baseline groups were observed to have enough time to present and discuss their findings.

\[ \text{Mean score for the four item cohesion scale across the fifteen groups was 5.6 (std = 1.0) on a seven-point scale where 1 corresponds to low cohesion and 7 corresponds to high cohesion.} \]
findings within the 45 minute limit.

It is possible that some findings were not introduced by individuals during the Baseline sessions because of evaluation apprehension related to shyness or lack of confidence in specific findings. As indicated in Table 5.6, the subjects in the Baseline groups reported a greater degree of evaluation apprehension than subjects in the other treatment groups ($p=0.06$). The failure of some subjects to introduce their findings may have also contributed to another type of group process loss that was perceived by the subjects: free riding. A significant difference was found with respect to the responses to the questionnaire item relating to free riding with more free riding reported by the Baseline subjects ($p=0.005$). No significant differences across treatments were found with respect to questionnaire items relating to domination, information overload and sidetracking (see Table 5.6).

In addition to the measured perceived differences in evaluation apprehension and free riding, some individuals may have held back a finding due to the incorrect assumption that their finding had already been introduced to the group. Since no visual aids were used during the Baseline sessions, group members had to listen carefully to what the others were saying to keep track of the overall group findings. Even though participants were observed to be attentive, it was apparent that there were times when group members were uncertain whether or not a defect had been introduced (e.g., group members
occasionally prefaced their contributions with a remark such as "I'm not sure if this has been mentioned already, but I found ... [description of defect] .... ").

<table>
<thead>
<tr>
<th>Questionnaire Item/Scale</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation Apprehension (Q14)</td>
<td>Mean</td>
<td>2.7</td>
<td>1.5</td>
<td>1.7</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>2.1</td>
<td>1.0</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Free riding (Q12)</td>
<td>Mean</td>
<td>4.3</td>
<td>5.8</td>
<td>5.5</td>
<td>5.83</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.9</td>
<td>1.5</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Domination (Q17)</td>
<td>Mean</td>
<td>4.2</td>
<td>3.3</td>
<td>3.7</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.7</td>
<td>2.1</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>Side-tracking (Q25)</td>
<td>Mean</td>
<td>3.2</td>
<td>2.3</td>
<td>2.8</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.8</td>
<td>1.6</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Information Overload (Q23)</td>
<td>Mean</td>
<td>2.8</td>
<td>2.6</td>
<td>2.6</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.8</td>
<td>1.6</td>
<td>1.5</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) See Appendix I (Post-Evaluation Session questionnaire) for specific questionnaire items cited in the above table
2) Questionnaire scale: Likert scale from 1 (strongly disagree) to 7 (strongly agree)
3) One-way ANOVA with (2,54) degrees of freedom

Table 5.6: Perceptions of potential sources of group process loss

5.2.4.2 Process gain

The presence of process gain for the treatment groups indicates that positive contributions can result from interaction during the group defect evaluation process (see Table 5.5). However, the process gain was not large for the Baseline groups (six percent of actual performance) and was negligible for the Manual and EMS groups (two and one percent,
respectively). The relatively low gain for the Manual and EMS groups was somewhat surprising, since all treatment groups in the study were encouraged to generate new defects during the group session. No significant difference was found across treatments with respect to gain.

Although there were no significant statistical differences across treatments, observations of the group behavior made during the review sessions suggest that there was a difference in group behavior across the treatments. The Manual and EMS groups were observed to be preoccupied with evaluating the findings presented on the visual display. The review sessions for these groups took on a very structured atmosphere as the groups stepped through each finding on the visual display and evaluated the acceptability of each finding. Since the focus was on evaluation, almost no new defects were detected. The atmosphere during the Baseline reviews was less structured since there was no visual display to impose order on the review process. Lacking a visual device to focus the attention of the group on the evaluation of specific findings, these groups would occasionally shift their attention to group discussions regarding the search for new defects. Based on these observations, it appears that the visual display served to restrict groups to an "evaluation mode" and thus hindered the identification of new findings during the review session. The Baseline groups were not restricted in this way and were therefore more likely to produce new findings.
5.2.5 Perceptions of the group process

Table 5.7 presents the results from the post-group review questionnaire. To assess whether the subjects had favorable perceptions with regard to satisfaction, acceptance, and effectiveness a series of t-tests were performed (i.e., mean perceptions were compared to a rating corresponding to "neutral"). Perceptions were found to be significantly more favorable than "neutral" for all measured variables across all treatments.

ANOVA tests were performed to investigate whether there were differences in perceptions across treatments. No significant differences across treatments were found with respect to satisfaction or acceptance. A marginal significant difference across treatments was found with respect to perceived group evaluation session effectiveness (p=0.09). Planned t-tests show that the effectiveness ratings for the Baseline groups were significantly lower than those for the Manual and EMS groups. There were no differences between the Manual and Baseline groups. Differences with respect to effectiveness may be due to the fact that subjects in the Manual and EMS groups were aware that all of their group resources had been considered by their group (due to the task support), while those in the Baseline groups may have had some doubts about whether all of their resources had been presented and received proper consideration.
The perceived effect of the scribe on the review sessions was measured to ensure that the study results were not impacted by scribe behavior. Survey findings indicate that the influence of the scribe was minimal and that there was no significant difference across treatments.

<table>
<thead>
<tr>
<th>Questionnaire Item/Scale</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction (Q1, Q2; alpha=0.90)</td>
<td>Mean</td>
<td>5.5</td>
<td>6.2</td>
<td>5.9</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.2</td>
<td>0.9</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Acceptance (Q3, Q4; alpha=0.80)</td>
<td>Mean</td>
<td>6.0</td>
<td>6.4</td>
<td>6.3</td>
<td>1.41*</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.1</td>
<td>0.7</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Effectiveness (Q7)</td>
<td>Mean</td>
<td>5.5</td>
<td>6.1</td>
<td>6.2</td>
<td>2.57</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.1</td>
<td>0.9</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Scribe Influence (Q27, Q28; alpha=0.70)</td>
<td>Mean</td>
<td>1.8</td>
<td>1.5</td>
<td>1.6</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.0</td>
<td>0.8</td>
<td>1.1</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) See Appendix I (Post-Evaluation Session questionnaire) for specific questionnaire items cited in the above table
2) Questionnaire scale: Likert scale from 1 (strongly disagree) to 7 (strongly agree)
3) For multi-item scales the mean score is reported
4) Scale reliability is reported using Cronbach’s alpha
5) Five groups per treatment cell
6) One-way ANOVA with (2,54) degrees of freedom
7) *: A Kruskal-Wallis nonparametric statistic with 2 degrees of freedom was used to evaluate the acceptance scale due to violation of the homogeneity of variance assumption

Table 5.7: Perceptions of the group process

5.2.6 Intangible benefits

Table 5.8 presents the responses to several questionnaire items that were posed on both the post-individual and post-group evaluation session questionnaires (individual responses
were aggregated across treatments). This data was collected to investigate the claims that intangible benefits such as increased understanding of defects and confidence in findings are associated with the group review process. For all items listed in Table 5.8, a significant gain was measured. For the types of subjects participating in this study (i.e., novices), these results support the claims made in the literature regarding the intangible benefits associated with the group review process.

A second analysis was performed to assess whether the observed gains were a function of a subject’s performance on the individual review. Based on their performance in the individual review, the subjects were grouped into three performance categories: low (28% of the subjects), medium (34%), and high (38%). T-test comparisons were made for subjects in each subgroup (see Table 5.8). The analysis shows that intangible benefits arising from the group process were not limited to the least proficient subjects, as significant gains were present for each subgroup for two out of the four items examined: understanding of defects and confidence that all defects have been identified.
<table>
<thead>
<tr>
<th>Questionnaire Item</th>
<th>Pre-Indiv Review</th>
<th>Post-Group Review</th>
<th>t-test Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (df=56)</td>
<td>Mean (df=14)</td>
<td>Low (df=19)</td>
</tr>
<tr>
<td>Understanding of Defects (Q18)</td>
<td>4.9 1.2</td>
<td>5.8 1.1</td>
<td>5.90 .000</td>
</tr>
<tr>
<td>Confidence w/regard to:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abilities as a reviewer (Q21)</td>
<td>Mean 5.2 1.1</td>
<td>5.6 1.1</td>
<td>3.03 .004</td>
</tr>
<tr>
<td>All defects have been identified (Q5)</td>
<td>Mean 3.4 1.8</td>
<td>4.7 1.6</td>
<td>5.27 .000</td>
</tr>
<tr>
<td>All findings are correct (Q6)</td>
<td>Mean 5.4 1.3</td>
<td>5.7 1.3</td>
<td>1.79 .08</td>
</tr>
</tbody>
</table>

Notes: 1) See Appendix I (Post-Evaluation Session questionnaire) for specific questionnaire items cited in the above table
2) Questionnaire scale: Likert scale from 1 (strongly disagree) to 7 (strongly agree)

Table 5.8: Perceived intangible impacts of the group process
CHAPTER 6: EXPERIMENTAL STUDY 2: METHOD AND RESULTS

This chapter describes the research method and the results of the second experimental study that was performed. This study had the following objectives: 1) Evaluate the impact of selected process support mechanisms on the performance and perceptions of inspection teams working interactively on the defect identification subtask, and 2) Evaluate the impact of a simple task support mechanism on the performance of perceptions of inspection teams conducting the evaluation subtask.

6.1 Method

6.1.1 Task

The task used for Experiment 2 was the same as the task used for Experiment 1 (see Appendix A). As noted earlier, the task required the subjects to identify defects existing within an "analysis specification document." The specification document described the user requirements for a subsystem of a fictional real estate company and included three main parts: a descriptive narrative section, a data dictionary, and three data flow diagrams (DFDs).
6.1.2 Subjects

The participants were 118 upper-division undergraduate MIS students from four sections of a semester-long course in Systems Analysis and Design at the University of Arizona. Students from this course were selected since the coursework provided an appropriate background for addressing the experimental task. Almost all (98 percent) of the students enrolled in the four class sections participated in the experiment. The mean age of the subjects was 24 years old (std=4.5). Seventy nine were male (67%) and thirty nine were female (33%). A large majority of the participants (82%) reported work experience (full-time or part-time).

The subjects were randomly assigned into thirty groups (twenty eight four-person teams and two three-person teams). As was the case for Experiment 1, team sizes of three to four people were used based on literature recommendations regarding the optimal size of software review teams (Fagan, 1976). Prior to the experiment, each group had spent six to seven weeks working together on an extensive group term project involving the analysis of a business system. The laboratory study was conducted several weeks after the teams had been assigned to ensure that each team had an opportunity to "develop as a team."

The four class sections participating in the experiment were taught by three different
instructors (two of the sections were taught by the author). To help minimize potential differences in the subject pool due to "instructor effects," the instructors coordinated their course material such that both sections of the course were taught using the same textbook and the lecture material for each course section was very similar. Groups from each course section were distributed as evenly as possible across treatments in a random fashion. Following the experiment, an ANOVA was performed to ascertain whether there was a difference in group performance (i.e., total number of defects found by group following the interactive group review) across course sections. No significant difference was found ($F(3,26)=1.42$, $p=0.26$), suggesting that the subject pool was relatively homogeneous with regard to ability to perform the task.

Although Experiment 1 was performed during the Fall semester of 1991 and Experiment 2 was performed during the Spring and Fall semesters of 1992, there is no reason to believe that the subject pools that participated in the two experiments were substantially different from one another with regard to demographics, educational background, or instruction. A comparison of the participant demographics shows that the mean age, gender mix, and work experience are similar across semesters. As all subjects for Experiments 1 and 2 were drawn from sections of the same MIS course (MIS 341: Systems Analysis and Design), their educational background was similar (since students at the University of Arizona typically take the MIS 341 course at a predefined point in their curriculum sequence). The instructional style and content of the course was also
very similar across semesters and instructors. The similarities between the subject pools offer qualitative support for the notion that the subject pools are essentially the same. If it can be assumed that the subject pools are the same across experiments, then it is possible to compare findings across experiments to gain insight into the study results (e.g., the mean potential performance of groups in Experiment 1 may be used as a baseline for potential performance of groups in Experiment 2).

6.1.3 Procedure

To address the research questions, this experiment was designed to allow measurement of the following: 1) The actual performance of each review team following the interactive defect identification session and the evaluation sessions, and 2) Individual perceptions regarding the inspection process. The experimental procedure used to capture this information consisted of two major portions: 1) An interactive defect identification session, and 2) A team defect evaluation session to discuss and evaluate the individual findings.

The specific review procedures followed for this study were based on Fagan’s inspection approach (Fagan, 1976). However, due to experimental design considerations there were two ways in which the experimental review procedure differed from Fagan’s approach: 1) Fagan’s "Reader" roleplayer was not included in the group reviews (however, Fagan’s
"Moderator" roleplayer was included), and 2) Two of the three experimental treatments involved the use of process support and task support mechanisms (Fagan's approach does not include these types of support mechanisms).

Experiments 1 and 2 were similar with regard to several aspects (e.g., task, subjects). However, due to differing research objectives the procedures for the experiments were different. The primary difference between the procedures for Experiments 1 and 2 was that the defect identification session was performed individually in Experiment 1, while the defect identification session was performed interactively for Experiment 2. The procedure for Experiment 2 was designed to correspond to Fagan's (1976) inspection procedure in which members of an inspection team do not examine a work product for defects until the group meeting.¹ Experiment 2 was thus concerned with two interactive review sessions per group of subjects (i.e., the defect identification and evaluation reviews), while Experiment 1 examined only one interactive review session (i.e., the evaluation review). Specific procedures regarding Experiment 2 are described below.

6.1.3.1 Experimental manipulations

The experimental manipulations were designed to investigate the impact of selected

¹ Fagan's (1976) inspection guidelines state that the objective of the preparation step is "education" regarding the work product, while the objective of the team inspection is to "find errors" (p. 200).
combinations of process and task support mechanisms on the performance of inspection teams. Support for the groups performing the review of the system specification varied across three experimental treatments. As the review process for each treatment was divided into two sessions -- defect identification and defect evaluation -- the type and amount of group support depended on the session. Different types of process support were provided for groups during the identification session, while different types of task support were provided for groups during the evaluation session.

**Defect identification session**: Process support was manipulated across three experimental treatments during the identification session. The "Baseline" treatment groups were not provided with any process support mechanisms during the interactive identification session. Baseline group members were required to speak up and introduce their findings to the group orally. For this treatment no visual aids were used to record group findings and no electronic media were used. The other two treatments did provide support for the teams. The "Manual" treatment provided an intermediate amount of process support through the use of a public display that was used to record findings generated by the group (i.e., group memory). Findings were recorded by a neutral moderator who wrote the findings onto sheets of white paper at the front of the review room. The "EMS" treatment provided more substantial process support by means of an EMS. Specifically, the EMS provided support for both parallel communication and group memory. The University of Arizona's GroupSystems "Group Outliner" tool was used to provide
process support (for a description of GroupSystems and Group Outliner see Nunamaker, et al., 1991). Groups in the EMS treatment used the EMS’s "interactive style" of process support (i.e., interaction via the EMS electronic communication channel only, no talking; see Nunamaker, et al, 1993) and had access to all findings generated by other members of their team.

**Defect evaluation session:** Task support was manipulated across the three experimental treatments for the evaluation session. The "Baseline" treatment did not provide any task support for review groups as group members were required to speak up and reintroduce the findings that they had identified during the identification session. For this treatment no task support mechanisms (e.g., visual aids) were used and there was no guarantee that all findings identified during the identification session would be introduced to the group. The other two treatments did provide a task support mechanism. The "Manual" treatment provided task support by means of the handwritten lists on the white sheets that were developed during the identification session. The "EMS" treatment provided task support by providing group members with access to the electronic textual files generated during the identification session. Access was provided in two ways: 1) Group members could peruse the files at their individual workstations, and 2) The files were projected onto an overhead screen using the EMS and a BARCO projector. The Manual and EMS treatments were designed to ensure that all findings generated during the individual sessions were brought before the group through the use of visual displays.
6.1.3.2 Pre-experimental procedure

Each group was randomly assigned to one of the three treatment groups described above. Upon reporting to the experimental site, each group member completed a pre-exercise questionnaire (see Appendix M). Following completion of the questionnaire, the team was provided with a brief overview relating to the objective and procedures associated with a software inspection (all students had received a class lecture on this subject prior to the experiment). See Appendix C for the handout packet that was reviewed during the overview (the same overview was used for Experiments 1 and 2). During the overview, it was emphasized that the purpose of a review was to detect defects in the specification and not to correct defects. The subjects were told that the review would be divided into two interactive sessions: 1) An defect identification session, and 2) A defect evaluation session.

6.1.3.3 Defect identification session

Following the overview, each group performed the interactive defect identification review. The purpose of the identification session was to generate a list of potential defects associated with the system specification. Each team received written and verbal

---

2 The questionnaires used for Experiment 2 were more in-depth than those used for Experiment 1, as more questionnaire items were included.
instructions describing the procedure to be followed for the identification session. See Appendices J, K, and L for the instructions provided for each treatment. Subjects were told that their efforts would be evaluated to determine course credit for the exercise. Any questions regarding the procedure were answered. The identification session was divided into two portions: 1) A 15 minute individual review of the specification document, and 2) A 30 minute interactive review. No groups were permitted to exceed the time limits.

The purpose of the individual review was to allow individuals an opportunity to read through the document and become familiar with it. Fifteen minutes for the individual review was allotted as observations of subjects participating in the Pilot Study and Experiment 1 indicated that most people required about 15 minutes to read through the document before they were prepared to detect defects. The 15 minute time limit for the individual review was established to ensure consistency across groups with respect to preparation time.

Following the 15 minute individual review portion, groups were instructed to begin the interactive defect identification review of the specification document. Participants were encouraged to identify as many defects as possible without being concerned about whether their findings were correct, as the findings would be discussed later during the evaluation session. As was done for Experiment 1, each subject was provided with a
checklist of generic types of defects to aid in the detection of defects (see Appendix G). The author served as both the scribe and moderator for all experimental sessions. The degree of process support associated with each treatment varied across treatments (see below). All identification sessions were audio tape recorded. At the end of the interactive defect identification session, each participant completed the post-defect identification session questionnaire (see Appendix N) and then was given a five minute break. Specific procedures for each treatment follow.

**Baseline treatment:** Members of the group were instructed to examine the system specification for potential defects and to orally introduce the potential defect to the group whenever a new potential defect was uncovered. If needed, the subjects were allowed to discuss the nature of their finding to the others for purposes of clarification. The subjects in the Baseline treatment were asked to keep a personal list of findings that they had introduced to the group. Each entry on the list included a short description of the defect plus a code indicating the defect classification defined in the checklist (e.g., completeness, consistency). (Note: The subjects were asked to record only their own findings; they did not record defects identified by other members of their group). Each personal list was recorded using pencil and paper. The moderator was present during

---

3 It has been noted by Bouchard (1969) that group performance for brainstorming tasks may be influenced by whether the participants verbalize or write down their findings. To reduce the chances for this type of confound, in this study a written record was made for each finding generated during the identification session. For the Baseline and EMS groups, the written record was made by the participant who generated the finding. For the Manual groups, the written record was made by the scribe on a public display device.
the session but did not speak unless it was necessary to enforce procedure.

**Manual treatment:** Members of the group were instructed to examine the system specification for potential defects and to orally introduce their findings to the group whenever a new potential defect was uncovered. If needed, the subjects were allowed to discuss the nature of their finding to the others for purposes of clarification. Findings for the Manual groups were recorded by a scribe (the researcher) on a public display consisting of large sheets of white paper taped to the wall at the front of the meeting room. To help organize the findings, a separate white sheet was used for each of the four categories of defects defined in the checklist. Subjects in the Manual treatment relied on the scribe to record potential defects identified by the group and did not maintain a personal list of their findings.

**EMS treatment:** Groups in the EMS treatment were only permitted to interact using the electronic communication channel of the EMS (i.e., no oral interaction). Each member of the review group was seated in front of an EMS workstation and group members were seated spaced apart at every other workstation to minimize the potential for verbal interaction. Members of the group were instructed to examine the system specification for potential defects and to orally introduce the potential defect to the group whenever a new potential defect was uncovered. Subjects in the EMS groups typed their findings into window screens using the Group Outliner interface. To help organize the findings,
a separate Outliner window (i.e., repository) was used for each of the four categories of defects defined in the checklist. Subjects were instructed to enter their finding into the window representing the most appropriate defect category. Additionally, subjects were asked to precede all electronic comments with their name so that it was clear to all participants who was responsible for each finding (i.e., elimination of anonymity). During the identification session, subjects could browse through the windows to view the findings that had been generated by the group.

6.1.3.4 Defect evaluation session

Following the break, each group performed the defect evaluation portion of the experiment. The purpose of the evaluation review was to develop a group list of defects based on the findings from the identification session. Each team received written and verbal instructions describing the procedure to be followed. For each potential defect that had been identified during the earlier session, the team was told to decide whether the defect was valid and belonged on the group list. Each team was told that they were not limited to the findings that were identified during the identification session; they were encouraged to identify new defects during the evaluation session. Teams were informed that their performance would be based on the number of defects identified on the group list, with deductions for any incorrectly diagnosed defects. The primary author served as both the scribe and moderator for all experimental sessions. Each team was allotted
45 minutes for the evaluation session. All groups were able to complete the evaluation review within the allotted time. All evaluation sessions were audio tape recorded. Following the evaluation session, the subjects completed the post-evaluation session questionnaire (see Appendix O) and were released. Specific procedures for each treatment follow.

**Baseline treatment:** Different members of the group took turns verbally introducing and describing one of their findings to the group for approval. To help organize the session, group members were asked to introduce their defects in the order of the categories listed in the checklist, beginning with completeness defects. Once all the defects that were judged by the group to be completeness defects were introduced, the group proceeded to introduce (in order) consistency, communicability, and correctness defects. If needed, the member who introduced a finding clarified the nature of the defect for the others. The group then decided whether the defect belonged on the group list. In the event that the finding was accepted, the moderator recorded the defect on the team’s group list along with its defect category. The team’s group list was recorded on a standard size of paper and was not recorded on a public display. This procedure was continued until the team had no more defects to discuss. For this study the moderator followed the accepted practice of frequently soliciting contributions and input from all review participants (Fagan, 1976). To help group members keep track of their group’s findings, the moderator read aloud the most recent findings at regular intervals (every 5-10 minutes).
Manual treatment: The public display list generated during the identification session was reviewed during the Manual evaluation session. The moderator facilitated the session by reading off the items (i.e., potential defect) on the public display list one at a time and asking for comments regarding the group's evaluation of each item. The moderator went through the public display list in order of defect category, beginning with completeness defects and continuing through consistency, communicability, and correctness defects. Based on the discussion, the group decided whether each potential defect belonged on the group list. This procedure was continued until there were no more potential defects on the list to evaluate. For this study the moderator followed the accepted practice of frequently soliciting contributions and input from all review participants (Fagan, 1976).

EMS treatment: The electronic Group Outliner files (i.e., windows) generated during the identification session were reviewed during the EMS evaluation session. The Group Outliner files were available to the review group in two ways: 1) Through a central display projected onto a screen at the front of the room (this display was controlled by the moderator), and 2) Through each workstation (each of the workstation displays were independent of the central display and could be controlled by the group member). The moderator facilitated the session by reading off the items (i.e., potential defect) listed in each file one at a time and asking for comments regarding the group's evaluation of each item. The moderator went through the electronic files in order of defect category,
beginning with file containing defects designated as completeness defects and continuing through the files containing consistency, communicability, and correctness defects. Based on the discussion, the group decided whether each potential defect belonged on the group list. This procedure was continued until there were no more potential defects on the list to evaluate. For this study the scribe followed the accepted practice of frequently soliciting contributions and input from all review participants (Fagan, 1976).

6.1.4 Dependent variables

The dependent variables for Experiment 2 were group performance and selected perceptions. Group performance following the defect identification and evaluation sessions was measured by counting the number of distinct defects identified in the analysis specification document. The master list of defects for the specification document was used as the "answer key." The amount of process loss and gain occurring during the evaluation session was measured for each group by comparing the list of defects that had been found following the identification session with the list of defects that were reported on each team's "Action List" (i.e., the list of defects that each team turned in following the evaluation session as its final product for the exercise). Perceptions regarding the exercise were measured using three questionnaires: the pre-exercise, post-identification session, and post-evaluation session questionnaires. The questionnaire items and scales were aimed at assessing various aspects of the subject's perceptions of the group process
including confound checks, process loss and gain issues, and general perceptions (e.g., satisfaction, effectiveness of team).

6.2 Results

The results from the experiment are organized into three major sections: 1) Performance and perceptions regarding the defect identification session, 2) Comparison of group versus individual defect identification (i.e., a comparison of identification findings from Experiments 1 and 2), and 3) Performance and perceptions regarding the defect evaluation session.

6.2.1 Defect identification session: Performance and perceptions

6.2.1.1 Confound checks

Random assignment of subjects to groups and groups to treatments was done to control for differences across groups with respect to individual and group characteristics. Selected items were included on the pre-exercise questionnaire to provide a check for differences across treatment groups with regard to each subject’s perception of their ability to review an analysis specification document and the cohesion of their group (a four-item scale described by Price and Mueller (1986) was used for group cohesion;
Cronbach’s alpha = 0.77). A one-way analysis of variance (ANOVA) test indicated that there were no significant differences across treatment groups for either of these factors (see Table 6.1). While a check of these selected measures does not ensure homogeneity across treatments, it does help to increase confidence that the randomization procedure helped to minimize differences across treatment groups with regard to these measures.

<table>
<thead>
<tr>
<th>Questionnaire Item/Scale</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have confidence in abilities as a reviewer (Q9)</td>
<td>Mean</td>
<td>5.8</td>
<td>5.4</td>
<td>5.6</td>
<td>1.71</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.0</td>
<td>0.9</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Team cohesion (Q1, Q2, Q3, Q4; alpha = 0.77)</td>
<td>Mean</td>
<td>5.7</td>
<td>5.5</td>
<td>5.8</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.0</td>
<td>0.8</td>
<td>0.8</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) See Appendix M (Pre-Exercise questionnaire) for specific questionnaire items cited in the above table  
2) Questionnaire scale: Likert scale from 1 (strongly disagree) to 7 (strongly agree)  
3) For multi-item scales the mean score is reported  
4) Scale reliability is reported using Cronbach’s alpha  
5) One-way ANOVA with (2, 115) degrees of freedom

Table 6.1: Pre-exercise perceptions

Efforts were made during the design and conduct of the experiment to minimize effects due to the following potential confounds on group performance: the influence of the scribe/moderator, anonymity, degree of personal effort, adequacy of review time, task focus during the session, typing skills, and ease of use of the EMS. Survey items were included on the post-identification questionnaire to help assess the subjects’ perceptions.
regarding these factors. Table 6.2 presents the means, standard deviations, and ANOVA results of subjects' perceptions regarding the questionnaire items. As discussed below, the questionnaire results suggest that the potential confounds identified above were not present in the experiment.

Influence of scribe/moderator: Subjects' perceptions regarding the influence of the scribe/moderator were measured using a two-item scale (Cronbach's alpha=0.82). For all treatment groups the scribe/moderator was viewed as having a minimal influence on the team inspection and no significant difference was found with regard to this measure.

Anonymity: The presence of anonymity during the identification session was checked by one item which asked subjects to indicate whether they "always knew the identity of the person who had introduced a finding to the team." The mean scores for this item indicate that anonymity was generally not present in any of the treatment groups. This finding was expected since the Baseline and Manual treatments involved face-to-face interaction of small intact (i.e., "ongoing") groups and the subjects in the EMS treatment attached their names to all contributions. An ANOVA test found that there was a significant difference across treatments with regard to this measure (p=0.05). A post-hoc Tukey test (alpha=0.05) indicates that subjects in the Baseline treatment were less likely to know the identity of the person who introduced a finding to the group than subjects in the other two treatments. A significant difference across treatments was not
expected. One explanation for this result may be that the lack of support mechanisms for the Baseline groups made it a bit more difficult for these groups to keep track of who contributed specific findings to the session.

<table>
<thead>
<tr>
<th>Questionnaire Item/Scale</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scribe/moderator influence (Q37, Q38; alpha = 0.76)</td>
<td>Mean</td>
<td>1.7</td>
<td>2.0</td>
<td>2.1</td>
<td>1.24</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.2</td>
<td>1.0</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Knew identity of contributor (Q31)</td>
<td>Mean</td>
<td>5.7</td>
<td>6.3</td>
<td>6.4</td>
<td>3.08</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.4</td>
<td>1.1</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>I put forth my best effort (Q16)</td>
<td>Mean</td>
<td>5.8</td>
<td>6.2</td>
<td>6.1</td>
<td>1.15</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.1</td>
<td>0.6</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Had adequate time for identification review (Q32)</td>
<td>Mean</td>
<td>3.5</td>
<td>3.7</td>
<td>3.9</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>2.1</td>
<td>1.8</td>
<td>1.9</td>
<td></td>
</tr>
<tr>
<td>Team's effort was spent on defect detection (Q33)</td>
<td>Mean</td>
<td>5.8</td>
<td>5.7</td>
<td>5.9</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.8</td>
<td>1.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>Team's effort was spent on defect correction (Q34)</td>
<td>Mean</td>
<td>2.9</td>
<td>2.1</td>
<td>2.9</td>
<td>3.42</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.6</td>
<td>1.1</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>Team's effort was spent on defect evaluation (Q35)</td>
<td>Mean</td>
<td>3.6</td>
<td>3.1</td>
<td>3.3</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.7</td>
<td>1.4</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>Typing skills did not impair ability to record (Q54)</td>
<td>Mean</td>
<td>NA</td>
<td>NA</td>
<td>6.0</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>NA</td>
<td>1.5</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Ease of use of EMS (Q55, Q56; alpha = 0.80)</td>
<td>Mean</td>
<td>NA</td>
<td>NA</td>
<td>6.6</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>NA</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) See Appendix N (Post-Identification Session questionnaire) for specific questionnaire items cited in the above table
2) Questionnaire scale (Q1 to Q40): Likert scale from 1 (strongly disagree) to 7 (strongly agree)
3) For multi-item scales the mean score is reported
4) Scale reliability is reported using Cronbach’s alpha
5) The Likert scale for Q56 was reversed to make the 2-item scale for “Ease of use of EMS” consistent
6) NA = not applicable
7) One-way ANOVA with (2, 115) degrees of freedom

Table 6.2: Perceptions of the identification session: Confound checks
Degree of personal effort: Subjects from all treatment groups indicated that they put forth a strong effort during the defect identification session. No significant difference across treatments was found.

Adequacy of review time: The mean ratings regarding the adequacy of review time suggest that the time allotted to the task was appropriate, as groups did not feel an undue amount of time pressure during the identification review. There was no significant difference with regard to adequacy of review time which suggests that the slight variations in procedure across treatments did not significantly impact the amount of time that the subjects could devote to the detection of defects.

Task focus: Questionnaire items were included to check whether the teams in each treatment group were focussed on the proper group task (i.e., defect detection). Subjects were asked to indicate whether their team spent a significant amount of effort on each of the following: defect detection, correction, and evaluation. As desired, the responses indicate that the teams focussed most of their effort on defect detection. There were no significant differences across treatments with respect to perceived effort devoted to defect detection and evaluation. However, there was a significant difference with respect to defect correction (p=0.04). A post-hoc Tukey test (alpha=0.05) indicated that subjects in the Manual groups perceived that they had expended less effort on defect correction than was reported by subjects in the other treatment groups.
Typing skills: The method used to record findings varied across treatments. While a hand written approach was used to record findings for the Baseline and Manual groups, the subjects in the EMS groups needed to type their findings into an electronic file. A questionnaire item was included to assess whether typing impaired the recording of findings for the EMS subjects. On a seven-point scale anchored at 1 (strongly disagree) and 7 (strongly agree), the EMS subjects averaged a score of 6.0 (std=1.5) for "My typing skills did not impair my ability to record my findings." Based on these perceptions, it appears that typing skills was not a confound for this study.

Ease of use of EMS: The usability of the EMS could also potentially impact the findings of this study by making it more difficult for the EMS subjects to record their findings (i.e., using a computerized recording system may be more difficult than using a hand written approach). A two-item scale was used to assess the ease of use of the EMS (Cronbach’s alpha=0.80). The mean score for this scale (averaged across the two items) was 6.6 (std=0.6) on a scale of 1 (strongly disagree) to 7 (strongly agree), which indicates that the subjects strongly agreed that the computerized system was easy to use. Hence, EMS usability does not appear to have been a confound for this study.4

4 It may also be noted that a large majority of the EMS subjects (73 percent) had used the GroupSystems EMS at least once prior to this experiment. The subjects’ familiarity with the EMS environment may have contributed to the high usability ratings.
6.2.1.2 Group performance and perceptions

Group performance

Table 6.3 presents the mean number of defects identified for each treatment group, standard deviations, and the ANOVA results. On average, the EMS groups found the most defects (18.0 defects; 62 percent of the twenty nine defects existing in the specification), followed by the Manual groups (14.8 defects; 51 percent), and the Baseline groups (13.3 defects; 46 percent). As predicted, a significant treatment effect was found with respect to the number of defects identified by the groups during the defect identification session (p=0.005). When viewed in relative terms, the difference in the mean number of defects detected across the Baseline and EMS groups was sizeable, as the EMS groups found 35 percent more defects than the Baseline groups. A post-hoc Tukey test (alpha=0.05) showed that the EMS groups identified significantly more defects than the Baseline groups. The Tukey tests indicated that there were no significant differences between the Baseline and Manual groups or the Manual and EMS groups.
### Group Performance

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Post-Identification Session Defects</td>
<td>Mean 13.3 Std 2.8</td>
<td>14.8 2.7</td>
<td>18.0 3.4</td>
<td>6.49</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Mean 13.0 Std 3.1</td>
<td>15.3 2.8</td>
<td>17.7 3.3</td>
<td>5.78</td>
<td>0.008</td>
</tr>
<tr>
<td>Process Loss</td>
<td>Mean 0.9 Std 0.3</td>
<td>0 0</td>
<td>0.5 0.7</td>
<td>10.17</td>
<td>0.001</td>
</tr>
<tr>
<td>Process Gain</td>
<td>Mean 0.6 Std 0.5</td>
<td>0.5 0.7</td>
<td>0.2 0.4</td>
<td>1.38</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Notes: 1) Ten groups per treatment cell  
2) One-way ANOVA with (2,27) degrees of freedom

**Table 6.3: Number of defects detected**

### Perceptions of potential sources of process loss and gain

To gain insight into the reasons underlying the productivity differences across treatments, the post-identification session questionnaire asked subjects to respond to several survey items relating to the sources of process loss and gain. Table 6.4 presents the means, standard deviations, and ANOVA results of subjects' perceptions regarding the pertinent questionnaire items. The findings for specific measures are discussed below.

**Domination**: Domination was measured using a two-item scale (Cronbach's alpha=0.80). There was a significant difference across treatments (p=0.03). A post-hoc Tukey test (alpha=0.05) showed that the differences across treatments partially
### Table 6.4: Perceptions of potential sources of group process loss and gain for the identification session

<table>
<thead>
<tr>
<th>Questionnaire Item/Scale</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Potential sources of process loss:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domination (Q21,Q22; alpha=0.80)</td>
<td>Mean</td>
<td>3.8</td>
<td>3.2</td>
<td>2.9</td>
<td>3.75</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.5</td>
<td>1.6</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>Side-tracking (Q30)</td>
<td>Mean</td>
<td>3.5</td>
<td>2.3</td>
<td>2.5</td>
<td>7.91</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.7</td>
<td>1.2</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Failure to remember (Q39,Q40; alpha=0.87)</td>
<td>Mean</td>
<td>2.3</td>
<td>2.9</td>
<td>4.2</td>
<td>8.65</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.2</td>
<td>1.4</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>Information overload (Q27,Q28; alpha=0.74)</td>
<td>Mean</td>
<td>3.3</td>
<td>3.1</td>
<td>3.9</td>
<td>2.97</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.4</td>
<td>1.3</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td>Production blocking (Q41,Q42,Q43; alpha=0.73)</td>
<td>Mean</td>
<td>2.7</td>
<td>3.0</td>
<td>2.2</td>
<td>2.21</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.4</td>
<td>1.2</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Evaluation apprehension (Q17,Q44,Q45; alpha=0.89)</td>
<td>Mean</td>
<td>6.0</td>
<td>5.9</td>
<td>5.8</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.2</td>
<td>1.4</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Free riding (Q15)</td>
<td>Mean</td>
<td>4.5</td>
<td>4.6</td>
<td>5.1</td>
<td>1.63</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.5</td>
<td>1.7</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td><strong>Potential sources of process gain:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synergy (Q9,Q10; alpha=0.71)</td>
<td>Mean</td>
<td>5.8</td>
<td>5.6</td>
<td>5.1</td>
<td>3.51</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.0</td>
<td>0.7</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>More information (Q8)</td>
<td>Mean</td>
<td>5.6</td>
<td>5.5</td>
<td>5.9</td>
<td>1.52</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.2</td>
<td>0.9</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>Learning (Q24,Q25; alpha=0.74)</td>
<td>Mean</td>
<td>6.1</td>
<td>5.8</td>
<td>5.9</td>
<td>1.46</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.9</td>
<td>1.0</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) See Appendix N (Post-Identification Session questionnaire) for specific questionnaire items cited in the above table
2) Questionnaire scale (Q1 to Q40): Likert scale from 1 (strongly disagree) to 7 (strongly agree)
3) For multi-item scales the mean score is reported
4) Scale reliability is reported using Cronbach's alpha
5) The Likert scale for Q17 was reversed to make the 3-item scale for "Evaluation apprehension" consistent; The Likert scale for Q25 was reversed to make the 2-item scale for "Learning" consistent
6) One-way ANOVA with (2,115) degrees of freedom
matched expectations, as the EMS group members reported significantly less domination than the Baseline group members. However, contrary to expectations, there was no significant difference between the EMS and the Manual treatments.

**Side-tracking:** Side-tracking was measured using a single item measure. A significant difference across treatments was found (p=0.001). A post-hoc Tukey test (alpha=0.05) showed that the findings partially matched expectations, as the Baseline group members reported more side-tracking than those in the EMS and Manual treatment groups. Although it was expected that the Manual group members would also report significantly more side-tracking than those in the EMS groups (due to the lack of a parallel communication channel), the data indicates that there was no difference between these treatments.

**Failure to remember:** Failure to remember was measured using a two-item measure (Cronbach’s alpha=0.87). A significant difference across treatments was found (p<0.001). However, the direction of the difference was the opposite of what was expected. While it was expected that the Baseline group members would rate this scale higher than subjects in the other two treatments (due to lack of group memory support for the Baseline groups), the Tukey test (alpha=0.05) shows that the Baseline subjects actually perceived significantly less "failure to remember" than the other two groups. Additionally, the Manual group members reported significantly lower scores on this scale.
than the EMS subjects.

**Information overload:** Information overload was measured using a two-item scale (Cronbach’s alpha=0.74). A marginally significant difference was found across treatments (p=0.06). The findings matched expectations, as a post-hoc Tukey test (alpha=0.05) found that the EMS group members reported significantly more information overload than members in the other two treatments. There was no significant difference between the Baseline and Manual treatments.

**Production blocking:** Production blocking was measured using a three-item scale (Cronbach’s alpha=0.73). Although it was expected that the EMS group members would report significantly less production blocking than the other two treatments (due to the parallel communication channel), no significant difference was found.

**Evaluation apprehension:** Evaluation apprehension was measured using a three-item scale (Cronbach’s alpha=0.89). The subjects reported a low degree of evaluation apprehension and no significant difference was found across the treatments. As the experiment was designed to eliminate anonymity for all treatments, it was expected that there would be no significant difference across treatments with respect to this measure.
Free riding: Free riding was measured using a single item question which asked subjects to report on the degree to which everyone on the team contributed the same amount of effort. No significant difference was found across the treatments, as subjects in all treatments felt that there was a bit of free riding in their group. No significant difference was found across the treatments. This finding was anticipated since the experiment was designed to control for free riding by eliminating anonymity across treatments.

Synergy: Synergy was measured using a two-item scale (Cronbach's alpha=0.71). The mean scores indicate that synergy was perceived to be present by subjects in all treatments. A significant difference across treatments was found (p=0.03). Although a significant difference was expected, the direction of the difference was contrary to what was expected. A post-hoc Tukey test found that the Baseline group members reported significantly more synergy than the EMS subjects.

More information: "More information" was measured using a single-item question that asked each subject whether other members of their team had found defects that the subject would not have found on his/her own. The mean values for this measure indicate that subjects in all treatments felt that their team members made new findings available to the team. Although it was expected that the EMS subjects would report a higher amount of "more information" than the other treatments, no significant differences across treatments were found.
Learning: Learning was measured using a two-item scale (Cronbach’s alpha = 0.74). The mean values for this scale indicate that subjects in all treatments felt that learning was a by-product of their group session. It was expected that the EMS groups would report a higher amount of learning than the other treatment groups. However, no significant difference across teams was found.

Perceptions of the group process and the EMS approach

Several measures were included on the post-identification questionnaire to assess the subjects’ perceptions regarding their group, findings, effectiveness, and the EMS approach. The means, standard deviations, and ANOVA results are presented in Table 6.5.

Satisfaction: A two-item scale was used to measure satisfaction with the group (Cronbach’s alpha = 0.77). Subjects in all treatments were satisfied with their group. A marginally significant difference across treatments was found (p = 0.07). The Baseline subjects were the most satisfied, while the Manual and the EMS subjects were somewhat less satisfied.

Acceptance: A two-item scale was used to measure acceptance of the team findings (Cronbach’s alpha = 0.72). Subjects in all treatments expressed strong acceptance of their
team's findings. A significant difference across treatments was found (p=0.02). A post-hoc Tukey test (alpha=0.05) found that the Baseline group members were significantly more accepting of their findings than those in the EMS groups.

**Effectiveness:** Team effectiveness was assessed using a single item measure. Subjects in all treatments generally viewed their team as being very effective and no significant difference across treatments was found.

<table>
<thead>
<tr>
<th>Questionnaire Item/Scale</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction (Q1,Q2; alpha=0.77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.9</td>
<td>5.7</td>
<td>5.4</td>
<td>2.74</td>
<td>0.07</td>
</tr>
<tr>
<td>Std</td>
<td>0.9</td>
<td>0.9</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acceptance (Q3,Q4; alpha=0.72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>6.1</td>
<td>5.8</td>
<td>5.5</td>
<td>3.95</td>
<td>0.02</td>
</tr>
<tr>
<td>Std</td>
<td>0.7</td>
<td>0.9</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effectiveness (Q7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>5.7</td>
<td>5.3</td>
<td>5.3</td>
<td>1.39</td>
<td>0.25</td>
</tr>
<tr>
<td>Std</td>
<td>0.9</td>
<td>1.1</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction with EMS (Q57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>NA</td>
<td>NA</td>
<td>6.0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Std</td>
<td>NA</td>
<td>NA</td>
<td>1.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effectiveness of EMS approach (Q58,Q59; alpha=0.71)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>NA</td>
<td>NA</td>
<td>5.4</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Std</td>
<td>NA</td>
<td>NA</td>
<td>1.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) See Appendix N (Post-Identification Session questionnaire) for specific questionnaire items cited in the above table
2) Questionnaire scale (Q1 to Q40): Likert scale from 1 (strongly disagree) to 7 (strongly agree)
3) For multi-item scales the mean score is reported
4) Scale reliability is reported using Cronbach's alpha
5) NA = not applicable
6) One-way ANOVA with (2,117) degrees of freedom

Table 6.5: Perceptions of the group process for the identification session
Satisfaction with EMS: Satisfaction with the EMS was evaluated using a single measure. On a seven-point scale ranging from 1 (strongly disagree) to 7 (strongly agree), subjects averaged a rating of 6.0 (std=1.4) for "I was satisfied with the computer system used for this exercise." This finding suggests that the EMS subjects were very satisfied with the EMS system.

Effectiveness of the EMS Approach: Subjects in the EMS groups were asked to evaluate the effectiveness of the computer-based approach using a two-item scale (Cronbach’s alpha=0.71). The mean score for this scale (averaged across the two items) was 5.4 (std=1.4) on a scale of 1 (strongly disagree) to 7 (strongly agree), which indicates that the subjects considered the EMS approach to be effective for the defect identification task.

6.2.2 Comparison of performance: Individual versus interactive defect identification

Since the same task, subject pool, and basic procedures (e.g., experimental overview, time allowed for defect identification, group size) were used for both Experiment 1 and Experiment 2, it is possible to compare the defect identification findings across experiments. In Experiment 1, subjects in each group worked independently to identify defects (and then the findings were pooled), while in Experiment 2, the group members worked interactively to identify defects. For the fifteen groups in Experiment 1, the
mean number of distinct defects identified was 15.9 defects (std=2.2). By referring to Table 6.3 we see that the mean performance of the groups in Experiment 1 is better than that of the Baseline (13.3 defects) and Manual (14.8 defects) treatment groups of Experiment 2, but less than the EMS groups (18.0 defects). A one-way ANOVA indicated that there was a significant difference across the four sets of groups (F(3,41)=5.28, p<0.01). However, a post-hoc Tukey test (alpha=0.05) found that the only significant difference was between the EMS and Baseline treatment groups of Experiment 2. Even though there appear to be sizeable performance differences between the Experiment 1 groups and the Baseline groups (2.6 defect difference) and the Experiment 1 groups and the EMS groups (2.1 defect difference), the differences were not large enough to be statistically significant.

6.2.3 Defect evaluation session: Performance and perceptions

6.2.3.1 Confound checks

As was done for the defect identification session, efforts were made during the design and conduct of the experiment to minimize effects due to the following potential confounds on group performance: the influence of the scribe/moderator, anonymity, degree of personal effort, adequacy of review time, and task focus during the session. The same survey items that were discussed in the previous section were also included on
the post-evaluation questionnaire to help assess the subjects’ perceptions regarding these factors. Table 6.6 presents the means, standard deviations, and ANOVA results of subjects’ perceptions regarding the questionnaire items. As discussed below, the questionnaire results suggest that the potential confounds identified above were not present in the experiment.

**Influence of scribe/moderator:** Subjects’ perceptions regarding the influence of the scribe/moderator were measured using a two-item scale (Cronbach’s alpha=0.77). For all treatment groups the scribe/moderator was viewed as having a minimal influence on the team inspection and no significant difference was found with regard to this measure.

**Anonymity:** The mean scores for this item indicate that anonymity was generally not present in any of the treatment groups. This finding was expected since the evaluation session involved face-to-face interaction for each of the treatment groups. An ANOVA test found that there was a significant difference across treatments with regard to this measure (p=0.002). A post-hoc Tukey test (alpha=0.05) indicates that subjects in the EMS treatment were more likely to know the identity of the person who introduced a finding to the group than subjects in the other two treatments. This result may be due to the fact that one could determine who had contributed a finding during the identification session by simply reading the name that was attached to the contribution. For the other treatments, the subjects needed to remember who had introduced the finding.
### Table 6.6: Perceptions of the evaluation session: Confound checks

<table>
<thead>
<tr>
<th>Questionnaire Item/Scale</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scribe/moderator influence (Q32, Q33; alpha = 0.77)</td>
<td>Mean</td>
<td>1.8</td>
<td>2.1</td>
<td>2.0</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.1</td>
<td>1.1</td>
<td>1.2</td>
<td>0.53</td>
</tr>
<tr>
<td>Knew identity of contributor (Q26)</td>
<td>Mean</td>
<td>5.9</td>
<td>6.1</td>
<td>6.8</td>
<td>6.43</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.1</td>
<td>1.4</td>
<td>0.7</td>
<td>0.002</td>
</tr>
<tr>
<td>I put forth my best effort (Q11)</td>
<td>Mean</td>
<td>6.0</td>
<td>6.1</td>
<td>6.2</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.9</td>
<td>0.6</td>
<td>0.9</td>
<td>0.61</td>
</tr>
<tr>
<td>Had adequate time for identification review (Q27)</td>
<td>Mean</td>
<td>2.3</td>
<td>1.9</td>
<td>2.4</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.6</td>
<td>1.1</td>
<td>1.6</td>
<td>0.42</td>
</tr>
<tr>
<td>Team's effort was spent on defect detection (Q28)</td>
<td>Mean</td>
<td>4.4</td>
<td>3.9</td>
<td>5.4</td>
<td>6.02</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>2.0</td>
<td>2.2</td>
<td>1.8</td>
<td>0.003</td>
</tr>
<tr>
<td>Team's effort was spent on defect correction (Q29)</td>
<td>Mean</td>
<td>2.9</td>
<td>2.2</td>
<td>3.2</td>
<td>4.16</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.6</td>
<td>1.4</td>
<td>1.8</td>
<td>0.02</td>
</tr>
<tr>
<td>Team's effort was spent on defect evaluation (Q30)</td>
<td>Mean</td>
<td>4.8</td>
<td>5.4</td>
<td>5.2</td>
<td>1.65</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.5</td>
<td>1.6</td>
<td>1.5</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Notes:  
1) See Appendix O (Post-Evaluation Session questionnaire) for specific questionnaire items cited in the above table  
2) Questionnaire scale (Q1 to Q35): Likert scale from 1 (strongly disagree) to 7 (strongly agree)  
3) For multi-item scales the mean score is reported  
4) Scale reliability is reported using Cronbach's alpha  
5) One-way ANOVA with (2, 115) degrees of freedom

**Degree of personal effort:** Subjects from all treatment groups indicated that they put forth a strong effort during the defect identification session. No significant difference across treatments was found.

**Adequacy of review time:** The mean ratings regarding the adequacy of review time suggest that the time allotted to the task was appropriate, as groups did not feel an undue
amount of time pressure during the evaluation review. There was no significant
difference with regard to adequacy of review time which suggests that the slight
variations in procedure across treatments did not significantly impact the amount of time
that the subjects could devote to the detection of defects.

Task focus: Questionnaire items were included to check whether the teams in each
treatment group were focussed on the proper group task (i.e., defect evaluation).
Subjects were asked to indicate whether their team spent a significant amount of effort
on each of the following: defect detection, correction, and evaluation. Based on the
survey responses, it appears that subjects in all treatments focussed on evaluation and
avoided expending effort on correction. However, defect detection also was a relatively
strong focus for the Baseline and EMS groups. There were no significant differences
across treatments with respect to perceived effort devoted to defect evaluation. There
was a significant difference across treatments with respect to defect detection (p=0.003)
and correction (p=0.02). Post-hoc Tukey tests (alpha=0.05) indicated that subjects in
the Baseline groups felt that they had expended significantly more effort on detection than
was reported by subjects in the other treatments, while the Manual group members
perceived that they had expended less effort on defect correction than was reported by
subjects in the other groups.
6.2.3.2 Group performance and perceptions

Group performance

Table 6.3 presents the mean number of defects reported by the groups following the evaluation session, along with standard deviations and the ANOVA results. The relative performances across treatments tracks the post-identification session performance as the EMS groups reported the most defects (17.7 defects), followed by the Manual groups (15.3 defects), and the Baseline groups (13.0 defects). As predicted, a significant treatment effect was found with respect to the number of defects reported by the groups following the defect evaluation session (p=0.008). A post-hoc Tukey test (alpha=0.05) showed that the EMS groups identified significantly more defects than the Baseline groups. The Tukey tests indicated that there were no significant differences between the Baseline and Manual groups or the Manual and EMS groups.

Process loss and gain

The impact of the task support mechanisms on the performance of the inspection teams can be assessed by examining the relative amount of process loss and gain across the treatments. Process loss and gain were measured for each group by comparing the list of defects that had been found following the identification session with the list of defects
that were reported on each team's "Action List" (i.e., the list of defects that each team turned in as its final product for the exercise). To aid in the interpretation of the findings, the mean number of post-identification session defects, process loss, and process gain are reported as a fraction of post-evaluation session defects (see Table 6.7).

The ratio of post-identification session defects to post-evaluation session defects may be used to determine whether there was a net process loss or gain during the evaluation session: a ratio greater than one indicates a net loss, while a ratio less than one indicates a net gain. As indicated in Table 6.7, the Baseline (ratio=1.03) and EMS (ratio=1.02) treatment groups experienced a small net loss during the evaluation session, while the Manual groups (ratio=0.97) experienced a small net gain. An ANOVA test indicated that there was a significant difference across treatments with respect to this ratio (p=0.02). A post-hoc Tukey test (alpha=0.05) indicated that the Manual treatment groups had a significantly lower ratio than the other two treatment groups.

The significant difference for the ratio across treatments can be traced to differences in process loss across the treatments. A significant difference existed across treatments with respect to the process loss ratio (p=0.001) (see Table 6.7). The Manual groups experienced no process loss, while loss did occur for the EMS (ratio=0.03) and Baseline groups (ratio=0.08). As expected, a post-hoc Tukey test indicated that the Manual and EMS groups had significantly less loss than the Baseline groups. No significant
difference across treatments was found with respect to the process gain ratio (see Table 6.7).

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post-Identification Defects</td>
<td>Mean</td>
<td>1.03</td>
<td>0.97</td>
<td>1.02</td>
<td>4.44</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>.06</td>
<td>.04</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>Post-Evaluation Defects</td>
<td>Mean</td>
<td>0.08</td>
<td>0</td>
<td>0.03</td>
<td>16.12</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>.04</td>
<td>0</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>Process Loss</td>
<td>Mean</td>
<td>0.04</td>
<td>0.03</td>
<td>0.01</td>
<td>2.40</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>.04</td>
<td>.04</td>
<td>.02</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) Ten groups per treatment cell  
2) One-way ANOVA with (2,27) degrees of freedom  
3) The values in the "Baseline" column do not add up due to round-off error

Table 6.7: Ratios of number of post-identification session defects, loss, and gain to number of post-evaluation session defects

Perceptions of potential sources of process loss and gain

To learn more about the sources of process loss and gain during the evaluation session, the post-evaluation session questionnaire asked subjects to respond to several survey items relating to the sources of process loss and gain (the questionnaire items were the same as those used for the post-identification questionnaire). Table 6.8 presents the means, standard deviations, and ANOVA results of subjects’ perceptions regarding the pertinent questionnaire items. The findings for specific measures are discussed below.
Process loss: The mean values for the survey items related to process loss suggest that there were not many group problems associated with the measured process losses. Subjects (across all treatments) generally reported a limited degree of domination, minimal sidetracking, a low level of "failure to remember," little information overload, low production blocking, and very low evaluation apprehension. Free riding is the only measured source of process loss that was perceived to be present to any degree. Information overload is the only process loss measure which significantly varied across treatments (p=0.03). A post-hoc Tukey test (alpha=0.05) found that the Manual subjects reported significantly less information overload than subjects in the other two treatments.

Process gain: The subjects reported the presence of process gain for each of the gain measures included on the post-evaluation session questionnaire: synergy and learning. Significant differences across treatments were found for synergy (p=0.04) and for learning (p=0.009). For each measure, post-hoc Tukey tests (alpha=0.05) indicated that the reported process gains were significantly lower for the subjects in the Manual treatment groups.
<table>
<thead>
<tr>
<th>Questionnaire Item/Scale</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential sources of process loss:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domination (Q16, Q17; alpha=0.82)</td>
<td>Mean</td>
<td>3.4</td>
<td>3.3</td>
<td>3.3</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.6</td>
<td>1.6</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>Side-tracking (Q25)</td>
<td>Mean</td>
<td>2.5</td>
<td>2.4</td>
<td>2.4</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.4</td>
<td>1.2</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Failure to remember (Q34, Q35; alpha=0.71)</td>
<td>Mean</td>
<td>2.1</td>
<td>1.9</td>
<td>2.1</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.0</td>
<td>0.9</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Information overload (Q22, Q23; alpha=0.71)</td>
<td>Mean</td>
<td>2.5</td>
<td>1.9</td>
<td>2.7</td>
<td>3.66</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.2</td>
<td>0.9</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td>Production blocking (Q36, Q37, Q38; alpha=0.84)</td>
<td>Mean</td>
<td>2.4</td>
<td>2.2</td>
<td>2.0</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.2</td>
<td>1.0</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Evaluation apprehension (Q39, Q40; alpha=0.79)</td>
<td>Mean</td>
<td>6.2</td>
<td>6.0</td>
<td>5.7</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.0</td>
<td>0.9</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>Free riding (Q10)</td>
<td>Mean</td>
<td>4.9</td>
<td>4.9</td>
<td>5.4</td>
<td>1.13</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.5</td>
<td>1.7</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Potential sources of process gain:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synergy (Q9)</td>
<td>Mean</td>
<td>4.9</td>
<td>4.3</td>
<td>5.2</td>
<td>3.30</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>1.5</td>
<td>1.7</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>Learning (Q19, Q20; alpha=0.72)</td>
<td>Mean</td>
<td>6.4</td>
<td>5.8</td>
<td>6.3</td>
<td>4.86</td>
</tr>
<tr>
<td></td>
<td>Std</td>
<td>0.7</td>
<td>1.1</td>
<td>0.6</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) See Appendix O (Post-Evaluation Session questionnaire) for specific questionnaire items cited in the above table
2) Questionnaire scale (Q1 to Q35): Likert scale from 1 (strongly disagree) to 7 (strongly agree)
3) For multi-item scales the mean score is reported
4) Scale reliability is reported using Cronbach's alpha
5) Although "Synergy" was measured using a two-item scale on the Post-Identification Session questionnaire, it was measured using only one item on the Post-Evaluation Session questionnaire since one of the items was not appropriate for the evaluation session.
6) The Likert scale for Q20 was reversed to make the 2-item scale for "Learning" consistent
7) One-way ANOVA with (2,115) degrees of freedom

Table 6.8: Perceptions of potential sources of group process loss and gain for the evaluation session
Perceptions of the group process

Several measures were included on the post-evaluation questionnaire to assess the subjects’ perceptions regarding their group process. The means, standard deviations, and ANOVA results are presented in Table 6.9.

**Satisfaction:** A two-item scale was used to measure group satisfaction (Cronbach’s alpha=0.83). Subjects in all treatments reported high satisfaction with their group. No significant difference across treatments was found.

<table>
<thead>
<tr>
<th>Questionnaire Item/Scale</th>
<th>Baseline</th>
<th>Manual</th>
<th>EMS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction</td>
<td>Mean</td>
<td>6.2</td>
<td>6.0</td>
<td>6.0</td>
<td>0.90</td>
</tr>
<tr>
<td>(Q1,Q2; alpha=0.83)</td>
<td>Std</td>
<td>0.7</td>
<td>0.7</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Acceptance</td>
<td>Mean</td>
<td>6.2</td>
<td>6.1</td>
<td>6.3</td>
<td>0.91</td>
</tr>
<tr>
<td>(Q3,Q4; alpha=0.83)</td>
<td>Std</td>
<td>0.6</td>
<td>0.6</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>Effectiveness</td>
<td>Mean</td>
<td>6.1</td>
<td>5.9</td>
<td>6.2</td>
<td>1.07</td>
</tr>
<tr>
<td>(Q7)</td>
<td>Std</td>
<td>0.7</td>
<td>0.8</td>
<td>0.8</td>
<td></td>
</tr>
</tbody>
</table>

Notes: 1) See Appendix O (Post-Evaluation Session questionnaire) for specific questionnaire items cited in the above table
2) Questionnaire scale (Q1 to Q35): Likert scale from 1 (strongly disagree) to 7 (strongly agree)  
3) For multi-item scales the mean score is reported
4) Scale reliability is reported using Cronbach’s alpha
5) One-way ANOVA with (2,117) degrees of freedom

Table 6.9: Perceptions of the group process for the evaluation session
Acceptance: A two-item scale was used to measure acceptance of the team findings (Cronbach’s alpha = 0.82). Subjects in all treatments expressed strong acceptance of their team’s findings. No significant difference across treatments was found.

Effectiveness: Team effectiveness was assessed using a single item measure. Subjects in all treatments generally viewed their team as being very effective. No significant difference across treatments was found.

6.3 Summary

In this chapter and the preceding chapter, the method and results of the experiments were presented. The next step is to discuss the findings in the context of the expectations of the study and to examine the implications of the findings. This discussion is presented in the next chapter.
The previous two chapters presented the results of the two experiments. The first part of this chapter discusses the findings in an effort to explain why the results came out the way that they did. Second, the implications of the study for practice and research are reviewed. The chapter ends with a discussion of the study's limitations and some concluding remarks.

7.1 Discussion

7.1.1 Experiment 1

This section summarizes and discusses the findings of Experiment 1. A summary of the findings from Experiment 1 are presented in Tables 7.1 and 7.2.

7.1.1.1 Nature of the task

Based on the results, the overall review task may be characterized as a complementary task comprised of disjunctive subtasks. The complementary nature of the task was demonstrated by the finding that different members generally contributed different resources (i.e., defects) to their team. On average, a large proportion of each team's
Table 7.1: Experiment 1: Summary of findings for each research question

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Question</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What are the characteristics of the review task?</td>
<td>o Task may be described as a complementary task comprised of Eureka subtasks</td>
</tr>
<tr>
<td>2</td>
<td>What is the variation in defect detection performance?</td>
<td>o Group variation (best:worst): 1.54:1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>o Individual variation (best:worst): 3.5:1</td>
</tr>
</tbody>
</table>

group resources (40 percent) were unique defects. The unique defects were generally spread about each team, as one capable person was usually not responsible for all the unique defects. The finding regarding the complementary nature of the review task supports the belief expressed in the literature (e.g., Yourdon, 1989) that a group of people (rather than a single individual) should be employed for the review task.

It is interesting to note that unique defects typically comprised a large proportion of group resources, despite the fact that all subjects had similar training and experience. Reasons for the large proportion of unique defects are not clear. Possible explanations may be that different people are adept at finding different types of errors (as suggested by Parnas and Weiss, 1987), or that the subjects were relatively inexperienced in defect identification (and hence there were frequently situations in which only one group member had the knowledge to detect a defect).
The task of detecting a specific defect in the work product was found to be a Eureka-type of disjunctive task, based on the high acceptance rates of properly diagnosed defects across treatments (98 to 100 percent for overlapping defects, 85 to 94 percent for unique defects). Groups did not generally discuss the evaluation of each defect extensively, as observations of the groups indicated that group members usually accepted a properly diagnosed defect without much debate. The Eureka nature of the task suggests that the defect evaluation subtask primarily requires a support mechanism (such as task support) which can ensure that all identification subtask findings are made available for evaluation. Since groups performing defect evaluation rarely needed to negotiate their positions or resolve conflicts, it appears that group support aimed at addressing these issues is not critical.

7.1.1.2 Variation in individual and group resources

A wide variation in individual defect detection ability was found, as the defect detection ratio between the best to worst individual performers (across the 57 subjects) was 3.5:1. Although random assignment of subjects to teams may have helped to minimize differences in group resources across experimental review teams, the ratio of group resources between the best and worst groups (across 15 teams) was still sizeable: 1.54:1. The large variations in ability across individuals and groups found in this study is consistent with previous findings reported in the software engineering literature for
programming and inspection tasks (e.g., Curtis, 1980; Schneider, et al., 1992).

7.1.1.3 Correlation of group resources and process to performance

Group resources (i.e., the aggregated defect detection ability of a group) appears to be a very important determinant of group review productivity. As predicted by Steiner's theory, the regression analysis performed for the data in Experiment 1 found that group resources correlated significantly ($p < 0.001$) with group performance. The high value for $R^2$ (0.92) indicates that group resources explained a great deal of the performance variance across groups. Contrary to expectations, no significant correlation was found with regard to the type of group process used for defect evaluation (see below for a discussion of the impacts of task support). These findings are not surprising given the strong Eureka nature of the task and the conditions of Experiment 1. As noted earlier, if a properly diagnosed defect was introduced to the group then the chances for acceptance were high. Assuming that most defects (i.e., resources) were introduced to the group, then one would expect a strong correlation between group resources and performance.

7.1.1.4 Impacts of task support

As expected, the study found that task support can significantly reduce process loss for
groups performing the evaluation subtask. As indicated in Table 5.5, the fractional process loss for the unsupported Baseline groups was 15 percent of the actual group findings, which is not a trivial amount of loss. The fractional loss for the supported groups was much less (four to five percent). As detailed in Chapter 5, most of the loss was due to the failure of groups to utilize all of their unique findings. There was very little process loss associated with overlapping findings.

<table>
<thead>
<tr>
<th>Review Subtask/ Support Mechanism</th>
<th>Hypothesis</th>
<th>Dependent Variable / Test</th>
<th>Significant findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defect evaluation: Task support</td>
<td>1</td>
<td>Regression</td>
<td>Performance correlates with group resources, but not treatment</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Process loss (measured)</td>
<td>Supported &lt; Unsupported</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Process gain (measured)</td>
<td>Not significant</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Satisfaction, Acceptance</td>
<td>Not significant</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Perceptions of understanding, confidence</td>
<td>Post-group review &gt; Post-individual review</td>
</tr>
</tbody>
</table>

Table 7.2: Experiment 1: Summary of findings for each hypothesis

While task support helped to reduce process loss, it did little to promote process gain. This null finding was not expected. The average amount of fractional process gain for the supported groups was almost negligible, which indicates that the hypothesized gains that may arise from task support (such as synergy and "more information") were not
present. While the fractional process gain was not significantly larger for the unsupported groups, it could be characterized as somewhat more than negligible (six percent). It appears that the presence of the public displays (i.e., the task support) caused the supported review teams to focus on evaluation of previously found defects and did not encourage the reviewers to identify additional defects.

7.1.1.5 Perceptions of satisfaction and acceptance

As expected, there were no significant differences with respect to perceptions of satisfaction with the group or the acceptance of the meeting outcome. These findings indicate that the application of task support had neither a positive or negative impact on the subjects' perceptions regarding these issues. Given this, one would not expect the implementation of these type of task support mechanisms to be disruptive to groups with regard to these concerns.

7.1.1.6 Perceptions of understanding and confidence

The claims made in the literature regarding the beneficial impacts of the review process on learning (e.g., Yourdon, 1989) were supported in this study as group members reported a higher level of understanding and confidence in their findings following the group review (compared to their perceptions following the individual review). These
findings underscore the intangible benefits that may take place when software reviews are performed.

7.1.2 Experiment 2

This section summarizes and discusses the findings of Experiment 2. A summary of the findings from Experiment 2 are presented in Table 7.3.

7.1.2.1 Impacts of process support

The results of the study suggest that EMS process support (in the form of parallel communication and group memory) can help to improve defect detection performance for interactive teams performing the defect identification subtask. As hypothesized, the results indicate that EMS-supported groups performed significantly better than unsupported groups. While the EMS groups also outperformed the Manual groups by a large amount (22 percent), the difference between these groups was not significant. Additionally, there was no significant difference between the Manual and Baseline groups.
<table>
<thead>
<tr>
<th>Review Subtask/Support Mechanism</th>
<th>Hypothesis</th>
<th>Dependent Variable / Test</th>
<th>Significant findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defect identification: Process support</td>
<td>5</td>
<td>No. of defects identified</td>
<td>EMS &gt; Baseline</td>
</tr>
</tbody>
</table>
|                                   | 6          | Perceptions of loss      | o Domination: EMS < Baseline  
|                                   |           |                          | o Side-tracking: EMS = Manual < Baseline  
|                                   |           |                          | o Failure to remember: Baseline = Manual < EMS  
|                                   |           |                          | o Information overload: Manual < EMS       |
|                                   | 7          | Perceptions of gain      | o Synergy: Baseline > EMS |
|                                   | 8          | Satisfaction, Acceptance | o Satisfaction: Baseline > EMS  
|                                   |           |                          | o Acceptance: Baseline > EMS               |
|                                   | 9          | Interactive vs. Nominal  | Not significant       |
| Defect evaluation: Task support   | 10         | Process loss (measured)  | Manual < Baseline     |
|                                   |            | Process gain (measured)  | Not significant       |
|                                   | 11         | Perceptions of loss      | o Information overload: Manual < EMS   |
|                                   | 12         | Perceptions of gain      | o Synergy: EMS > Manual  
|                                   |           |                          | o Learning: Baseline = EMS > Manual         |
|                                   | 13         | Satisfaction, Acceptance | Not significant       |

Table 7.3: Experiment 2: Summary of findings for each hypothesis
The difference between the EMS and Baseline groups was more than an artifact of statistics, as the size of the treatment effect was considerable: on average, EMS groups identified 35 percent more defects than the Baseline groups. These findings are in agreement with other EMS studies which have found that process support in the form of a parallel communication channel can improve group performance for divergent tasks (e.g., see Gallupe, et al., 1992). This finding also offers support for Fagan's (1986) assertion that "group process" can be a key influence on inspection team performance.

In an effort to explain the performance findings, this study used a number of perceptual measures to assess the impact of the support mechanisms on the potential sources of process loss and gain. Overall, the perceptual findings regarding the sources of process loss and gain for the identification subtask suggest that the performance differences between the EMS and the unsupported groups can be primarily attributed to reductions in process loss, rather than increases in process gain.

**Process loss**

A summary of the findings regarding perceptions of process loss is presented in Table 7.3. Group members using the EMS reported significantly more favorable perceptions with respect to domination and side-tracking. The difference in responses between the EMS and Baseline group members with respect to domination is likely due to parallel
communication, since the parallel (and nonverbal) electronic channel is a poor forum for domination. The fact that there was not a significant difference between the Manual and Baseline group members with respect to domination indicates that the group memory feature does not account for the difference. Side-tracking was significantly lower for the EMS and Manual groups versus the Baseline groups. Since the Manual groups only had group memory process support, this finding suggests that group memory plays a role in reducing loss due to side-tracking. Group memory may help to reduce this type of loss by helping group members to stay more task focussed (i.e., members may be focussed on "filling up" the whiteboard or the electronic file with new findings) and by cutting down on wasted time due to repeating the introduction of defects that had already been identified (Baseline groups were observed to occasionally introduce a defect more than once).

While the perceptual data indicates that the EMS process support helped to reduce some types of process loss, the survey responses also indicate that the EMS may have increased other types of process loss: information overload and "failure to remember." The perceived increase in process loss due to information overload was expected and can be attributed to the greater amount of meeting input that is generated through the use of the parallel communication channel. The ratings for "failure to remember" were significantly higher for the EMS members than they were for the other two groups. This indicates that the EMS members were less likely to keep track of the contributions made
by others in their group. While this finding was the opposite of what was expected, observations made during the experiment indicated that this unexpected finding may be due to the window-based design of the EMS tool used for the study (i.e., Group Outliner). Although it was not difficult to view all the contributions made by others via Group Outliner, the interface did not seem to encourage users to view others' contributions, as the EMS group members did not frequently scan contributions made in windows other than the one that they were working in.

Process gain

As indicated in Table 7.3, the only significant difference across groups with regard to perceptions of process gain involved the measure of synergy. Contrary to predictions, the Baseline group members reported a significantly higher level of synergy than those in the EMS groups. Although this finding was unexpected it may be consistent with the earlier finding regarding "failure to remember." As noted above, the interface of the EMS used for the study may have discouraged EMS subjects from keeping track of the findings made by others, and hence may have reduced the opportunities for synergy.

Contrary to expectations, no significant differences were found with respect to learning or "more information." The lack of significance is somewhat surprising in light of the large performance difference between the EMS and Baseline groups. One reason for the
lack of significance is that the EMS members failed to keep track of the input made by others and thus failed to fully recognize all of the inputs made by their team.

As noted above, attempts were made to identify the sources for process loss and gain for groups performing the identification subtask by using the perceptual measures. Although the perceptual measures may not accurately reflect actual process loss and gain, the findings suggest that the EMS used in this study had both favorable and unfavorable impacts on the groups. Specifically, while the perceptual data indicates that the EMS was beneficial since it reduced some types of process loss (i.e., domination, sidetracking), the EMS may have been somewhat detrimental in that subjects perceived higher degrees of other types of loss (failure to remember, information overload) and less gain due to synergy. While the overall impact of the EMS on defect identification was positive, the perceptual data regarding sources of loss and gain suggest that the application of EMS to this task may result in tradeoffs with respect to different sources of process loss and gain.

Satisfaction, Acceptance, and Effectiveness

Ratings for satisfaction and acceptance were favorable across treatments. However, the Baseline group members reported a significantly higher level of satisfaction and acceptance than the EMS subjects. Reasons for this are not clear. It may be that the
Baseline subjects were more comfortable with a "traditional" meeting and thus rated their treatment more highly. Also, the fact that EMS subjects did not communicate orally may have been a source of discomfort or frustration for the EMS subjects, as the subjects may have desired the richer communication channel of speech at times. However, the fact that the scores for satisfaction and acceptance were favorable for EMS groups suggests that the EMS approach would not be problematic to implement with respect to these issues.

As expected, despite the large difference in actual performance between the Baseline and EMS groups, there was no difference with respect to perceived group effectiveness. This finding is consistent with the finding reported by others that group members can be poor judges of group effectiveness (e.g., Bouchard, 1969; Connolly, et al., 1990). While it may be difficult to generalize this finding to the field, it raises the possibility that some members of software review teams in industry may also be overestimating the effectiveness of their group review process. If this is true, then industry review teams using a sub-optimal review process might not be aware that their review process is experiencing process loss that is reducing their review effectiveness.

7.1.2.2 Impacts of task support

In general, the magnitude of the process loss was not large for teams performing the
defect evaluation subtask (range: zero to eight percent of the final number of group defects detected). As noted earlier, the relatively low level of process loss can probably be attributed to the Eureka nature of the task. As expected, the results of this study suggest that task support (in the form of public display) can help to reduce process loss for teams performing the defect evaluation subtask (see Table 7.3). In particular, the results indicate that the Manual groups had significantly less process loss than the unsupported groups. However (contrary to expectations), while the EMS groups (which had task support) had less loss than the Baseline groups, there was no significant difference between these treatments.

Unlike the Manual groups (which experienced no loss), there was some process loss for the EMS teams. One explanation for the extra process loss experienced by EMS teams (versus the Manual teams) may be information overload: the EMS subjects perceived significantly more information overload than the Manual groups. The list of defects evaluated by the EMS teams typically included many more entries than the lists evaluated by the Manual teams, since the EMS lists included many repeated defects. The presence of many repeat entries on the typical EMS list likely contributed to the perceived overload and may have caused the EMS subjects to be a bit more careless during the evaluation subtask. Another reason underlying the loss exhibited by EMS groups may be that subjects may have had occasional difficulty in expressing the defects they found using written language. It is possible that a poorly articulated defect (even if properly
diagnosed) might not have received support during the evaluation session. This type of problem probably didn’t happen for the Manual groups since defects raised during the Manual identification sessions were usually clarified orally prior to being put on the public display list (i.e., white board).

No significant differences were found across the groups with respect to measured process gain, as gains were small for all groups (range: 1 to 4 percent of the final number of group defects detected). While no gains across groups were measured, the perceptual data indicated that the subjects perceived differences with respect to synergy and learning (see Table 7.3). Specifically, the Manual group members rated process gain for these two factors lower than the members of the other groups. The low process gain for the Manual group may be attributed to the rather brisk nature of the Manual evaluation meetings. Observations showed that members of the Manual groups moved very quickly through the list of potential defects developed during the identification session. As the Manual groups had already orally discussed the defects in a focused way, they appeared to gain less from the evaluation session than members of the other groups.

As expected, no differences across treatments was found with respect to satisfaction or acceptance. Members from all treatments appeared to have favorable feeling regarding their group process. The fact that perceptions were statistically the same across treatments has practical significance, since it suggests that the application of task support
may be readily accepted by software review teams (i.e., the data do not suggest that review team members would resist the use of a task support mechanism).

7.1.2.3 Examination of the "novelty effect"

In Experiment 2, strong attempts were made to control and test for confounding factors (e.g., motivation). Based on the discussion regarding "confound checks" in Chapter 6, it does not appear that any of the confounds examined earlier existed for this study. However, an effect which has not been considered is the "novelty effect." A novelty effect might have occurred if subjects unfamiliar with the EMS technology were motivated to expend more effort due to the novelty of the EMS environment. However, the chances for a novelty effect are slim, as 73 percent of the EMS subjects had worked in the GroupSystems EMS environment at least once before participating in this study. Additionally, the members of EMS groups did not appear to have been more motivated than those in other treatments, as there was no significant difference across treatments with regard to responses to a question on the post-identification session which asked whether the subjects had "put forth my best effort" (see Table 6.2). In light of this supporting data, it is doubtful that a novelty effect impacted the findings of the study.
7.2 Implications

The findings from this study dealt with the nature of the review task, the impact of support mechanisms on review team performance, and the importance of group resources to review team performance. These findings have a number of implications for practitioners and researchers.

7.2.1 Implications for practice

A major implication of this study for practitioners is that unsupported software review teams are probably not detecting as many defects as they should be finding due to group process loss. Based on this study, practitioners should strongly consider using group support mechanisms to improve review team performance. While the most dramatic (and significant) impacts were found for the application of process support to the identification subtask, the use of task support for the evaluation subtask was also found to have significant effects on performance.

While this study suggests that the EMS approach to reviews is the best of those examined, it appears that non-automated techniques may also be useful. Given the high cost of EMS, those in industry may wish to consider using the non-automated techniques described in this study. For example, this study found that a simple non-automated task
support mechanism consisting of a visual display is one technique that may be used to significantly reduce process loss associated with the failure of group members to introduce all unique findings to their team during defect evaluation.

While the task support mechanism helped reduce a portion of the loss associated with unique findings, it did not have an effect on the process loss associated with unique findings which were poorly articulated to the group. To help reduce this type of process loss, review groups may be advised to explicitly focus on making sure that all findings introduced by group members are understood by the group before being discarded as invalid contributions. The use of facilitation techniques (as practiced by the moderator) may help to ensure that all findings are fully articulated to the team.

With regard to the application of EMS in organizations, a related issue to consider is the implementation of EMS into the software review process. In particular, how easy would it be to implement the EMS approach for software reviews into business organizations? The findings from this study suggest that users may be willing to accept the EMS approach, as the EMS system used for this study received very favorable ratings with regard to user satisfaction and perceived system effectiveness (see Table 6.5).
7.2.2 Implications for research

This study demonstrated that net process loss can occur for review teams involved with both the defect identification and evaluation subtasks and that support mechanisms can be used to improve the balance of process loss and gain. Given the importance of the software review to the software development process, the findings from this study should "raise a flag" with researchers and draw attention to the need for additional group-oriented research aimed at supporting review teams. In particular, researchers may wish to explore further the impact of group support mechanisms on review team performance.

This study found that the group process associated with the defect identification subtask was subject to a relatively larger net process loss than that of the evaluation subtask. Since the identification subtask appears to be a more pressing concern, researchers may wish to focus more of their efforts on investigating mechanisms to support this subtask.

There are several issues that researchers may wish to explore. For example, this study looked at only two ways of delivering process support to review teams: 1) A manual approach using a whiteboard, and 2) An EMS approach using the Group Outliner tool from the GroupSystems EMS toolkit. Researchers may wish to investigate the impacts of other approaches to process support for review teams (e.g., other manual approaches, different GroupSystems tools, different EMS packages). For example, as suggested in the discussion of the results, the type of EMS interface used for a software review may
have a direct impact on the sources of process loss and gain. Researchers may wish to investigate different EMS interfaces to determine which interfaces can support the most desirable balance of process loss and gain.

The software review literature suggests that review groups should be relatively small, and should not involve more than five people (e.g., Fagan, 1986; Yourdon, 1989). However, the findings from some studies suggest that larger groups of reviewers may be useful since more input may result in more defects being detected (Ackerman, 1988; Martin and Tsai, 1990; Schneider, et al., 1992). Historically, researchers have found that large group sizes can be dysfunctional (McGrath, 1984; Steiner, 1972). However, recent research has shown that EMS technology may be used to improve the performance of large groups performing divergent tasks (Gallupe, et al., 1992). Since large group sizes for software reviews may be desirable in certain situations (e.g., user reviews, JAD reviews), it would be worthwhile to investigate the impact of EMS process support on group review performance as a function of group size. Research of this type may show that group size limits for review teams can be increased.

While net process loss was relatively greater for the identification subtask, researchers may still wish to examine support for the evaluation subtask. The findings from this study indicate that task support was successful in reducing process loss for the evaluation subtask to a very low level. However, the issue of encouraging process gain during the
evaluation subtask is still unresolved, as the task support mechanism used in the study did not support the generation of process gain. In fact, the task support appeared to have a restrictive effect on review groups as groups without task support generated more gain. Thus the advantages of using task support to reduce process loss were somewhat counterbalanced by the failure to encourage process gain. It may actually be dysfunctional for groups to focus on detecting new defects (i.e., process gain) during the evaluation session. However, it is possible that the defect evaluation process could trigger new thinking regarding the identification of undiscovered defects. Hence, it may be worthwhile to consider another "round" of defect identification and evaluation after the first evaluation session in order to capture as many gains as possible from the review process. Researchers may wish to examine the value of conducting a second round of review (i.e., identification and evaluation).

Historically, software reviews have been held as meetings in which team members meet together at the same place and time. However, given the apparent Eureka nature of the review task, it may be possible to conduct relatively productive asynchronous (i.e., "different time") team reviews for teams which have difficulty in finding a common time to meet. EMS technology could serve as the vehicle for implementing the asynchronous review: individual findings could be stored in a common EMS repository and then examined and evaluated by members of the review team when free time was available. An electronic voting tool could be used to poll review team members on the validity of
specific findings. In addition, team members could electronically append comments to the repository during their evaluation. As with any review, the input from the review team would then be used by the producer of the work product as the basis for product revisions.

While the asynchronous review may hold promise, the findings from this study do raise some potential pitfalls with the asynchronous approach. First, the study showed that the articulation of a valid finding is not always clear to others. This problem is likely to be exacerbated for asynchronous review teams since it will require more effort for a team to clarify the gist of a reviewer’s findings without the benefit of face-to-face communication. Also, there may be limited opportunities for process gains unless special techniques are devised to stimulate the identification of new findings in an asynchronous environment. Lastly, the potential for intangible benefits due to the review process (e.g., educational benefits) may be limited due to the relatively low richness of electronic media versus that of face-to-face communication.

The software review literature suggests that synergy is a common by-product of team reviews and is stimulated by the moderator (Fagan, 1976). However, since no controlled studies of reviews have attempted to measure process gain due to a facilitator, it is not clear whether such claims for process gain are fact or fiction. In this study, measured process gain was very small for groups performing the defect evaluation subtask, and was
probably relatively limited (if it even existed at all) for groups performing interactive
defect identification. Since the results of this study regarding process gain are not
consistent with the claims made by the literature, it may be worthwhile to follow up on
this issue and investigate how key roleplayers in the review process (e.g., moderator,
reader; see (Fagan, 1979)) may optimally influence the productivity of the review
process. Research in this area could provide perspective regarding the importance of
roleplayers to team review productivity.

Lastly, researchers should consider replicating this study in a field setting using
professional subjects. While the underlying reasons given for the study results should
be applicable to professional subjects as well as student subjects (i.e., process loss and
gain are likely to occur for all types of subjects), a field investigation is required to allow
us to generalize the findings from this study to the field. In addition, the study should
be replicated for other type of software work products to determine whether group
support is more or less important for other types of reviews.

7.3 Limitations

The generalizability of the study is limited to the extent that the subjects and
experimental task are representative of software reviews in the field. Several steps were
taken to make the experimental exercise realistic: 1) The experimental task involved the
review of a document characteristic of an analysis specification review; 2) The experimental task was reasonably complex (the task involved 45 minutes for defect identification and 45 minutes for defect evaluation); 3) The teams participating in the study had worked together in the past; and 4) The subjects in the study were relatively mature (mean age = 24 years) and most had work experience.

However, there remain several limitations to the study. First, the relative inexperience of subjects with regard to the review task and the group review process may have affected the findings. Subjects with more experience performing software reviews may behave differently and have different perceptions regarding the review process. In addition, although the teams had worked together for several weeks prior to the review, the behavior of the experimental teams may still not have been representative of teams that have spent more time together. Third, while the number of groups used for the study appeared adequate based on the number of groups used by researchers in other published experimental studies of this type (e.g., see Connolly, et al., 1990), a larger pool of subjects would serve to strengthen the findings of the study and increase statistical power. Fourth, the subjects in this study were asked to review an analysis specification document. There are many other types of work products that may be reviewed in a software review. The results of this study may or may not generalize to reviews involving work products such as design documents, program code, or testing plans. While efforts were made to develop a realistic analysis specification document,
the characteristics of specifications (e.g., style, complexity) can vary widely in the workplace. Fifth, we note that the experimental task was somewhat simplified due to the fact that the subjects were working under the conditions of known documentation standards and user requirements (the narrative part of the specification was assumed to reflect the correct user requirements). In practice, some analysis specification reviews may have less of a Eureka nature due to debates arising from uncertainties regarding standards and user requirements. Lastly, it should be recognized that the type of EMS (and EMS tool) used can influence the findings for a study of this type (Dennis, et al., 1990-1991). An EMS with a different type of user interface and design might have had different impacts on process loss and gain for the groups.

7.4 Concluding remarks

The findings from this study indicate that software inspection teams may experience group process loss. Since any amount of process loss during inspections can ultimately result in an increase in project costs (i.e., the cost to correct a defect is greater if it is handled later in development rather than earlier), it behooves inspection teams to reduce process loss as much as possible by means of some form of group support. This study, the first controlled laboratory experiment to assess the comparative effectiveness of group support mechanisms for the software review task, suggests that group support can improve the performance of interactive review teams. Specifically, the results indicate
that: 1) EMS process support (in the form of parallel communication and group memory) can help to improve defect detection performance for teams identifying defects in an analysis specification, and 2) Task support can help to reduce group process loss for defect evaluation. It is hoped that the positive findings arising from this study may encourage others to pursue research aimed at improving the practice of software reviews.
APPENDIX A:

EXPERIMENTAL INSPECTION TASK: SYSTEM SPECIFICATION
SYSTEM SPECIFICATION SUMMARY

Project Name:
Ashland Real Estate Listings Information System

Producer:
Joe Sundevil

System Overview:
The purpose of the Ashland Real Estate system is to provide automated support for the firm's operations related to the listings of residential homes. (Note: A "listing" refers to a house which has been put on the housing market.) Operations include interfacing with home sellers and buyers, storing data in the firm's data bases (e.g., listings file, buyer file), and generating reports for the management and buyers. A context data flow diagram (DFD) and top-level DFD are provided to show the key processes and the external entities to the system (see Figures 1 and 2).

Overview of Processes:

- Process 1.0 (Process Seller Information):
The purpose of this subprocess is to process the information received from potential home sellers. Information includes data about the homes which are for sale (see data dictionary for specifics). Information is input into the system by data entry personnel and then stored in the listings file. This process is a functional primitive.

- Process 2.0 (Process Buyer Information):
The purpose of this subprocess is to process the information received from potential home buyers. Information includes data about the buyers and their requirements for a new house (see data dictionary for specifics). Information is input into the system by data entry personnel and then stored in the buyer file. This process is a functional primitive.

- Process 3.0 (Generate Reports):
The purpose of the Generate Reports subprocess is to generate two types of reports: 1) a summarized listing report for the firm's management (i.e., a report which summarizes all of the firm's house listings), and 2) customized listing reports for the potential home buyers. This process can be decomposed into several subprocesses (see Figure 3).
Narrative Description of the Generate Reports Subprocess:
Based on interviews with the users and management at Ashland Real Estate, the following description of the reporting process has been prepared:

- Two types of reports are generated every week: the management report and the customized reports for potential buyers.
- The management report is prepared by obtaining the listings records from the listings file. The listings file includes the homes which have been listed with the Ashland Real Estate firm. The listings are summarized and a report is generated for management.
- The customized listing report for each potential buyer (i.e., client) is prepared by accessing data from the listings file and the buyer file. The contents of the listings file is compared with each potential buyer's stated requirements for a new home (e.g., number of bedrooms and bathrooms, size of lot). Each buyer's requirements are accessed from the buyer file. Homes listed by Ashland Real Estate which are suitable for each buyer are identified and summarized into a customized listing report.
- A mailing label is prepared for each customized listing report by accessing the appropriate potential buyer's address from the buyer file. Each customized report is then mailed to the appropriate potential buyer after affixing the mailing label to the report.
- A DFD describing the Generate Reports process is attached (see Figure 3).

Data Dictionary Entries

- External Entities (Sources/Sinks)
  - Seller: Sellers enlist the services of Ashland Real Estate to help sell their homes.
  - Buyer: Buyers enlist the services of Ashland Real Estate to help find a home to purchase.
  - Management: The management of Ashland Real Estate.
o **Data Stores (including associated data structures)**

  o **Buyer File = \{Buyer_file_record\}**

    ```
    Buyer_file_record = Buyer_name +
    Buyer_requirements_record
    
    Buyer_name = First_name + Last_name
    
    Buyer_requirements_record = Number_of_bedrooms +
    Number_of_bathrooms +
    Square_foot_size +
    House_style +
    Lot_size +
    School_district
    ```

  o **Listings File = \{Listings_file_record\}**

    ```
    Listings_file_record = Listing_number +
    Listing_date +
    Listing_sales_terms +
    Listing_price +
    House_address_record +
    Number_of_bedrooms +
    Number_of_bathrooms +
    Square_foot_size +
    House_style +
    Lot_size +
    School_district
    
    House_address_record = Street_address +
    City +
    State +
    Zip_code
    ```

    (Note: the brackets (\{\}) represent a repeating group.)

  o **Input/Output Data Flows (shown on Figure 3)**

    o **Listings = Listing_number +
    Listing_date +
    Listing_sales_price +
    House_address_record +
    Number_of_bedrooms +
    Number_of_bathrooms +
    Square_foot_size +
    House_style**
Input/Output Data Flows (cont.)

- Management_report = {Listing_number +
  Listing_date +
  Listing_sales_price +
  House_address_record +
  Number_of_bedrooms +
  Number_of_bathrooms +
  Square_foot_size +
  House_style}

- Buyer_requirements = Buyer_name + Buyer_requirements_record

- Buyer_address = Buyer_name + Buyer_address_record

- Customized_report = Buyer_name +
  {Listing_number +
  Listing_date +
  Listing_sales_price +
  House_address_record +
  Number_of_bedrooms +
  Number_of_bathrooms +
  Square_foot_size +
  House_style}

- Customized_report_with_mailing_label =
  Buyer_name +
  {Listing_number +
  Listing_date +
  Listing_sales_price +
  House_address_record +
  Number_of_bedrooms +
  Number_of_bathrooms +
  Square_foot_size +
  House_style} +
  Buyer_address_record

Mini-Specifications

Mini-specs have not yet been prepared for functional primitives.
Figure 1: Context DFD
Figure 2: Top-Level DFD
Figure 3: Diagram 3.0
APPENDIX B:

EXPERIMENTAL INSPECTION TASK: LIST OF DEFECTS
Inspection Exercise: List of Defects
Spring/Fall 1992

Attached is a list of defects which exist the "Systems Specification" for the Ashland Real Estate System (i.e., the specification given to subjects for the experiments in Fall 1991, Spring 1992, and Fall 1992). Defects are categorized by the basic type of error: completeness, consistency, communicability, and correctness. This list is used to summarize the defects detected by individuals and groups in the experimental setting.

Group Session (Date/Time):

Analysis Date:

Group Member Codes: A, B, C, D, E (if needed)

Group Decision Codes:
Defect: circle
Suggestion: diamond
Open Issue: square

Summary of Results:

<table>
<thead>
<tr>
<th>Group Member</th>
<th>No. of Errors Detected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C1</td>
</tr>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
</tr>
</tbody>
</table>

Group: Defect

Suggest

Open
1. Completeness Errors

A. Defects:

- **Data Dictionary**
  - DF: Missing entries for DFs from Figure 3 (i.e., not all DFs from Fig 3 are listed in the I/O section: listing report, homes..., etc.) DD is also missing entries for DFs from Figs 1 and 2.
  - DF: "Listings" DF should include the elements 'lot size' and 'school district'.
  - DF: The "Customized" and "Management report" DFs should include the elements 'lot size' and 'school district'.
  - DS: Missing a definition for "Buyer address" (which is mentioned in the Buyer_file_record definition).
  - DS and DF: Information should be provided regarding the limits on repeating groups (e.g., 1[File_record]*).

- **Data Elements**: Values & meanings information is not included.
- **MiniSpecs**: Missing for all functional primitives.

- **Diagrams**
  - Figs 1-3: Diagram labeling is vague, poorly named, could be more descriptive (e.g., "Top Level DFD" label for Fig. 2 could be "Level-0 DFD: Ashland Real Estate System.")
  - Figs 2, 3: Process bubbles are missing a "*" or "P" to indicate functional primitive processes (which are defined in the narrative, e.g., Processes 1.0, 2.0 in Fig. 2).

2. Consistency Errors

A. Defects:

- **Data Dictionary**
  - DS vs DF: DS: Listings_file: listing price
  - DF: Listings: listing sales price

- **Diagrams**
  - DF: Fig 1, Missing a DF "customized report" to Buyer BB
  - DF: Fig 2, Label for DF from Seller (SS) should be sales_info (not seller_info)
  - DF: Fig 3, arrow is going in wrong direction for "mgmt rpt"
  - SS: Fig 3, Label 'Client' should be 'Buyer'.

-2-
3. Communicability Errors

A. Defects:

- **Data Dictionary**
  - DSs "Management_report," "Customized_report," and "Customized_report_with_mailing_label" have the same contents as "Listings". Why not use "Listings" in DD entries to make it easier to read?
  - DS: Contents of Listings_File and Buyer_File overlap. Any way to avoid this (via common record defns)?
  - DS: Should partition Listings_File into aggregated records for easier reading (4-16-6pm, EMS)
  - DS: Listing and Buyer labels are too general (e.g., rename Listing_File as Home Listing_File for clarity).

- **Data Flows (on Figs)**
  - Labeling in Fig. 2: Why are different DFs labels used for the DFs entering and leaving the Listings_File? Is there a difference between "Listings_information" and "Listings"? Also, why are DFs with different labels entering and leaving the Buyer_File?
  - DF labeling (general): Poor/odd hyphenation on several labels (e.g., EE: Manage-ment on Figs 1-3). Also, several labels seem rather cryptic (abbreviated and use "w/"). DF label for "Cust_report_with_mailing_label" is rather long.
  - Label positioning (general): Labels for some DFs are too far from the lines.
  - Labeling: Label for "Listings" and "Management_report" could be improved (these label names are a bit vague).

- **Processes**
  - Labeling in Fig. 2: Labels for 1.0, 2.0, and 3.0 are not descriptive enough (e.g., label for 3.0 should include "produce labels").
  - Labeling in Fig. 3: Process labels for 3.1 (should be "Prepare Mgmt Report", not "Print Listing Report"), 3.3 (should be "Produce Customized Report"), and 3.4 (should be "Prepare Mailing Labels") are not correct (based on the narrative specification).
4. Correctness Errors

A. Defects

○ Data Flows

- Labeling in Fig. 1: Are the "update" DFs needed? What is the difference between the "information" DFs and the "updated information" DFs? (There is no explicit mention of updates in the narrative.)

- Fig 2: Should the DFD split up the "Buyer_rqts&_address" DF into two DFs?

- Needed: Fig 3, DF "Buyer_address_info" from Buyer_File to process 3.4.

- Needed: Fig 3, DF going from Listing DS to process 3.2. Not needed: Fig 3, "management report" from 3.1 to 3.2.

○ Processes

- Fig 3: Could combine/drop some processes, e.g., Process 3.4 does not need to be separate (it can be combined with 3.3); could combine 3.2 and 3.3; could drop 3.4. (Note: All of these are subjective, but are worth looking into as defects.)
APPENDIX C:

OVERVIEW HANDOUT FOR INSPECTION EXERCISE
INSPECTION REVIEW EXERCISE

Objective:

- The major objective of this exercise is to introduce the student to the concept of the "inspection review" through the use of a brief overview and "hands-on" experience.

- A secondary objective of the exercise is to help researchers at the University of Arizona to better understand the application of inspections to systems analysis and design.

Course Credit:

- This exercise is a required part of your course in systems analysis and design. Your performance on this exercise will be evaluated, so do your best!

- Since this exercise will be evaluated, it is important that you do NOT discuss the exercise with other students currently taking the MIS 141 systems analysis and design course.

Agenda:

- Pre-exercise questionnaire
- Overview of inspections
- Inspection exercise
- Questionnaires
SOFTWARE INSPECTION REVIEWS: OVERVIEW

What is an inspection?

An inspection is a peer group review of a work product for the purpose of identifying errors or deficiencies. Errors may include those relating to completeness, consistency, correctness, communicability, or adherence to standards. When approaching an inspection, reviewers should consider the product guilty until proven innocent. (However, the producer of the product is always innocent because he/she is not on trial.)

What an inspection is NOT

The major purpose of an inspection is error detection, NOT error correction. Let the producer of the work product determine how to resolve the error. Also, the inspection is NOT used for individual performance appraisal: it is performed to improve the quality of a work product.

Why perform inspections?

Inspections are justified on an economic basis. Inspections have proven to be very effective in identifying errors early in the system development process. Since the cost of fixing an error rises dramatically as a project proceeds in the development cycle, it is worthwhile to perform inspections to catch errors early in the process.

Types of inspections

Within the area of systems analysis and design there may be several different types of inspections depending on the type of work product being reviewed: analysis inspections, design inspections, code inspections, and test inspections.

When should an inspection be conducted?

Inspections should be conducted frequently, so that errors may be found as early as possible in the development process. When a unit of documentation (e.g., a DFD, structure chart) is completed and available for review, an inspection should be scheduled.
Who participates in an inspection?

The types of participants depend on the purpose of the inspection and may include users, analysts, designers, programmers, consultants, and managers. Key roles for participants include the moderator, the coordinator, the producer, and the scribe.

Guidelines for conducting the inspection

- The number of participants should be manageable (3-6 people)
- Participants must agree to abide by a set of "ground rules"
  - Emphasize error detection, not correction
  - Participants should be charitable critics
  - Participants should be objective
  - Use standards to avoid disagreements
- Make the inspection only as formal as necessary
- Keep the inspection as short as possible
INDIVIDUAL REVIEW: PROBLEM LIST

Reviewer: ____________________________________________

Project: _____________________________________________

Product/Producer: _____________________________________

Date of Individual Review: ______________________________

Problem List

1. List all problems or possible problems that you can identify regarding the specification.
2. Do NOT list problems identified by other team members.
3. Designate "problem type" using one of the following codes:
   - 1 = completeness
   - 2 = consistency
   - 3 = communicability
   - 4 = correctness

4. Describe each problem in enough detail to clearly identify the specific nature of the problem. Include a reference to the specific document in question (e.g., Figure 1, Data Dict.).

<table>
<thead>
<tr>
<th>Problem Type</th>
<th>Description (plus document ref.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
</tr>
</tbody>
</table>
INSPECTION REPORT FORM

Coordinator: ____________________________________________

Project: ________________________________________________

Product/Producer: ________________________________________

Coordinator's Checklist:

1. Confirm with producer(s) that material is ready and stable: ____

2. Issue invitations, schedule meeting room, and distribute materials

Date: ________ Time: ________ Duration: ________

Place: ________________________________________________

Participants Can Received
Can Received
attend? materials?

1. ________________________________________________ ________

2. ________________________________________________ ________

3. ________________________________________________ ________

4. ________________________________________________ ________

5. ________________________________________________ ________

6. ________________________________________________ ________

Agenda:

1. All participants agree to follow the "ground rules"

2. New project: Perform inspection of the materials
   Old project: Perform item-by-item checkoff of the previous action list

3. Creation of new action list

4. Group decision

5. Deliver copy of this form to project management

Decision: ________ Accept product "as-is"

______ Revise (no additional inspection required)

______ Revise and schedule another inspection
INSPECTION ACTION LIST

Coordinator: ________________________________
Project: ________________________________
Product/Producer: __________________________
Date of Inspection: __________________________

Action Items

Notes:

1. Designate "problem type" using one of the following codes:
   1 = completeness
   2 = consistency
   3 = communicability
   4 = correctness

2. Designate the "disposition" code using one of the following:
   D = defect
   S = suggestion
   O = open issue (i.e., an issue/problem that requires more information and discussion before classification as a defect)

3. Describe each problem such that the producer will be able to easily recollect/understand the specific nature of the problem. Include a reference to the document.

<table>
<thead>
<tr>
<th>Problem Type Code</th>
<th>Disposition Code</th>
<th>Description (plus document ref.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX D:

EXPERIMENT 1: DIRECTIONS FOR BASELINE TREATMENT
STRUCTURED WALKTHROUGH EXERCISE:
DIRECTIONS

Individual Review Exercise

- The purpose of the individual review exercise is to provide you with an opportunity to become familiar with the analysis specification prior to the team walkthrough.
- Your task is to review the analysis specification that is provided and identify all the errors that you find.
- Use the "checklist" to guide you as you review the specification.
- Use the "Individual Review: Problem List" to record your findings.
- There may be some cases in which you are not sure whether you have identified a problem or not. In these cases, record the (possible) problem on your list. You may discuss the problem with your team during the team walkthrough.
- Remember that the goal of a walkthrough is detection, NOT correction.
- You will be evaluated on the overall quality of your review (types of errors detected) and number of errors detected.
- Do NOT make any changes to your individual problem list once the individual review period has ended.

Team Walkthrough Exercise

- The purpose of the team walkthrough exercise is to give your team some experience in performing a group walkthrough.
- Review the "Walkthrough Report Form"
- Conducting the walkthrough:
  - Use the "checklist" to help organize the walkthrough process.
  - Use the findings from the individual reviews.
  - Use the "Action List" to record your team's findings (the scribe will record the findings).
- Before ending the walkthrough determine the group "decision."
- Your team will be evaluated on the overall quality of the review (types of errors detected) and the number of errors that are correctly detected and classified.
- At the conclusion of the team walkthrough exercise turn in your individual problem list and the system specification summary.
APPENDIX E:

EXPERIMENT 1: DIRECTIONS FOR MANUAL TREATMENT
STRUCTURED WALKTHROUGH EXERCISE:
DIRECTIONS

Individual Review Exercise

- The purpose of the individual review exercise is to provide you with an opportunity to become familiar with the analysis specification prior to the team walkthrough.

- Your task is to review the analysis specification that is provided and identify all the errors that you find.

- Use the "checklist" to guide you as you review the specification.

- Use the plastic overhead sheets to record your findings. Follow the same format as indicated in the "Individual Review: Problem List."

- There may be some cases in which you are not sure whether you have identified a problem or not. In these cases, record the (possible) problem on your list. You may discuss the problem with your team during the team walkthrough.

- Remember that the goal of a walkthrough is detection, NOT correction.

- You will be evaluated on the overall quality of your review (types of errors detected) and number of errors detected.

- Do NOT make any changes to your individual problem list once the individual review period has ended.

Team Walkthrough Exercise

- The purpose of the team walkthrough exercise is to give your team some experience in performing a group walkthrough.

- Review the "Walkthrough Report Form."

- Conducting the walkthrough:
  - Use the "checklist" to help organize the walkthrough process.
  - Use the findings from the individual reviews.
  - Use the "Action List" to record your team's findings (the moderator/scribe will record the findings).

- Before ending the walkthrough determine the group "decision."

- Your team will be evaluated on the overall quality of the review (types of errors detected) and the number of errors that are correctly detected and classified.

- At the conclusion of the team walkthrough exercise turn in your individual problem list and the system specification summary.
APPENDIX F:

EXPERIMENT 1: DIRECTIONS FOR EMS TREATMENT
STRUCTURED WALKTHROUGH EXERCISE:  
DIRECTIONS

**Individual Review**

- The purpose of the individual review is to provide you with an opportunity to become familiar with the analysis specification prior to the team walkthrough.
- Your task is to review the analysis specification that is provided and identify all the errors that you find.
- Use the "checklist" to guide you as you review the specification.
- Use the GroupSystems software to record your findings (see attachment for directions and procedure).
- There may be some cases in which you are not sure whether you have identified a problem or not. In these cases, record the (possible) problem on your list. You may discuss the problem with your team during the team walkthrough.
- Remember that the goal of a walkthrough is detection, NOT correction.

**Team Walkthrough**

- The purpose of the team walkthrough exercise is to give your team some experience in performing a group walkthrough.
- Review the "Walkthrough Report Form"
- Conducting the walkthrough:
  - Walkthrough the errors/problem items identified by the members of the team.
  - Team members may record new electronic comments on the meeting record.
  - Moderator will record the team’s assessment of specific errors (i.e., disposition code, etc.)
- Before ending the walkthrough determine the group "decision."
- Your team will be evaluated on the overall quality of the review (types of errors detected) and the number of errors that are correctly detected and classified.
DIRECTIONS FOR USING GROUPSYSTEMS: GROUP OUTLINER

- Initially the user screen will show an outline that looks something like this:

  Session: Ashland Real Estate 10/20/91 4 pm

  Jack
  -- completeness
  -- consistency
  -- communicability
  -- correctness

- You will need to categorize the errors that you find into different categories (as defined by the checklist discussed earlier). Enter the errors that you find into the appropriate electronic window (i.e., completeness, consistency, communicability, correctness):

  - Initially, only you will have access to each window.
  - To enter a window use the up or down arrow keys to move the orange cursor around on the outline tree and place the cursor on the desired window heading. Then hit "enter."
  - Note: Do NOT enter comments into the window represented by the root node of the tree (e.g., "Jack" in the example above).

  - You will see a split screen with two windows:
    + The top (blue) window shows the comments already existing in the file.
    + The bottom (black) window is an input window for contributing new comments to the file.

  - Type in your comments in the bottom window.
    + Precede each comment with an "identification code" consisting of the letter and the number of the comment (e.g., "A.1. DFD is not complete," "A.2. Data dictionary is not complete").
    + Describe each problem in enough detail to clearly identify the specific nature of the problem to the rest of the team. Include a reference to the specific document in question (e.g., Figure 1, Data Dictionary).
    + Submit/save your comments using the F9 ("send") key.

  - To exit one window and enter a window for another category of problem:
    + Close the window that you are working within by using the F3 ("send/close") key.
    + You will return to the outline figure.
    + Move orange cursor to the desired window category.
    + Hit enter key to enter the window.
APPENDIX G:

CHECKLIST FOR A SOFTWARE ANALYSIS INSPECTION
Guidelines for an Analysis Inspection: A Checklist

This handout provides specific guidelines to aid the reviewer in performing a systems analysis inspection (i.e., an inspection of a specification which has been prepared to describe the users' requirements for a system). Four relevant criteria for evaluating an analysis system specification are: 1) completeness, 2) consistency, 3) communicability, and 4) correctness (the "4 Cs"). A checklist of factors relevant to each criteria is provided below.

1) Completeness

It is important to ensure that the specification is complete. The reviewer needs to consider the following questions when examining a specification for completeness. (Note: Data dictionary = DD.)

- For each diagram in a set of DFDs:
  - Is the diagram number and name correct?
- For each component of a diagram:
  - Process "bubble":
    - Are all bubbles labeled?
    - Are all bubbles numbered?
    - Is the process numbered?
    - If the process is a functional primitive:
      --> - Does the mini-specification (MS) exist in the DD?
      - Are all data elements, data flows and data stores referenced in the MS defined in the DD?
    - If the process is not a primitive:
      --> - Does the "child" (i.e. lower-level) DFD exist?
  - External entity (EE):
    - Are all EEs labeled?
    - Does a DD entry exist for each EE?
  - Data flow (DF):
    - Are all DFs labeled?
    - Does a DD entry exist for each DF?
  - Data store (DS):
    - Are all DSs labeled?
    - Does a DD entry exist for each DS?
2) Consistency

After ensuring that the specification is complete, it is necessary to examine the specification for inconsistencies and contradictions. The reviewer needs to consider the following questions when examining a specification for consistency:

- Are the labels on the DFD symbols unique?:
  - Process bubble
  - External entity
  - Data flow
  - Data store

- Do the labels appearing in the DD match the labels on the DFD symbols?

- Do superfluous entries exist in the DD?

- Are the DFDs balanced? (i.e., do the input and output data flows associated with a "parent" process bubble match the net inputs/outputs for that bubble's "child" DFD?)

- For each functional primitive process shown on a DFD: Are the input/output data flows shown on the DFD consistent with the inputs/outputs described in the associated mini-spec?

3) Communicability

Once a specification has been examined for completeness and consistency it is necessary to evaluate whether the specification will help to communicate the user requirements to those involved with the analysis and design process (e.g., users, analysts, designers). Issues to consider include the following:

- Is each DFD easy to understand?
  - Has the DFD been prepared in a readable style?
    - Is the diagram legible?
    - Are there any crossing lines?
    - Are the objects too close together?
    - Is it difficult to determine which labels go with which data flows?
    - Are there too many process bubbles on one diagram (i.e., more than seven bubbles per DFD)?

- Have the label names for the DFD symbols and data elements been well chosen?
  - Are the label names hard to understand and cryptic?
- Has each process bubble been labeled with the appropriate "verb-object" name?
- Is the name of an external entity too vague?
- Does the name of a data flow reflect its contents?
- Does the name of a data store reflect its contents?

4) Correctness

Lastly the reviewer needs to consider the following questions when examining a specification for correctness:

- Overall, does the specification correctly represent the system of interest? (i.e., Does the spec reflect the user requirements as defined by the narrative?)

- With regard to the process bubbles:
  - Is the transformation performed by each process bubble essential?
  - Are there any process bubbles which are missing?
  - Are there any process bubbles which are:
    - "Black holes" (i.e., incoming data flows, but NO outgoing data flows)
    - "Miracles" (i.e., outgoing data flows, but NO incoming data flows)

  - Conservation of data: Are the outputs produced by each process bubble derivable from the inputs? (In other words, what comes out must have gone in.)
  --> Are there any missing data flows?

  - Are there any unnecessary or superfluous data flows?

- With regard to the mini-specifications:
  - Is the mini-spec correct?
  - Is the logic represented properly?
  - Is the structured English/pseudocode written properly?

- With regard to the data stores (OS):
  - Is the DS essential? Could it be replaced by a data flow?

  - Is the DS shown at the proper level?
    - Is the DS accessed by two or more process bubbles?
    - Is the DS associated with a functional primitive?

- With regard to the data elements in the data dictionary:
  - Is the specification of "values and meanings" or "range of values" correct?
APPENDIX H:

EXPERIMENT 1: PRE-SESSION QUESTIONNAIRE
INDIVIDUAL QUESTIONNAIRE

Social Security Number: ________________________________

For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. I am confident that I have identified all of the errors in the system specification. 1 2 3 4 5 6 7
2. I am confident that all of my findings are correct. 1 2 3 4 5 6 7
3. I have a clear understanding of the problems existing within the systems specification reviewed during the exercise. 1 2 3 4 5 6 7
4. The people in my project team are friendly. 1 2 3 4 5 6 7
5. The people in my project team take a personal interest in me. 1 2 3 4 5 6 7
6. I trust the people in my project team. 1 2 3 4 5 6 7
7. I look forward to meetings with my project team. 1 2 3 4 5 6 7
8. During meetings with my project team, I generally feel uncomfortable in expressing my findings to the rest of the team. 1 2 3 4 5 6 7
9. During meetings with my project team, I feel that I always have a fair opportunity to express my findings and ideas to the rest of the team. 1 2 3 4 5 6 7
10. During meetings with my project team, I feel that it is difficult to get others to listen to me. 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

11. During meetings with my project team, I feel that there is usually at least one person in the team who tends to dominate other team members. 1 2 3 4 5 6 7

12. As a result of participating in the individual review portion of this exercise, I gained knowledge that will help me to prepare higher quality systems specifications in the future. 1 2 3 4 5 6 7

13. I am confident of my abilities to serve as an effective reviewer of an analysis specification. 1 2 3 4 5 6 7

The following questions concern your interaction with your project team. For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Very much</th>
<th>Medium</th>
<th>Very Little</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

14. How frequently are YOUR ideas and opinions used by the team? 1 2 3 4 5 6 7

15. How much do YOU typically ask questions or interact with the team? 1 2 3 4 5 6 7

16. How would you assess YOUR influence with respect to the team’s activities? 1 2 3 4 5 6 7

17. To what extent are you able to have YOUR opinions actually considered by others? 1 2 3 4 5 6 7

18. How successful are YOU in asserting your opinion? 1 2 3 4 5 6 7

19. To what extent does the team’s work on the team project to date reflect YOUR contributions? 1 2 3 4 5 6 7
If you used a computer for this exercise, please answer questions 20-21. Otherwise skip these questions and turn in the questionnaire.

20. My typing skills have NOT impaired my ability to record my findings for the individual review.  
   1 2 3 4 5

21. How would you describe your typing skills (circle one)?
   1. None
   2. Hunt and peck
   3. Casual (rough draft with errors)
   4. Good (can do 25 words per minute without errors)
   5. Excellent (can do 40 words per minute without errors)
APPENDIX I:

EXPERIMENT 1: POST-SESSION QUESTIONNAIRE
POST-SESSION QUESTIONNAIRE

Social Security Number: ______________________

The following questions relate to your perceptions relating specifically to the group portion of the exercise (i.e., the part of the exercise which occurred after the individual review).

For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Performing this exercise with my project team was a pleasant experience. 1 2 3 4 5 6 7

2. I would be willing to perform other exercises of this type with my project team in the future. 1 2 3 4 5 6 7

3. I am committed to my team’s findings and recommendations resulting from this exercise. 1 2 3 4 5 6 7

4. I am satisfied with my team’s findings and recommendations resulting from this exercise. 1 2 3 4 5 6 7

5. I am confident that my team has identified all of the errors in the system specification. 1 2 3 4 5 6 7

6. I am confident that all of my team’s findings are correct. 1 2 3 4 5 6 7

7. The team was very effective in performing this exercise. 1 2 3 4 5 6 7

8. The group portion of this exercise uncovered valid errors that I did NOT detect in my individual review. 1 2 3 4 5 6 7

9. The group portion of this exercise made me reevaluate the findings that I developed in my individual review. 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10. The process that we went through during the group portion of this exercise helped ME to generate additional findings. 1 2 3 4 5 6 7

11. The process that we went through during the group portion of this exercise helped OTHER members of the team to generate additional findings. 1 2 3 4 5 6 7

12. Everyone on the team contributed about the same amount to this group exercise. 1 2 3 4 5 6 7

13. At least one person on the team was a "free loader" during this exercise. 1 2 3 4 5 6 7

14. During the the group portion of this exercise, I felt uncomfortable in expressing my findings to the rest of the team. 1 2 3 4 5 6 7

15. During the group portion of this exercise, I felt that I always had a fair opportunity to express my findings and ideas to the rest of the team. 1 2 3 4 5 6 7

16. During the group portion of this exercise, I felt that it was difficult to get others to listen to me. 1 2 3 4 5 6 7

17. During the group portion of this exercise, I felt that there was at least one person in the team who tended to dominate other team members. 1 2 3 4 5 6 7

18. I have a clear understanding of the problems existing within the systems specification reviewed during the exercise. 1 2 3 4 5 6 7

19. As a result of participating in the group portion of this exercise, I gained knowledge that will help me to prepare higher quality systems specifications in the future. 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

20. I did not learn anything new in the group portion of this exercise. 1 2 3 4 5 6 7

21. I am confident of my abilities to serve as an effective reviewer of an analysis specification. 1 2 3 4 5 6 7

22. During the group portion of the exercise, it was frequently difficult to keep track of what was being discussed. 1 2 3 4 5 6 7

23. During the group portion of the exercise, I found that I was experiencing "information overload." 1 2 3 4 5 6 7

24. The group portion of this exercise was generally conducted in a productive and efficient fashion. 1 2 3 4 5 6 7

25. There were several times during the group portion of the exercise when the team got side-tracked on issues that were "off the point." 1 2 3 4 5 6 7

26. The scribe/moderator played a neutral role during the group portion of this exercise. 1 2 3 4 5 6 7

27. The findings resulting from the group portion of the exercise were strongly influenced by the scribe/moderator. 1 2 3 4 5 6 7

28. If the scribe/moderator had not participated in the group portion of this exercise then our group would have developed a different set of findings. 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Very much</th>
<th>Medium</th>
<th>Very Little</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

29. How would you rate the level of YOUR participation in this exercise? 1 2 3 4 5 6 7

30. How frequently were YOUR ideas and opinions used by the team? 1 2 3 4 5 6 7

31. How much did YOU ask questions or interact with the team? 1 2 3 4 5 6 7

32. How would you assess YOUR influence with respect to the team's development of the walkthrough Action List? 1 2 3 4 5 6 7

33. To what extent were you able to have YOUR opinions actually considered by others? 1 2 3 4 5 6 7

34. How successful were YOU in asserting your opinion? 1 2 3 4 5 6 7

35. To what extent does the team's final output from this exercise reflect YOUR contributions? 1 2 3 4 5 6 7

36. During the team walkthrough how much conflict was there between yourself and others? 1 2 3 4 5 6 7

37. How much disagreement were YOU directly involved in? 1 2 3 4 5 6 7

38. To what extent were differences of opinion resolved to the mutual satisfaction of conflicting parties? 1 2 3 4 5 6 7

39. To what extent were conflicts resolved to YOUR satisfaction? 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

Very much: 1 2 3 4 5 6 7

If you used a computer for this exercise, please answer questions 40-42. Otherwise skip these questions and turn in the questionnaire.

40. I was satisfied with the computer system used for this exercise. 1 2 3 4 5 6 7

41. The computer system was very helpful in enabling the team to perform the group portion of this exercise. 1 2 3 4 5 6 7

42. For this type of group task, the computer-based approach is more effective than a comparable non-automated approach. 1 2 3 4 5 6 7
APPENDIX J:

EXPERIMENT 2: DIRECTIONS FOR BASELINE TREATMENT
INSPECTION REVIEW EXERCISE:  
AGENDA AND DIRECTIONS (BASE)  
Spring/Fall 1992  

Introduction  
- Pre-exercise Questionnaire  
- Inspections: Overview  
- "Inspection Report Form"  
- Two key activities: 1) Identification of Errors  
  2) Evaluation of Errors  

Group Session: Identification  
- **Purpose**: Detect all potential errors existing within the system specification document  
- **Time Alloted**: 45 minutes  
  - Individual review: 15 min  
  - Interactive team review: 30 min  
- **Individual Review**: Take 15 minutes to read through the analysis specification and become familiar with it.  
- **Interactive Team Review**: Your team will have 30 min to identify the errors existing within the specification.  
  1. Use the "checklist" as a guide for identifying potential problems  
  2. **Assume** that the written narrative section of the specification is correct.  
  3. Whenever YOU identify a potential error:  
     - **Record** the potential error that YOU find on the Problem List which has been provided (**categorize** your findings into the appropriate "4 Cs" category).  
     - **Verbally** tell the other members of your team about your finding.
- Do NOT attempt to figure out how to correct the error. You only need to DETECT the error.

- "Free-wheeling" is welcome. Even if you are not sure whether you have identified an actual problem, go ahead and make a note of it on your Problem List AND mention it to your team. (Your team will have the chance to discuss the problem later during the Evaluation Session.)

4. Whenever SOMEONE ELSE in your team identifies an error:

- Pay attention to your teammate's verbal remarks. Perhaps THEIR findings will help to stimulate YOUR search for errors.

- Do NOT criticize or judge your teammate's comments.

- Do ask for clarification if needed.

- Do NOT record someone else's contribution on your Problem List. (Allow each person to be responsible for their OWN list.)

Notes: The instructor will be present to serve as a moderator and help your team with procedural questions.

Identification Sessions will be tape-recorded.

- Post-Session Questionnaire: Team members are required to complete the questionnaire at the conclusion of the Identification Session.
Group Session: Evaluation

- **Purpose:** Develop the walkthrough "Action List" which lists all problems with the specification document plus the appropriate "disposition codes" (determined by the team).

- **Time Allotted:** 45 minutes

- **Interactive Team Review:**
  1. Team members will verbally *introduce* and *discuss* each of the errors that they identified in the Identification Session.
     
     Organize discussion by error type: Completeness, ...

  2. The team will need to decide on the "disposition code" for each error (i.e., defect, suggestion, open issue).

  3. The instructor will serve as the moderator and scribe for this Session.

  4. Review the Inspection Report Form checklist and the "group decision"

**Notes:** The instructor will be present to serve as a moderator and help your team with procedural questions.

Evaluation Sessions will be tape-recorded.

- **Post-Session Questionnaire:** Team members are required to complete the questionnaire at the conclusion of the Evaluation Session.

- **Turn in** your Questionnaires, Problem List, and Handouts.
APPENDIX K:

EXPERIMENT 2: DIRECTIONS FOR MANUAL TREATMENT
INSPECTION REVIEW EXERCISE:
AGENDA AND DIRECTIONS (MANUAL)

Spring/Fall 1992

Introduction

- Pre-exercise Questionnaire
- Inspections: Overview
- "Inspection Report Form"
- Two key activities: 1) Identification of Errors 2) Evaluation of Errors

Group Session: Identification

- **Purpose**: Detect all potential errors existing within the system specification document.

- **Time Allotted**: 45 minutes
  
  Individual review: 15 min
  Interactive team review: 30 min

- **Individual Review**: Take 15 minutes to read through the analysis specification and become familiar with it.

- **Interactive Team Review**: Your team will have 30 min to identify the errors existing within the specification.

  1. Use the "checklist" as a guide for identifying potential problems
  2. **ASSUME** that the written narrative section of the specification is correct.
  3. Whenever **YOU** identify a potential error:

     - Try to **categorize** your finding into one of the "4 Cs" categories.

     - Verbally tell the other members of your team and the scribe/moderator about your finding. The scribe/moderator will record your finding for public display.
+ Do NOT attempt to figure out how to correct the error. You only need to DETECT the error.

+ "Free-wheeling" is welcome. Even if you are not sure whether you have identified an actual problem, go ahead and make a note of it on your Problem List AND mention it to your team. (Your team will have the chance to discuss the problem later during the Evaluation Session.)

4. Whenever SOMEONE ELSE in your team identifies an error:

+ Pay attention to your teammate's contributions. Perhaps THEIR findings will help to stimulate YOUR search for errors.

+ Do NOT criticize or judge your teammate's contribution.

+ DO ask for clarification if needed.

**Notes:** The instructor will be present to serve as a moderator and help your team with procedural questions.

Identification Sessions will be tape-recorded.

- **Post-Session Questionnaire:** Team members are required to complete the questionnaire at the conclusion of the Identification Session.
Group Session: Evaluation

ο Purpose: Develop the Inspection Action List which lists all problems with the specification document plus the appropriate "disposition codes" (determined by the team).

ο Time Allotted: 45 minutes

ο Interactive Team Review:

1. Team members will "walkthrough" and verbally discuss the list of errors that were identified in the Identification Session.
   - Organize discussion by error type: Completeness, ...

2. The team will need to decide on the "disposition code" for each error (i.e., defect, suggestion, open issue).

3. The instructor will serve as the scribe for this Session.

4. Review the Walkthrough Report Form checklist and the "group decision"

Notes: The instructor will be present to serve as a moderator and help your team with procedural questions.

Evaluation Sessions will be tape-recorded.

ο Post-Session Questionnaire: Team members are required to complete the questionnaire at the conclusion of the Evaluation Session.

ο Turn in your Questionnaires, Problem List, and reusable handouts.
APPENDIX L:

EXPERIMENT 2: DIRECTIONS FOR EMS TREATMENT
INSPECTION REVIEW EXERCISE:
AGENDA AND DIRECTIONS (EMS)

Spring/Fall 1992

Introduction

- Pre-exercise Questionnaire
- Inspections: Overview
- "Inspection Report Form"
- Two key activities: 1) Identification of Errors
  2) Evaluation of Errors

Group Session: Identification

- **Purpose**: Detect all potential errors existing within the system specification document.

- **Time Allotted**: 45 minutes
  
  Individual review: 15 min
  Interactive team review: 30 min

- **Individual Review**: Take 15 minutes to read through the analysis specification and become familiar with it.

- **Interactive Team Review**: Your team will have 30 min to identify the errors existing within the specification.
  1. Do NOT interact verbally with your other team members. ONLY interact via the electronic media.
  2. ASSUME that the written narrative section of the specification is correct.
  3. Use the "checklist" as a guide for identifying potential problems.
4. Whenever YOU identify a potential error:
   + Record the potential error using the GroupSystems software (see attachment for directions and procedure).
   - Categorize your findings into the appropriate "4 Cs" category by using the windows feature.
   - Identify yourself when making a comment by typing your name before your comment.

5. + "Free-wheeling" is welcome. Even if you are not sure whether you have identified an actual problem, go ahead and make a note of it using the GroupSystems software. (Your team will have the chance to discuss the problem later during the Evaluation Session.)

6. Whenever SOMEONE ELSE in your team identifies an error:
   + Pay attention to your other team member's contributions by scanning the electronic windows. Perhaps THEIR findings will help to stimulate YOUR search for errors.
   - Do NOT criticize or judge your teammate's contribution.

Note: The instructor will be present to serve as a moderator and help your team with procedural questions.

Post-Session Questionnaire: Team members are required to complete the questionnaire at the conclusion of the Identification Session.
Group Session: Evaluation

- **Purpose:** Develop the inspection "Action List" which lists all problems with the specification document plus the appropriate "disposition codes" (determined by the team).

- **Time Allotted:** 45 minutes

- **Interactive Team Review:**
  1. Team members will "walkthrough" and verbally discuss the list of errors that were identified in the Identification Session.
     
     + Organize discussion by error type: Completeness, ...
  2. The team will need to decide on the "disposition code" for each error (i.e., defect, suggestion, open issue).
  3. The instructor will serve as the moderator and scribe for this Session.
  4. Review the Inspection Report Form checklist and the "group decision"

**Notes:** The instructor will be present to serve as a moderator and help your team with procedural questions.

Evaluation Sessions will be tape-recorded.

- **Post-Session Questionnaire:** Team members are required to complete the questionnaire at the conclusion of the Evaluation Session.

- **Turn in** your Questionnaires and reusable handouts.
DIRECTIONS FOR USING GROUPSYSTEMS: GROUP OUTLINER

- Initially the user screen will show an outline that looks something like this:

  Session: Ashland Real Estate 10/20/92 4 pm

  Ashland_Real_Estate_10/20/92_4pm
  |--- completeness
  |--- consistency
  |--- communicability
  |--- correctness

- You will need to categorize the errors that you find into different categories (as defined by the inspection checklist). Enter the errors that you find into the appropriate electronic window (i.e., completeness, consistency, etc.):
  - All team members have "write" (append) access to each window
  - To enter a window use the up or down arrow keys to move the orange cursor around on the outline tree and place the cursor on the desired window heading. Then hit "enter."
  - Note: Do NOT enter comments into the window represented by the root node of the tree (e.g., "Ashland_Real_Estate..." in the example outline above).
  - You will see a split screen with two windows:
    + The top (blue) window shows the comments already existing in the group file.
    + The bottom (black) window is an input window for contributing new comments to the group file.
  - Type in your comments in the bottom window.
    + Precede each comment with you name and the number of the comment (e.g., "Bob.1: DFD is not complete").
    + Describe each problem in enough detail to clearly identify the specific nature of the problem to the rest of the team. Include a reference to the specific document in question (e.g., Figure 1, Data Dictionary)
    + Submit/save your comments using the F9 ("send") key.
  - To exit one window and enter a window for another category of problem:
    + Close the window that you are working within by using the F3 ("send/close") key.
    + You will return to the outline figure.
    + Move orange cursor to the desired window category.
    + Hit enter key to enter the window.
APPENDIX M:

EXPERIMENT 2: PRE-SESSION QUESTIONNAIRE
PRE-EXERCISE QUESTIONNAIRE
Spring/Fall 1992

Social Security Number: ____________________________

For the following questions, please rate your response using the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. The people in my project team are friendly. 1 2 3 4 5 6 7

2. The people in my project team take a personal interest in me. 1 2 3 4 5 6 7

3. I trust the people in my project team. 1 2 3 4 5 6 7

4. I look forward to meetings with my project team. 1 2 3 4 5 6 7

5. During meetings with my project team, I generally feel uncomfortable in expressing my findings to the rest of the team. 1 2 3 4 5 6 7

6. During meetings with my project team, I feel that I always have a fair opportunity to express my findings and ideas to the rest of the team. 1 2 3 4 5 6 7

7. During meetings with my project team, I feel that it is difficult to get others to listen to me. 1 2 3 4 5 6 7

8. During meetings with my project team, I feel that there is usually at least one person in the team who tends to dominate other team members. 1 2 3 4 5 6 7

9. I am confident of my abilities to serve as an effective reviewer of an analysis specification. 1 2 3 4 5 6 7
For the following questions, please rate your response using the following scale:

<table>
<thead>
<tr>
<th>Very much</th>
<th>Medium</th>
<th>Very Little</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4</td>
<td>5 6 7</td>
<td></td>
</tr>
</tbody>
</table>

10. How frequently are YOUR ideas and opinions used by the team? 1 2 3 4 5 6 7

11. How much do YOU typically ask questions or interact with the team? 1 2 3 4 5 6 7

12. How would you assess YOUR influence with respect to the team's activities? 1 2 3 4 5 6 7

13. To what extent are you able to have YOUR opinions actually considered by others? 1 2 3 4 5 6 7

14. How successful are YOU in asserting your opinion? 1 2 3 4 5 6 7

15. To what extent does the team's work on the team project to date reflect YOUR contributions? 1 2 3 4 5 6 7

16. How many times have you used the University of Arizona's computerized group support system? (circle one)
   1. Never
   2. 1-2 times
   3. 3-4 times
   4. 5-6 times
   5. 7 or more times

17. How would you describe your typing skills? (circle one)
   1. None
   2. Hunt and peck
   3. Casual (rough draft with errors)
   4. Good (can do 25 words per minute without errors)
   5. Excellent (can do 40 words per minute without errors)
For each of the following questions, please circle the answer which is the most appropriate for you.

18. At a party do you
   1. interact with many people, including strangers
   2. interact with only a few people that you know

19. At parties do you
   1. leave early, with decreasing energy
   2. stay late, with increasing energy

20. In your social groups do you
   1. keep abreast of other's happenings
   2. get behind on the news

21. When making a phone call do you
   1. rehearse what you'll say before making the call
   2. simply make the call

22. In company do you
   1. wait to be approached
   2. initiate conversation

23. Does new and non-routine interaction with others
   1. stimulate and energize you
   2. tire you out

24. Do you prefer
   1. many friends with brief contact
   2. a few friends with more lengthy contact

25. Do you
   1. find that you have little to say to strangers
   2. speak easily and at length with strangers

26. When the phone rings do you
   1. try to get to the phone first so that you can answer it
   2. hope that someone else will answer the phone

27. Are you more inclined to be
   1. somewhat reserved
   2. easy to approach
APPENDIX N:

EXPERIMENT 2: POST-IDENTIFICATION SESSION QUESTIONNAIRE
QUESTIONNAIRE: POST-IDENTIFICATION SESSION
Spring/Fall 1992

Social Security Number: ____________________________

The following questions relate to your perceptions regarding the Identification Session that your team has just completed (i.e., the part of the exercise relating to the identification of errors and deficiencies in the system specification).

For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. Performing this exercise with my project team was a pleasant experience.    1 2 3 4 5 6 7

2. I would be willing to perform another identification session of this type with my project team in the future.    1 2 3 4 5 6 7

3. I am committed to my team's findings resulting from this exercise.    1 2 3 4 5 6 7

4. I am satisfied with my team's findings resulting from this exercise.    1 2 3 4 5 6 7

5. I am confident that my team has identified ALL of the errors in the system specification.    1 2 3 4 5 6 7

6. I am confident that all of my team's findings are correct.    1 2 3 4 5 6 7

7. The team was very effective in performing this exercise.    1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. My team members uncovered valid problems that I would not have detected on my own.  

4. The process that we went through during the group session helped me to more rapidly generate additional findings.

5. Findings introduced by other members of my team prompted me to identify potential errors that I would not have identified if I had been working alone.

6. I frequently spent time recording my own findings.

7. I frequently spent time dictating my findings to the Scribe so that he could record my findings on the public display.

8. I frequently spent time listening to the verbal remarks made by other members of my team.

9. I frequently spent time reviewing the written/typed listings of remarks made by other members of my team.

10. Everyone on the team contributed about the same amount to this group session.

11. I put forward my best effort during the group session.
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

17. I felt uncomfortable in expressing my findings to the rest of the team. 1 2 3 4 5 6 7

18. I was reluctant to introduce one (or more) of my findings to the group since I thought that it might be perceived negatively. 1 2 3 4 5 6 7

19. During the group session, I felt that I always had a fair opportunity to express my findings and ideas to the rest of the team. 1 2 3 4 5 6 7

20. During the group session, I felt that it was difficult to get others to listen to me. 1 2 3 4 5 6 7

21. During the group session, I felt that there was at least one person in the team who tended to dominate other team members. 1 2 3 4 5 6 7

22. During the group session, I felt that there was at least one person in the team who tended to participate much more than the other team members. 1 2 3 4 5 6 7

23. I have a clear understanding of the problems existing within the systems specification reviewed during the exercise. 1 2 3 4 5 6 7

24. As a result of participating in the group session, I gained knowledge that will help me to prepare higher quality systems specifications in the future. 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

25. I did NOT learn anything new during the group session. 1 2 3 4 5 6 7

26. I am confident of my abilities to serve as an effective reviewer of an analysis specification. 1 2 3 4 5 6 7

27. During the group session, it was frequently difficult to keep track of what was being discussed. 1 2 3 4 5 6 7

28. During the group session, I found that I was experiencing "information overload." 1 2 3 4 5 6 7

29. The group session was generally conducted in a productive and efficient fashion. 1 2 3 4 5 6 7

30. There were several times during the group session when the team got side-tracked on issues that were "off the point." 1 2 3 4 5 6 7

31. I always knew the identity of the person who had introduced a finding to the team. 1 2 3 4 5 6 7

32. My team did not have enough time to adequately review the system specification. 1 2 3 4 5 6 7

33. A significant amount of our team's effort was spent on detecting errors/deficiencies in the system specification. 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

34. A significant amount of our team's effort was spent on correcting errors/deficiencies in the system specification. 1 2 3 4 5 6 7

35. A significant amount of our team's effort was spent on evaluating the errors/deficiencies in the system specification (i.e., deciding on the proper "disposition code"). 1 2 3 4 5 6 7

36. The scribe/moderator played a neutral role during the group session. 1 2 3 4 5 6 7

37. The findings resulting from the group session were strongly influenced by the scribe/moderator. 1 2 3 4 5 6 7

38. Our team would have developed a different set of findings if the scribe/moderator had NOT participated in the group session. 1 2 3 4 5 6 7

39. There were frequently times during the session when I did NOT pay attention to the contributions made by other members of my team. 1 2 3 4 5 6 7

40. There were frequently times during the session when I forgot the contributions made by other members of my team. 1 2 3 4 5 6 7
41. When you found a defect
Could you express it immediately

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

Did you have to wait to express it

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

42. Did you express
All of the findings that occurred to you

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

None of the findings that occurred to you

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

43. Did you express your findings
Soon after you found them

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

After waiting a while

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

44. How at ease were you during the error detection session?
Definitely not at ease

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

Neutral/Undecided

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

Very much at ease

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

45. Did you feel any apprehension about revealing your findings to the group?
A lot of apprehension

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

Neutral/Undecided

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
</table>

No apprehension

|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>This Situation</th>
<th>This Situation</th>
<th>This Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never Happened to Me</td>
<td>Occasionally Happened to Me</td>
<td>Frequently Happened to Me</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Please indicate HOW OFTEN the following situations happened to YOU:
I found an error/deficiency that I planned to introduce to the team, BUT I never did introduce my finding to my team BECAUSE:

46. The error/deficiency that I found was mentioned by somebody else before I had a chance to introduce it to the team. 1 2 3 4 5 6 7

47. I forgot what I wanted to say while waiting for someone else to finish communicating their remarks to the team. 1 2 3 4 5 6 7

48. A remark made by someone in the team made me think that my finding would not be worth mentioning. 1 2 3 4 5 6 7

For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Very much</th>
<th>Medium</th>
<th>Very Little</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

49. How would you rate the level of YOUR participation in this exercise? 1 2 3 4 5 6 7

50. How frequently were YOUR contributions used by the team? 1 2 3 4 5 6 7

51. How much did YOU ask questions or interact with the team? 1 2 3 4 5 6 7

52. To what extent does the team's output from the group session reflect YOUR contributions? 1 2 3 4 5 6 7

53. During the group session how much conflict was there between yourself and others? 1 2 3 4 5 6 7
If you used a computer for this exercise, please answer the following questions. Otherwise skip these questions and turn in the questionnaire.

For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

54. My typing skills did NOT impair my ability to record my findings.  
1 2 3 4 5 6 7

55. The computerized system was easy to use.  
1 2 3 4 5 6 7

56. I wasted time trying to figure out how to operate the computerized system.  
1 2 3 4 5 6 7

57. I was satisfied with the computer system used for this exercise.  
1 2 3 4 5 6 7

58. The computer system was very helpful in enabling the team to perform the tasks associated with the identification session.  
1 2 3 4 5 6 7

59. For this type of group task, the computer-based approach is more effective than a comparable non-automated approach.  
1 2 3 4 5 6 7
APPENDIX O:

EXPERIMENT 2: POST-EVALUATION SESSION QUESTIONNAIRE
QUESTIONNAIRE: POST-EVALUATION SESSION

Spring/Fall 1992

Social Security Number: __________________________

The following questions relate to your perceptions regarding the Evaluation Session that your group has just completed (i.e., the part of the exercise relating to the evaluation of errors and deficiencies in the system specification).

For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

1. Performing the evaluation exercise was a pleasant experience. 1 2 3 4 5 6 7

2. I would be willing to perform another evaluation session of this type with my project team in the future. 1 2 3 4 5 6 7

3. I am committed to my team's findings resulting from this exercise. 1 2 3 4 5 6 7

4. I am satisfied with my team's findings resulting from this exercise. 1 2 3 4 5 6 7

5. I am confident that my team has identified ALL of the errors in the system specification. 1 2 3 4 5 6 7

6. I am confident that all of my team's findings are correct. 1 2 3 4 5 6 7

7. The team was very effective in performing this exercise. 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4</td>
<td>5 6 7</td>
</tr>
</tbody>
</table>

8. The Evaluation Session encouraged me to **reassess** the findings that were generated during the identification session. 1 2 3 4 5 6 7

9. The process that we went through during the Evaluation Session helped ME to **generate** additional findings. 1 2 3 4 5 6 7

10. Everyone on the team contributed about the **same amount** to this the evaluation session. 1 2 3 4 5 6 7

11. I put forward my **best effort** during the evaluation session. 1 2 3 4 5 6 7

12. I felt uncomfortable in expressing my findings to the rest of the team during the evaluation session. 1 2 3 4 5 6 7

13. I was **reluctant** to introduce one (or more) of my comments to the group since I thought that it might be perceived **negatively**. 1 2 3 4 5 6 7

14. During the group session, I felt that I always had a **fair opportunity** to express my findings and ideas to the rest of the team. 1 2 3 4 5 6 7

15. During the evaluation session, I felt that it was **difficult** to get others to listen to me. 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

16. During the evaluation session, I felt that there was at least one person in the team who tended to dominate other team members. 1 2 3 4 5 6 7

17. During the evaluation session, I felt that there was at least one person in the team who tended to participate much more than the other team members. 1 2 3 4 5 6 7

18. I have a clear understanding of the problems existing within the systems specification reviewed during the exercise. 1 2 3 4 5 6 7

19. As a result of participating in the evaluation session, I gained knowledge that will help me to prepare higher quality systems specifications in the future. 1 2 3 4 5 6 7

20. I did NOT learn anything new during the evaluation session. 1 2 3 4 5 6 7

21. I am confident of my abilities to serve as an effective reviewer of an analysis specification. 1 2 3 4 5 6 7

22. During the evaluation session, it was frequently difficult to keep track of what was being discussed. 1 2 3 4 5 6 7

23. During the evaluation session, I found that I was experiencing "information overload." 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

24. The evaluation session was generally conducted in a **productive and efficient** fashion. 1 2 3 4 5 6 7

25. There were several times during the evaluation session when the team got **side-tracked** on issues that were "off the point." 1 2 3 4 5 6 7

26. I always knew the **identity** of the person who had introduced a comment to the team. 1 2 3 4 5 6 7

27. During the evaluation session my team did **not have enough time** to adequately evaluate our findings from the first group session. 1 2 3 4 5 6 7

28. A significant amount of our team's effort during the evaluation session was spent on **detecting errors/deficiencies** in the system specification. 1 2 3 4 5 6 7

29. A significant amount of our team's effort during the evaluation session was spent on **correcting errors/deficiencies** in the system specification. 1 2 3 4 5 6 7

30. A significant amount of our team's effort during the evaluation session was spent on **evaluating** the errors/deficiencies in the system specification (i.e., deciding on the proper "disposition code"). 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Undecided</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

31. The scribe/moderator played a neutral role during the evaluation session.  1  2  3  4  5  6  7

32. The findings resulting from the evaluation session were strongly influenced by the scribe/moderator.  1  2  3  4  5  6  7

33. Our team would have developed a different set of findings if the scribe/moderator had NOT participated in the evaluation session.  1  2  3  4  5  6  7

34. There were frequently times during the evaluation session when I did NOT pay attention to the contributions made by other members of my team.  1  2  3  4  5  6  7

35. There were frequently times during the evaluation session when I forgot the contributions made by other members of my team.  1  2  3  4  5  6  7
36. When you had a comment to make regarding a potential defect

Could you express it immediately

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Did you have to wait to express it

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

37. Did you express

All of the comments that occurred to you

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

None of the comments that occurred to you

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

38. Did you express your comments

Soon after you found them

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

After waiting a while

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

39. How at ease were you during the evaluation session?

Definitely not at ease

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Neutral/Undecided

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Very much at ease

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

40. Did you feel any apprehension about revealing your comments to the group?

A lot of apprehension

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Neutral/Undecided

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

No apprehension

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>This Situation</th>
<th>This Situation</th>
<th>This Situation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never Happened to Me</td>
<td>Occasionally Happened to Me</td>
<td>Frequently Happened to Me</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please indicate HOW OFTEN the following situations happened to YOU:

During the Evaluation Session, I planned to make a comment to the team, BUT I never did introduce my comment to the team BECAUSE:

41. The comment that I was going to make was mentioned by somebody else before I had a chance to introduce it to the team. 1 2 3 4 5 6 7

42. I forgot what I wanted to say while waiting for someone else to finish communicating their remarks to the team. 1 2 3 4 5 6 7

43. A remark made by someone in the team made me think that my comment would not be worth mentioning. 1 2 3 4 5 6 7

For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Very much</th>
<th>Medium</th>
<th>Very Little</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

44. How would you rate the level of YOUR participation during the evaluation session? 1 2 3 4 5 6 7

45. How frequently were YOUR contributions used by the team during the evaluation session? 1 2 3 4 5 6 7

46. How much did YOU ask questions or interact with the team during the evaluation session? 1 2 3 4 5 6 7

47. How would you assess YOUR influence with respect to the team's development of the Action List? 1 2 3 4 5 6 7
For the following questions, please rate your response on the following scale:

<table>
<thead>
<tr>
<th>Very much</th>
<th>Medium</th>
<th>Very Little</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

48. To what extent were you able to have YOUR opinions actually considered by others during the evaluation session? 1 2 3 4 5 6 7

49. How successful were YOU in asserting yourself during the evaluation session? 1 2 3 4 5 6 7

50. To what extent does the team's output from the evaluation session reflect YOUR contributions? 1 2 3 4 5 6 7

51. During the evaluation session how much conflict was there between yourself and others? 1 2 3 4 5 6 7

52. How much disagreement were YOU directly involved in during the evaluation session? 1 2 3 4 5 6 7

53. To what extent were differences of opinion resolved to the mutual satisfaction of the conflicting parties? 1 2 3 4 5 6 7

54. To what extent were conflicts resolved to YOUR satisfaction during the evaluation session? 1 2 3 4 5 6 7
APPENDIX P:

LETTER OF APPROVAL FROM HUMAN/ANIMAL SUBJECTS COMMITTEE
May 17, 1993

Craig K. Tyran, Ph.D.
Department of Management Information Systems
Mcclellan Hall
Main Campus

RE: THE SOFTWARE INSPECTION: THE TASK AND MECHANISMS FOR GROUP SUPPORT

Dear Dr. Tyran:

I received Dr. J. George’s 13 May 1993 telephone call and follow-up letter and accompanying material relating to your doctoral dissertation [as cited above]. As stated in the abstract, this study was undertaken to develop more information about the nature of the software review task and to assess impacts of selected group support mechanisms on review team performance. It is my understanding that upper division undergraduates participated in the experiment concerning detection and evaluation of errors in a system specification, and the dissertation based on it has been completed and successfully defended. However, as Dr. George states, prior approval for the research was not obtained from the Human Subjects Committee since the nature of the experiment was relevant to classroom material.

Based on information received, regulations published by the U.S. Department of Health and Human Services [45 CFR Part 46.101(b) (2)] exempt this type of research from review by our Committee. Therefore, retroactive approval of this exempt project is granted for purpose of dissertation only.

If you have any questions concerning the above, please contact this office.

Sincerely yours,

William F. Denny, M.D.
Chairman
Human Subjects Committee

WFD:rs

xc: Joey F. George, Ph.D., Assistant Professor
Department of Management Information Systems
REFERENCES


Shaw, M.E. and Ashton, N. "Do assembly effects occur on disjunctive tasks?," Bulletin of the Psychonomic Society, 8(6), 1976, 469-471.


