8-15-1997

# Team Synergy in Software Inspections: A Group Behavior Analysis

Bruce C. Hungerford
*University of South Florida*

Alan R. Hevner
*University of South Florida*

# Team Synergy in Software Inspections: A Group Behavior Analysis

**Bruce C. Hungerford** and **Alan R. Hevner**
Information Systems and Decision Sciences
College of Business Administration
University of South Florida
Tampa, FL 33620

**ABSTRACT**

Inspections provide many benefits in the software development process. However, the cost effectiveness of inspections has been criticized. Also, many organizations simply do not have the time to perform complete inspections of all software artifacts within the development schedule. Due to its pragmatic and domain-specific nature, little formal research has been performed on inspections. We propose to begin a study of software development inspections by surveying several relevant research papers on group behavior theory. We apply this research to inspections and outline a laboratory experiment for future research.

## 1. Software Inspections

Software development is a very complex and intellectually challenging activity. Team-based inspections of software development products, including requirements, designs, and code, have been shown to have significant improvements on software quality and development team productivity [Wheeler et al. 1996]. Briefly, software inspections can be used to provide the following benefits:

- Contain defects as close to their origin as possible
- Identify root causes of defects
- Improve the development process
- Provide uniform application of standards
- Support group interaction and egoless development
- Train personnel in software development techniques
- Produce metrics for analyses and baselines

However, the inspection process has been criticized as not cost-effective [Votta 1993] and many variations have been proposed to improve inspection techniques [Fagan 1986, Wheeler et al. 1996]. Since existing research is based on particular situations and does not involve sampling from a general frame or process, statistical generalization cannot be employed. Generalization from case results requires a theory of the underlying processes, but inspection techniques have developed pragmatically and without relation to underlying theory. Software inspections necessarily involve complex cognitive and social activities, but an extensive review of the literature finds no research discussing inspections from the perspective of group behavior theory.

The inspection **meeting** is a key element of the inspection process. Meetings are expensive in terms of staff time and potential development delays due to the lag between completion of the software artifact and its inspection. Proponents of group inspections argue that inspection meetings generate a **synergy** that increases the percentage of program defects identified and are therefore cost-effective. However, Votta [1993] finds that most program defects are identified by individual reviewers before the meeting and his cost analysis indicates that it is generally not cost-effective to incur the cost of a meeting to detect a few additional faults. He therefore proposes substituting a "deposition" or nominal group approach.

Whether Votta's results are the due to factors unique to his research or reflect a more general problem is a question of generalizability which cannot be answered statistically. However, group behavior theory does provide a framework for identifying and analyzing factors which may lead to synergy. Our research attempts to describe software inspections in terms of group behavior theory with an emphasis on factors important to group synergy. We begin by briefly surveying several relevant papers from group behavior

theory. Then, we discuss the relevance of this theory to inspections. A laboratory experiment to test one of the predictions arising from that theory is proposed.

## 2. Summary of Group Behavior Research

Conceptually, the inspection process depends on the assumption that group decisions will be better than the decisions of the most knowledgeable member. Evidence from some empirical research, however, indicates that under certain circumstances the knowledge base of the most competent group member represents the upper limit of group performance and that performance is often actually **inferior** to that of the best individual. It has been argued that these results are due, in large part, to the artificial nature of the groups, tasks, or settings in which the research has been conducted [Hackman and Morris 1976]. Two studies which investigate this possibility and which are directly relevant to the current research are Michaelsen et al. [1989] and Watson et al. [1991].

Michaelsen et al. [1989] studied individual versus group decision making using data from 222 team learning groups involved in 25 organizational behavior courses. Subjects were assigned to groups by the instructors using a methodology which attempted to minimize differences in potential resources. Due to differences in class sizes, students dropping the class, etc., group sizes ranged from 3 to 8, with a mean size of 5.97. Data consisted of the cumulative scores from a series of six individual and group tests. Subjects first completed each test individually and turned in their answer sheets. The groups then met and completed the same test using a new answer sheet. Individual and group scores counted toward the course grade.

The primary focus of this study was the comparison of groups' cumulative scores over the six exams with the comparable scores of their average and most knowledgeable member. The mean group score was 21.2% above the average individual score and 8.8% above the best individual score. All 222 groups outperformed their average member and 215 groups outperformed their best member.

The Watson et al. [1991] study was a follow-up to Michaelsen and examined the extent to which increased experience working as a group would affect the ability of the group to perform synergistically. Data were gathered from 272 graduate students enrolled in an organizational behavior course. Using a methodology similar to that in Michaelsen, 50 heterogeneous groups were formed, each consisting of five or six members. Instruction included experiential exercises, individual exams, group exams, and group projects. Group work accounted for more than half of students' individual course grades. Analysis was based on three tests in three separate time periods which were given individually and as a group. The study found significant group added value and increased synergy ratio scores in all three time periods. Synergy ratio scores increased from .21 at Time 1, to .27 at Time 2, to .35 at Time 3. The synergy ratio score was a measure of the extent to which group input could compensate for deficiencies in the knowledge of the best member [Watson et al. 1991]. These results support the propositions that decision-making effectiveness increases over time and that newly formed groups are significantly different from groups in which members have worked together for some time.

## 3. Studying Group Dynamics in Software Inspections

An inspection meeting may be conceptualized as a series of group decisions regarding the correctness of each functional unit of the development product (e.g., requirements, design, code) culminating in a decision to accept the unit, accept it subject to minor revisions, or require another inspection. Decision issues fall on a continuum from purely *judgmental* (i.e., those for which there is no correct answer) to purely *intellective* (i.e., those which have a demonstrably correct answer). Groups generally decide judgmental issues using some form of consensus; groups will generally produce the correct answer to an intellective issue *if a member of the group proposes it*. Software correctness decisions are almost always intellective, and therefore any synergistic effects arising from a group meeting are due to correct answers (i.e., identified defects) proposed during the meeting that were not identified previously. Group behavior research indicates that new correct answers should be the result of group learning. Factors generally associated with group learning include the following:

1. Diversity of Knowledge - Diverse knowledge among the members permits them to learn from each other in the meeting. If all the members bring similar knowledge and viewpoints to the group, it is unlikely that group learning will occur.

1. Facilitative Leadership - Leadership which facilitates the free exchange of ideas is critical; authoritarian leadership styles generally discourages creativity and learning.

1. Sufficient Resources (e.g., meeting time) - McGrath [1990] writes that the "evidence also suggests that under such time pressure groups will eliminate much of the communicative activity by which they evaluate one another's task ideas."

1. Sufficient Group Duration - Watson et al. [1991] find "that decision-making effectiveness continued to improve over time [which] suggests that newly formed groups are quite different from groups in which members have worked together for some period of time."

## 4. Research Approach

An initial goal of our research is to develop a description of software inspections in terms of group behavior theory. As a test of that framework, one or more laboratory experiments are proposed to test the effects of group duration on synergy. Students from analysis and design classes will be divided randomly into groups of five. In each group, a moderator, reader, and "author" will be randomly selected and trained for each role; the remaining two students in each group will be trained as independent inspectors. A series of problems with known errors will be presented over a period of several weeks for the groups to inspect. Half of the groups will remain intact throughout the experiment; members of the other groups will be shuffled so as to create a new group for each inspection. Based on Factor #4 above, the intact groups should show significantly greater synergy than those which are constantly being reshuffled.

Development of a group behavior theory-based framework for inspections would represent a major theoretical contribution. From a practitioner perspective, three potential outcomes with regard to the synergy analysis would be of interest:

1. If synergy levels are generally low and it is not possible or cost-effective to raise them, the use of inspection meetings should be reduced due to meeting costs.

1. If synergy levels are generally low but can reasonably be raised to cost-effective levels (e.g., via training or increased group duration), the effectiveness of inspections would be increased significantly.

1. If synergy levels are generally adequate, it still may be possible to fine-tune them for optimal results. Also, one of the rationales given for not employing inspections would be removed.

## 5. References

Fagan, M. "Advances in Software Inspections," *IEEE Transactions in Software Engineering*, XII:7, July 1986, pp. 744-751.

Hackman, J. and Morris, C. "Group Tasks, Group Interaction Process, and Group Performance Effectiveness: A Review and Proposed Integration," in L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Academic Press, 1976, pp. 1-99.

McGrath, J. "Time Matters in Groups" in *Intellectual Teamwork*, Lawrence Erlbaum Associates, Publishers, 1990, pp. 42-43.

Michaelsen, L., Watson, W., and Black, R. "Realistic Test of Individual versus Group Decision Making," *Journal of Applied Psychology*, 74:5, 1989, pp. 834-839.

Votta, L. "Does Every Inspection Need a Meeting?", SIGSOFT '93 - Proc. 1st ACM SIGSOFT Symp. Software Development Eng., ACM Press, New York, NY, 1993, pp. 107-114.

Watson, W., Michaelsen, L., and Sharp, W. "Member Competence, Group Interaction, and Group Decision Making: A Longitudinal Study," *Journal of Applied Psychology*, 76:6, December 1991, p. 803-809.

Wheeler, D., Brykczynski, B., and Meeson, R. *Software Inspection: An Industry Best Practice*, IEEE Computer Society Press, Los Alamitos, CA, 1996.