Understanding Fit and Appropriation Effects in Group Support Systems via Meta-Analysis

Alan R. Dennis
Accounting and Information Systems Department
Kelley School of Business
Indiana University
Bloomington IN 47405
ardennis@indiana.edu

Barbara Haley Wixom
McIntire School of Commerce
University of Virginia
Charlottesville VA 22903

Robert J. Vandenberg
Department of Management
Terry College of Business
University of Georgia
Athens GA 30602

October 2, 2000

Acknowledgment. We would like to thank Frank Schmidt and Jose Cortina for their helpful assistance in designing the meta-analysis procedure, and Bob Bostrom, Rick Watson, Dale Goodhue, George Marakas and the associate editor and the reviewers for helpful comments and insights on previous drafts. An early version of this paper was presented at the International Conference on Information Systems, Cleveland, Ohio 1996.
Understanding Fit and Appropriation Effects in Group Support Systems via Meta Analysis

Abstract

Many previous papers have lamented the fact that the findings of past GSS research have been inconsistent. This paper develops a new model for interpreting GSS effects on performance (a Fit-Appropriation Model), which argues that GSS performance is affected by two factors. The first is the fit between the task and the GSS structures selected for use (i.e., communication support and information processing support). The second is the appropriation support the group receives in the form of training, facilitation, and software restrictiveness to help them effectively incorporate the selected GSS structures into their meeting process. A meta-analysis using this model to organize and classify past research found that when used appropriately (i.e., there is a fit between the GSS structures and the task and the group receives appropriation support), GSS use increased the number of ideas generated, took less time, and led to more satisfied participants than if the group worked without the GSS. Fitting the GSS to the task had the most impact on outcome effectiveness (decision quality and ideas), while appropriation support had the most impact on the process (time required and process satisfaction). We conclude that when using this theoretical lens, the results of GSS research do not appear inconsistent.

Key words: group support systems, GSS, groupware, meta-analysis, task technology fit, appropriation, adoption, structuration, meta-analysis, contingency theory

Categories: HA11, AI0114
Introduction

The effectiveness of group support systems (GSS) has been the subject of many published research studies over the past 15 years. Unfortunately, when taken at face value, research findings about GSS performance have been relatively inconsistent across studies (Benbasat et al. 1993; Bui and Sivasankaran 1990; Dennis and Gallupe 1993; Fjermestad and Hiltz 1999; Kline and McGrath 1999). When compared to manual support, GSS use has been shown to improve decision quality (e.g., Chidambaram and Jones 1993) and reduce it (Straus and McGrath 1994), and increase the number of ideas (e.g., Gallupe et al. 1992), and decrease them (e.g., Straus and McGrath 1994). GSS has helped groups make decisions faster (e.g., Dennis, et al., 1999b) and take longer (e.g., George, et al. 1990), and be more satisfied (e.g., Anson, Bostrom, and Wynn 1995), and less satisfied (e.g., Gallupe et al. 1992) than groups working manually.

Meta-analysis has often been used to better understand and interpret the results of putatively conflicting results because in many cases, what at first appears to be conflicting results are in fact are not conflicting when analyzed with the use of meta-analytic statistics (see Hunter and Schmidt, 1990, pp. 23-31). They conclude that "'Conflicting results in the literature' may be entirely artifactual" (p. 29); that is, due to underlying statistical variation, not true conflicts.

Unfortunately, conclusions from two meta-analyses done in the early 1990s have been equally equivocal. McLeod (1992) and Benbasat and Lim (1993) concluded that in general, GSS use slightly improved decision quality and the number of alternatives or ideas generated, but appeared to increase the time taken to complete the task and to reduce participant satisfaction. However, they also noted that effects were not consistent and suggested that there were likely many moderators that influenced the effects of GSS use on performance. Although they identified moderators (e.g., group size, group history, type of GSS), they did not propose an overall framework that offered reasons for the moderator effects and guidance in how GSS
should thus be applied. These inconsistencies make it challenging for researchers to feel confident about the overall direction of GSS research and to know how to build upon existing research. Practitioners similarly are affected by inconsistent findings because they have less confidence in the way research findings can be applied within real-world settings.

Since the early 1990s, several theories have emerged that have the potential to offer new insights into the effects of GSS use. Two major schools of thought have dominated the study of information technology in general and GSS in particular over the last decade: the decision theorist school and the institutionalist school (DeSanctis and Poole 1994). The decision theorist school adopts a rational model and is rooted in cognitive and social psychology. Much recent work in the decision theorist school has centered on the development of new task-technology fit models that attempt to better explain and predict performance by examining the fit between the needs of the task and capabilities that GSS offer (e.g., Dennis and Valacich 1999; Zigurs and Buckland 1998). In contrast, the institutionalist school sees technology as an opportunity for change -- not as a causal agent of change -- and focuses on the social evolution that occurs as technology is adopted into social organizations. Recent research in the institutionalist school has focused on understanding how groups adopt (or "appropriate") GSS into their work processes and the factors that influence the appropriation process (e.g., DeSanctis and Poole 1994; Wheeler and Valacich 1996). We believe that both schools of thought hold promise for GSS research and thus the key lies in integrating them, not holding them separate.

The purpose of this study is to conduct a meta-analysis to summarize and synthesize the results of the past 15 years of research. To accomplish this, we first integrate recent work from both the decision theorist and institutionalist schools into one integrated framework that we call the Fit-Appropriation Model (FAM). We then use this model to organize previous research to better understand and interpret GSS effects on performance by examining both the fit between
GSS capabilities and tasks, and the appropriation of these capabilities by groups.

**Previous Theory and Research**

In this section, we begin by first reviewing the capabilities that GSS may provide (called "structures" in the institutionalist school). Then we present two recent theoretical advances from the decision theorist school that examine how the fit between GSS structures and task characteristics may affect performance. Next, we turn our attention to two theories from the institutionalist school that examine the appropriation processes by which groups adopt GSS into their work processes. We then integrate these two sets of theories into our Fit-Appropriation Model, from which we develop four hypotheses about the effects of fit and appropriation on performance.

**GSS Structures**

GSS are software tools that have the potential to change the nature of information exchange and decision processes during group discussion (DeSanctis and Gallupe 1987; Nunamaker et al. 1991). There are many different kinds of GSS available; nonetheless, it is possible to categorize the tools by looking at the way in which they are designed to enable or encourage groups to interact in a certain manner -- that is, the social *structures* provided by the GSS that influence how the group chooses to work (DeSanctis and Poole 1994). *Structural features* are the specific capabilities that the GSS offers (DeSanctis and Poole 1994).

Three kinds of social structures provided by a GSS are particularly important and have been defined and named differently as our understanding of them has evolved (Zigurs and Buckland 1998) (see Table 1). The first is communication support, which includes aspects of the GSS that support or enhance communication among participants, such as parallelism (both input and display) and anonymity (Zigurs and Buckland 1998). Parallelism is the ability of group members to simultaneously enter information; no participant needs to wait for others to finish
before contributing information. The need to wait to speak (termed production blocking) has been identified as a major cause of poor performance in brainstorming groups (Diehl and Stroebe 1987; Lamm and Trommsdorff 1973) and may also be cause of poor information sharing in decision-making groups (Dennis 1996b; Harkins and Petty 1987). Parallelism mitigates production blocking, resulting in process gains (Gallupe et al. 1994; Valacich et al. 1994). However, it may also result in information overload, or reduce the attention to important pieces of information (Dennis 1996b; McLeod et al. 1997).

Anonymity enables group members to contribute comments without being identified, which may increase the motivation to participate. Without anonymity, individuals, particularly low status participants, may withhold ideas due to negative evaluation apprehension (Diehl and Stroebe 1987; Lamm and Trommsdorff 1973) or may feel pressured to conform to the group majority or senior participants' views (Hackman and Kaplan 1974). Anonymity reduces the reluctance to contribute information, because it shields the contributor from group reaction and deindividuates the interaction (Connolly et al., 1990). Of course, the provision of anonymity does not guarantee that anonymity can dampen status effects (Weisband et al. 1995) or that other group members will not be able to guess the author of a specific comment¹. Anonymity also may impair some members' influence (McLeod et al. 1997) and increase free riding.

The second important structure is information processing support (Zigurs and Buckland 1998). Features of information processing support include ways to evaluate, gather, and aggregate information (e.g., voting), as well as ways to organize and analyze information (e.g., modeling) (Zigurs and Buckland 1998). These affect processes and outcomes by helping groups

¹ However, evidence suggests that even among groups whose members know each other, anonymity does protect the author's identity for most comments (Hayne and Rice 1997).
to better understand the issues under discussion. They may show levels of agreement about facts and criteria, and can integrate differing opinions (DeSanctis and Gallupe 1987; Sambamurthy and Poole 1992). Information processing support may improve performance by reducing losses due to incomplete task analysis or by increasing gains due to synergy, encouraging more information to be shared, and promoting more objective evaluation (Nunamaker et al. 1991).

The third important structure is process structure (Nunamaker et al. 1991; Zigurs and Buckland 1998). Process structure defines the process by which the group works on the task to help the group resolve the task in the most effective and efficient way(s) possible. Process structure means developing a task performance agenda for the group to follow in performing the task and "enforcing" the agenda -- that is, ensuring the group follows the agenda (Zigurs and Buckland 1998). In established groups that have used GSS tools for long periods of time, process structure can arise from the development of group norms or habitual routines that evolve gradually from use and act to constrain the way in which the group thinks about and uses the GSS (DeSanctis and Poole 1994). For groups with little practical experience with GSS that lack established norms or routines for GSS use, process structure can be delivered using training that induces the group to follow a planned agenda, restrictive software that inhibits the use of some features, and facilitation (often provided by a person external to the group) that helps group members follow the agenda (Wheeler and Valacich 1996). Process structure makes it more likely that the group will use the GSS as intended by the designer (Wheeler and Valacich 1996).

**Understanding Fit**

Much of the recent theoretical work in the decision theorist school has been what one might call contingency theories. The development and use of contingency theories has been criticized in recent years, primarily for two reasons (Iivari, 1992). First, many contingency theories suffered from the lack of a solid theoretical basis; they were opportunistic and based on
empirical data, rather than sound theory, and their predictions failed to be supported by new data. We have attempted to address this in our model by building on sound theoretical work. Second, the "naïve" deterministic nature of many contingency theories proved unfounded; people simply do not behave in a deterministic manner. We have attempted to address this by incorporating the institutionalist school into our model. Of course, any time one develops hypotheses, one can be accused of being deterministic. In short, while acknowledging that no model is ever ideal, we have attempted not to repeat the error of the early contingency theorists. Like Iivari (1992), we are supportive of the use of contingency theories.

Task-technology fit theories are contingency theories that argue that use of a technology may result in different outcomes depending upon its configuration and the task for which it is used (Goodhue and Thompson 1995). Fit can be conceptualized in many different ways (Iivari, 1992; Venkatraman 1989). Much recent work in the decision theorist school has used Venkatraman's "fit as profile deviation" approach (similar to Iivari's systems approach), in which an ideal profile of GSS capabilities is defined for a set of different task profiles (Zigurs and Buckland 1998). When a GSS is configured to match the ideal profile for a task, then a fit exists between the GSS and the task, and performance should improve. Conversely, performance should be reduced when the GSS configuration does not match the ideal profile for the given task. The first step in using this "fit as profile deviation" approach is to develop the set of ideal profiles for different tasks. For this, we turn to two recent theories in the decision theorist school.

Media Synchronicity Theory (MST) focuses on the fit between the communication support capabilities of the GSS and the needs of the task (Dennis and Valacich 1999). MST argues that pure electronic communication is most suited for tasks in which group members are encouraged to convey information (i.e., generate ideas/alternatives), and possibly explore many divergent
possibilities (i.e., what McGrath (1984) would term "generation tasks") because it enables parallel communication and does not require high feedback (Dennis and Valacich 1999).

In contrast, verbal communication is best suited for tasks in which group members need to converge on and agree to the meaning of information or a course of action because it provides rapid feedback (Dennis and Valacich 1999). It is important to note that MST argues that most decision-making tasks require both the conveyance of information (e.g., facts about the situation, discussion of alternatives) and the convergence on a course of action. For example, a group may generate and discuss a series of alternatives electronically, but verbal communication may be needed to converge on a final decision. Thus a combination of both electronic and verbal communication best fits the needs of decision-making tasks (Dennis and Valacich 1999).

Task-Technology Fit Theory (TTFT) argues that it is also important to fit information processing support features to the task. For "simple" or single goal tasks such as idea generation (similar to McGrath's (1984) "generate tasks"), no information processing support is needed (Zigurs and Buckland 1998). Such generation tasks are additive in that the outputs of individual participants are aggregated to form the group output. Members are encouraged to diverge, or contribute a wide variety of ideas, and they need not select among ideas nor come to consensus on a shared understanding. As such, information processing support adds little value. However, for decision-making tasks (single or multi-criteria intellective tasks, judgement tasks with no "correct" answer, and "fuzzy tasks" such as sense-making) information processing support is important (Zigurs and Buckland 1998). Decision-making tasks require participants to develop a shared understanding of criteria and alternatives, and reach a consensus on which alternative(s) is best, activities for which information processing structures can add value.

In summary, we have defined fit in terms of two "ideal GSS profiles." For generation tasks, in which the goal is to produce many ideas or alternatives, a task-technology fit exists
when electronic communication is used. For decision-making tasks, in which the goal is to select among defined alternatives, a task-technology fit exists when a combination of electronic and verbal communication is used and information processing support is provided. For tasks that include both the generation of ideas/alternatives as well as a decision-making choice (for which the goal is to produce a set of alternatives and select among them) a task-technology fit exists when electronic communication is used for generation sub-task(s) and when verbal communication with information processing support are used for choice sub-task(s).

**Understanding Appropriation**

The institutionalist school argues that people appropriate technology in different ways; therefore, how people use a technology is at least as important as its fit with the task. Todd and Benbasat (2000) argue -- in the context of individual decision support systems -- that task-technology fit influences the way in which individuals and groups use the technology, and it is this use that affects performance, not the fit itself. A GSS is inherently a social technology, so the way in which a group chooses to use a GSS is affected not just by the task-technology fit, but also by the fit of the technology with the group's habitual routines -- the social structures that evolve slowly over time (DeSanctis and Poole 1994; Gersick and Hackman, 1990). Because GSS is a less widely used technology (compared to e-mail, for example), very little research has examined groups who have used GSS long enough to have developed habitual routines or stable social structures around the use of the GSS. Therefore, most of the GSS research to date in the institutionalist school has instead examined the appropriation process -- the process by which groups incorporate the new social structures offered by the GSS's communication support, information processing, and process structure capabilities into their work processes. We consider two recent theories in the institutionalist school that examine this appropriation process.

Adaptive Structuration Theory (AST) argues that when a GSS is first introduced into a
group, the group goes through an appropriation process by which members examine the structures and agree on when and how to use them, and eventually their existing norms and routines adapt to include some, all, or none of the new structures (DeSanctis and Poole 1994). This appropriation can be "faithful," in that group members adopt the GSS structures in the way that the designer intended the structures to be used, or it can be "unfaithful," in that group members use the structures in ways not intended (DeSanctis and Poole 1994, p. 130)\(^2\). A group's ultimate performance depends upon what GSS capabilities are appropriated by the group, and whether the appropriation is faithful or unfaithful. As groups appropriate GSS capabilities they begin to learn what appropriations are successful and what are not, and this leads to the repetition of some appropriations from which habitual norms for GSS use gradually emerge.

Process Restrictive Adaptive Structuration Theory (PRAST) (Wheeler and Valacich, 1996) argues that the faithfulness of appropriation is affected by process structure. In its simplest form, process structure can be a meeting agenda that specifies how the group should use the GSS to perform the task (Zigurs and Buckland 1998). The guidance provided by an agenda is passive; an agenda cannot intervene to affect a group's appropriation (Wheeler and Valacich 1996). In this section, we focus on three more active approaches to providing process support. The first is facilitation, whether provided by a group member (e.g., group leader) or an individual external to the group, which provides a direct intervention into the group's meeting process to dissuade unfaithful appropriation acts and encourage faithful appropriation. A facilitator can have a

\(^2\) In their original papers, DeSanctis and Poole used faithful vs. ironic, but ultimately changed this in the final version of AST to "faithful" vs. "unfaithful" (see DeSanctis and Poole, 1994). DeSanctis and Poole use faithful and unfaithful relative to the "spirit and structural feature design" of the GSS (p. 130). That is, relative to what the GSS software designer built. In PRAST, Wheeler and Valacich (1996, p. 435) revise "faithful appropriation" to focus not on the software designers but instead on the meeting designers who develop agendas (or "heuristics") for the ideal way in which the GSS structures should be applied to a given task. We should note that for much of the research behind AST, the same individuals were often both software designers and meeting designers, and thus the distinction may have been less obvious at that point. We believe that considering faithful and unfaithful appropriation relative to a meeting design is more sound from both a practical and a theoretical standpoint.
significant intentional (and unintentional) impact on the group's meeting process (Griffith et al. 1998). A second process structure factor, restrictiveness, refers to the extent to which the GSS constrains individual behavior and makes it more difficult for individuals to adopt the structures in ways unfaithful to the meeting agenda (Silver 1990). Highly restrictive software (e.g., the GroupSystems Electronic Brainstorming) increases the probability that the GSS will be faithfully appropriated. A third factor, which we term appropriation training\(^3\), educates members about the structures and when to use them. It moves the group more quickly down the gradual path to the stabilization of appropriate structures and habitual routines. Wheeler and Valacich (1996) found that as the amount of facilitation, restrictiveness, and training increased, so did the extent to which groups faithfully appropriated the GSS in the way intended by the meeting designer.

In summary, when a group new to the use of a GSS for a given task is first faced with the GSS, it must decide how to appropriate it. The way in which the meeting process is structured (in the form of a meeting design and its fit with the task, coupled with facilitation, restrictiveness, and/or appropriation training) affects the likelihood that the group will adopt the GSS structures as intended. Once a group has appropriated the GSS structures and has experience using them to perform a certain task, norms evolve, and process structure becomes less important.

**Towards a Fit-Appropriation Model**

Three social structures form the core capabilities offered by GSS (communication support, information processing support, process structure) but act to affect the processes and outcomes of GSS use in two very different ways. Communication support and information

---

\(^3\) It is important to note that appropriation training means training in the appropriation of the structures; that is, when to use the structure, not just technical training on how to use the structures. (A good analogy may be the difference between understanding how to use a hammer to drive in nails (technical training) versus knowing how many nails to put where (appropriation training)). In general, appropriation training requires a shared training experience in which group members work together and actually use the structures in which they are being trained on a task similar to what they are expected to perform (Wheeler and Valacich 1996).
processing support are similar in that they offer ways by which groups can choose to communicate and to process information. They are best thought of as enabling structures. The choices of these two structures affect the "fit" between a GSS and a task. It is possible to assess the extent to which a given GSS provides these two structures by examining the individual software tools they contain (Zigurs et al. 1999).

In contrast, process structure acts to constrain group behavior by guiding the group to adopt the communication support and information processing structures in the way intended by the designer (Wheeler and Valacich 1996). Process structure affects appropriation and thus is best conceptualized as appropriation support, not as just as structure for the meeting process. Therefore, for the rest of this article, we will replace the term process structure with the term appropriation support. Appropriation support can be provided both by the restrictiveness of the GSS itself, as well as by factors outside the GSS (such as training, facilitation, prior group appropriation experiences, and group and organizational policies and norms). Thus, we argue that appropriation support cannot be assessed just by examining the GSS alone.

By integrating the task-technology fit theories (TTFT: Zigurs and Buckland 1998; MST: Dennis and Valacich 1999) with the appropriation theories (AST: DeSanctis and Poole 1994; PRAST: Wheeler and Valacich 1996), we are able to better understand how GSS use is likely to affect group performance than if we were to consider the theories individually (see Figure 1). Communication support and information processing structures offered by the GSS and selected for use by the meeting designer(s) (which may be the group as a whole, the group leader, and/or a facilitator) either fit or do not fit the needs of the task. If the selected structures do fit the task, performance will be improved if the group appropriates the structures faithfully. For experienced GSS groups with well-established norms and habitual routines, those norms may be sufficient; however, if the group lacks established GSS use norms and routines, external appropriation
support may be needed to help the group appropriately adopt the structures. A good fit without the needed appropriation support is less likely to lead to improved performance. Conversely, a poor fit with faithful appropriation is similar to fitting a square peg into a round hole. Thus, although faithful appropriation is desired in situations of good fit, it may be unwise to force groups to use a GSS faithfully if the communication support and information processing support structures do not fit the task at hand.

-- Insert Figure 1 about here --

In summary, given the inconsistencies in past research, we developed a Fit-Appropriation Model (FAM) to help us organize past research to make our meta-analyses more precise. FAM argues that a task-technology fit is a necessary but not sufficient condition to improve performance. Without proper appropriation support (e.g., prior experience, habitual routines, facilitation, restrictive software, appropriation training), performance is less likely to improve even if there is a task-technology fit. That is, task-technology fit affects performance, as mediated by the appropriation process. Because we are using secondary data and are unable to include appropriation, we test the relationship between fit and performance, mediated by the provision of appropriation support. See Figure 1.

Hypotheses

To develop hypotheses, we must now define performance, which has been used in different ways by different researchers. We follow the approach of Drazin and Van de Ven (1985), Benbasat and Lim (1993), and Dennis and Kinney (1998), who defined performance in terms of three major factors: (i) effectiveness as defined by decision quality and/or number of ideas generated, (ii) efficiency as defined by the time to complete the task, and (iii) participants' satisfaction with the process and/or outcomes. While no one conception of performance is
perfect, we believe this is a reasonable set of factors to triangulate on the performance construct.

FAM leads to one omnibus, end-to-end hypothesis and several hypotheses that examine individual relationships within the model (see Table 2). First, a task-technology fit is necessary, but not sufficient for good performance; without support in appropriating the GSS (assuming the group has no habitual routines for using the GSS for the task\(^4\)) performance is less likely to improve. Therefore, we hypothesize that performance will be improved when there is a task-technology fit and when appropriation support is provided compared to situations in which there is no task-technology fit with or without appropriation support (i.e., cell I versus cells III and IV in Table 2):

\[ H1: \text{When there is a task-technology fit and appropriation support is provided when needed, a) decision quality will be improved, b) more ideas/alternatives will be produced, c) less time will be required to complete the task, d) participants will be more satisfied with the outcome, and e) participants will be more satisfied with the meeting process.} \]

We argued above that appropriation support would increase the chance that the group will appropriate the technology as intended. Thus, when there is a task-technology fit, performance should improve when appropriation support is provided (i.e., cell I versus cell II in Table 2):

\[ H2: \text{When there is a task-technology fit, the provision of appropriation support should result in a) improved decision quality, b) more ideas/alternatives, c) less time required to complete the task, d) improved participant satisfaction with the outcome, and e) improved participant satisfaction with the meeting process.} \]

We have argued above that task-technology fit -- even without appropriation support -- should improve performance relative to situations in which there is no task-technology fit,

---

4. While FAM considers the effects of habitual routines and examines faithful and unfaithful appropriation, our use of meta-analysis precludes the inclusion of these constructs in our tests because virtually no studies examined them.
whether appropriation support is provided to the non-fit conditions or not. Because appropriation support should improve the likelihood that the GSS is appropriated as intended, providing appropriation support in situations of bad fit should have little positive impact on performance (groups will simply use a less effective tool more reliably). In fact, stronger pressure to use a less effective tool may even impair performance relative to those groups that have less appropriation support. This leads to two hypotheses, that fit alone without appropriation support should improve performance relative to no-fit situations (cell II versus cells III and IV in Table 2), and that in no-fit situations, appropriation support has no impact on performance (i.e., cell III versus cell IV in Table 2).

**H3:** The presence of a task-technology fit without appropriation support should result in a) improved decision quality, b) more ideas/alternatives, c) less time required to complete the task, d) improved participant satisfaction with the outcome, and e) improved participant satisfaction with the meeting process compared to situations with no fit.

**H4:** When there is no task-technology fit, the provision of appropriation support should result in no change in a) decision quality, b) the number of ideas/alternatives, c) time required to complete the task, d) participant satisfaction with the outcome, and e) participant satisfaction with the meeting process.

**Method**

**Selection of Studies**

To locate studies, we performed computer searches on 11 databases, did manual searches through likely MIS, psychology and management journals and conference proceedings, read previous literature reviews, and posted messages soliciting studies from the subscribers of the GSS-L listserv. We focused only on published refereed papers, including journal articles and

---

conference papers but omitting dissertations and working papers. We also omitted studies of e-mail and "chat" software. This resulted in over 250 studies published between 1980 and 1999, inclusive. We selected all papers that reported the results of a test of a treatment group (GSS use) to a control group (interaction without a GSS) and provided means and standard deviations (or other statistics) for any of the five dependent variables (decision quality, number of ideas/alternatives, time, satisfaction with the process, or satisfaction with the outcome).

This resulted in 61 articles (indicated by an * in the reference list). Some useful and often cited articles are not included in our analyses because they did not report any of the five dependent variables (e.g., Watson et al. 1988; Zigurs et al. 1988) or because they offered no direct comparison of GSS to no-GSS groups (e.g., Connoly et al. 1990). About half of the studies crossed the GSS and non-GSS condition with levels of some other independent variable. Studies containing several such comparisons were disaggregated and treated as separate data points (Hunter and Schmidt 1990). For example, Gallupe, et al. (1992) reported the comparison between GSS and non-GSS groups for five levels of group size. Since individuals were randomly assigned to conditions, we treated the five levels of group size as independent studies for the analyses. This procedure resulted in 119 usable data points.

Four field studies were included in the meta-analysis because they compared a treatment group to a manual control group. Some researchers have argued that field studies should be excluded from meta-analyses in which the bulk of studies are lab experiments either because this is a case of trying to compare apples and oranges, or because the field studies may be weaker

6. The exclusion of unpublished studies and dissertations may bias the analyses because published studies tend to have significant, hypothesis-supporting results (McLeod 1992). We address this concern later in the paper by reporting the fail-safe-n, which predicts how many unpublished studies with nonsignificant effect sizes would be required to reduce the observed average effect size to 0.

7. We included only those field studies that had a comparison non-GSS group as a "control group": Dean et al. 1994; Dennis et al. 1999b; Jarvenpaa et al. 1988; Olson et al. 1993. While there are numerous well-done field studies,
methodologically. Hunter and Schmidt (1990, pp. 495-498) provide compelling arguments against this. First, even though the effect sizes may differ in magnitude between lab and field studies, they should be consistent in direction if there is any validity to the theoretical arguments (Cook and Campbell 1979); the underlying theoretical processes are the same so this is not a cases of apples and oranges. Second, judging methodological weakness is difficult (Hunter and Schmidt argue it requires six judges to achieve the reliability required for meta-analysis) and errors from improperly eliminating studies are at least as severe as errors from including "weak" studies in the analysis. Nonetheless, we follow Hunter and Schmidt's advice and conduct the same analyses without the field studies (see the discussion).

**Dependent Measures**

The final set of studies included one or more of the five dependent variables described above. The first dependent variable, decision quality, was defined by most researchers as the correctness or "goodness" of the group's choice, although in some cases it was measured as a zero-one variable (correct or incorrect choice) or the distance from the optimal solution. The number of ideas/alternatives was operationalized as the number of unique ideas (i.e., nonredundant) generated for idea generation tasks, or the number of nonredundant decision alternatives generated for decision-making tasks. The third dependent variable, efficiency, was defined as the time to complete the task. The final two variables, satisfaction with the process and satisfaction with the outcome, were measured via post-session questionnaires. Several studies reported one overall measure of satisfaction rather than separate measures of process satisfaction and outcome satisfaction; rather than attempting to categorize the overall measure as one or the other, we choose instead not to include it. As an aside, we note that there was some

---

only those four studies included comparison groups.
natural partitioning in the studies by task; studies that were purely idea generation did not have
decision quality measures or time measures, and most decision making studies did not include
idea generation measures.

Analytic Procedure

To test the hypotheses, we used the three-step partitioning comparison procedure developed
by Hunter and Schmidt (1990)\(^8\). With this procedure, the studies are first partitioned into two or
more sets. Second, a meta-analysis is performed separately on the studies within each set to
produce a set of statistics for each set. Third, t-tests are used to compare the statistics among the
sets to see if there are significant differences among them.

**Partitioning the studies.** We first partitioned the data into four sets: (i) one in which there
was a fit between the structures used and task, and appropriation support was provided. (ii) one
in which there was a fit between the structures used and task, but no appropriation support was
provided; (iii) one in which there was no fit between the GSS structures and task and
appropriation support was provided; and (iv) one in which there was no fit between the GSS
structures and task and no appropriation support was provided. See Table 2.

We determined if there was a fit using the "ideal profiles" developed in the summary
paragraph of the "Understanding Fit" section above (use of electronic communication for
generation tasks; use of a combination of electronic and verbal communication and information
processing support for decision-making tasks). One rater read the studies and coded the tasks

\(^8\) An alternative is to use the Hedges and Olkin (1985) multiple regression procedure to test for moderators.
However, with this approach, one is almost guaranteed that some moderators will be found statistically significant
when they are not. Hunter and Schmidt (1990, pp. 85-89) demonstrate that assuming some modest sampling error, a
multiple regression of 6 moderators that are not related to the dependent variable over a set of 40 studies will
produce a multiple R of .37, which will have a p-value less than .001, even though the true multiple R is .00 and
should not be found significant. Conversely, the true moderators are likely to account for only a moderate amount
of the variation in the dependent variable and thus sample sizes of only 20-40 will likely not provide sufficient
power to detect them in the face of modest sampling error. Therefore, we decided to use Hunter and Schmidt’s
partitioning approach.
used within each study as generation, decision-making, or a combination of idea/alternative generation and decision-making based upon the study’s description. The same rater then coded the way in which communication support was used as electronic communication only, verbal communication only (e.g., a chauffeured system), or a combination of electronic and verbal communication. Finally, the rater coded if information processing support was present (i.e., if groups actually used any level 2 structures as defined in Table 2 of DeSanctis and Gallupe (1987)). We contacted authors of four studies to clarify information in the study that was not clear to the coder. A second rater then independently performed the same procedure on 30 of the studies (approximately 50%). Three differences were found (which were resolved between the raters) giving an inter-rater reliability of .97 (3 differences / (30 studies * 3 codings per study). The results of the coding were further validated against a list of published GSS research design descriptions (Fjermestad and Hiltz 1999).

Those studies in which electronic communication was used for idea generation tasks were classified as a fit. Those studies in which a combination of electronic and verbal communication and level 2 structures were used for decision-making tasks were classified as a fit. If the task was a combination of generation task and a decision-making task, the processes used in both parts had to be a fit for the overall process to be classified as a fit. Studies for which there was no fit were classified as "no fit" (e.g., using verbal communication for idea generation). The studies were then examined to determine the extent to which appropriation support was provided. Appropriation support was coded as present if any of the three appropriation mediators defined

---

9. We also examined the studies to see if any groups had strong histories of using the GSS for the type of task under study and thus could be expected to have already appropriated the GSS and developed stable norms for use. However, this was not the case for any of the studies in the sample, although Hollingshead et al. (1993) comes the closest to a stable appropriation over a variety of different tasks. For this reason, we do not consider the effect of stable habitual norms in our analyses.
by Wheeler and Valacich (1996) were present (i.e., facilitation, restrictiveness, appropriation training). Facilitation was classified as present if a facilitator actively assisted the participants in appropriating the GSS into their process by directing the group in how to use the GSS (e.g., Anson, et al., 1995) as opposed to a passive technical facilitator (see Wheeler and Valacich 1996, p. 437). The software was considered restrictive only if it strongly constrained participants' behavior such that they could choose to use or not use the GSS, but had little control over which tools to use when or how those tools would be used (e.g., GroupSystems Electronic Brainstorming tool was considered restrictive because it controls what information participants have access to). Appropriation training was considered present if the group was trained on how to use the GSS and required to work as a group on a comparable task to the experimental task before performing the experimental task. Studies were then classified as having "appropriation support" or "no appropriation support." Inter-rater reliability over this set of codings was .93.

**Meta-analysis within sets.** The second step was to perform a separate meta-analysis within each of the three sets of studies. It is beyond the scope of this manuscript to describe in detail the statistical algorithm underlying a meta-analysis (see Hunter and Schmidt 1990). In short, the procedure computes a weighted average effect size for a given dependent variable across an entire set of studies. The first step is to calculate an effect size ($d$) for each dependent variable in each study. This was done using the equation presented by Cohen (1977). In brief, the mean value for the dependent variable for the non-GSS subjects is subtracted from the mean value for the GSS subjects, and this difference is divided by the pooled standard deviation from both conditions (see also Hedges and Olkin 1985; Hunter and Schmidt 1990). The net effect of this transformation is that the differences are now standardized to a common metric across all studies, and this in turn means that effect sizes may be statistically combined and evaluated. Positive effect sizes indicate that the mean impact of GSS use across all included studies was to
increase the measure (e.g., positive decision quality effect sizes mean improved quality, while positive time effect sizes mean longer times are required). Effect sizes can also easily be tested to see whether they are statistically significant from zero.

**T-tests.** Hypotheses H1, H2, and H3 were one-directional, so we used one-tailed t-tests. H4 is non-directional, so we used a two-tailed test.\(^\text{10}\)

### Results

**Overall Effects**

We began by calculating the overall effect sizes across all studies, without regard for our hypotheses (see Table 3). Across the entire set of studies, GSS use was associated with a statistically significantly greater number of ideas, more time to complete the task, and decreased satisfaction with the process, but had no effect on decision quality or satisfaction with the outcome. These overall values present a somewhat different pattern from the meta-analysis conducted seven years previously by McLeod (1992), who found GSS use to improve decision quality and by Benbasat and Lim (1993), who found GSS use to improve decision quality, and reduce outcome satisfaction, but have no effect on process satisfaction.

--- Insert Table 3 about here ---

Table 3 also presents the fail-safe-n, which is used to predict how many unpublished studies with nonsignificant effect sizes would be required to reduce the observed average effect size to 0. We reported this only for effect sizes that were statistically significant.

---

\(^{10}\) All results that were significant with a one-tailed test were also significant using a two-tailed test. We use a one-tailed test because it is theoretically appropriate for our hypotheses. Had we chosen to use a one-tailed test for H4, it would make the nonsignificant results for the number of ideas and time significant, in that providing appropriation support would result in fewer ideas and require more time. Because we have conducted multiple univariate tests (four hypotheses each with five t-tests for a total of 20 test) we must acknowledge that the family-wise type I error (the probability that at least one significant result is due to chance) is approximately .64 using the Bonferroni inequality.
The remaining two items in Table 3 are tests of homogeneity of variance. If homogeneity Q is statistically significant or if the percent explained is less than 75% of the variance for a given effect size, then there are important differences in variance among studies, likely due to moderating variables (Hunter and Schmidt 1990). Table 3 shows that there was significant heterogeneity of variance across all five dependent variables (Homogeneity Q was significant, and the percent explained was less than 75% for all five). Therefore, we conclude that there might be important moderating factors, such as those we hypothesized in the model.

**H1: Overall Results**

Table 4 presents the number of studies and mean effect sizes for each dependent variable within the different sets of studies. It also shows whether the mean effect size is significantly different from zero. H1 examines the overall, end-to-end test of situations in which there is a fit and appropriation support is provided (the first set of data in Table 4) to those in which there is no fit (the third set of data in the table).

--- Insert Table 4 about here ---

**Effectiveness.** There were no significant differences in decision quality (t(53)=1.14, p=ns). The mean effect sizes in both sets of studies were not significantly different from zero meaning that GSS groups produced decisions that were not different than non-GSS groups. There was, however, a significant difference in the number of ideas/alternatives generated (t(33)=2.33, p=.013). While the mean effect sizes in both sets were significantly greater than zero (meaning that with a fit and appropriation support, and without a fit, GSS groups produced more ideas than non-GSS groups), GSS groups working in situations with a fit and with appropriation support produced more ideas than groups working without a fit. H1a was not supported, but H1b was.

**Efficiency.** A t-test found that groups took less time when there was a fit with appropriation
support than when there was not (t(20)=4.77, p=.001). The mean effect size for the fit-with-appropriation-support groups is significantly less than zero, meaning that they took less time than non-GSS groups to complete the task. H1c was supported.

**Satisfaction.** There were no significant differences in outcome satisfaction (t(15)=0.22, p=.414), and the mean effect sizes in both sets of studies were not significantly different from zero. However, participants were more satisfied with the process when there was a fit with appropriation support than when there was no fit (t(28)=7.77, p=.001). A glance at the mean effect sizes within each of the three categories is again instructive. Although the mean effect size for the no-fit set is significantly negative (meaning that members of GSS groups were significantly less satisfied than non-GSS groups), the mean effect size for the fit-with-appropriation-support groups is significantly greater than zero, meaning that they were significantly more satisfied than non-GSS groups. H1d was not supported; H1e was.

**H2: Effect of Appropriation Support in the Presence of a Task-Technology Fit**

H2 argued that providing appropriation support in cases where there was a task-technology fit should improve performance. This set of analyses compares the first two sets of data in Table 4. The t-tests reported in this table show significant differences only for time (t(5)=5.03, p=.002) and process satisfaction (t(17)=8.44, p=.001). The provision of appropriation support resulted in less time and more satisfaction with the process, but had no effect on decision quality, number of ideas, or outcome satisfaction. H2c and H2e were supported; H1a, H2b and H2d were not.

**H3: Effect of a Task-Technology Fit**

H3 argued that providing task-technology fit (without appropriation support) should still improve performance relative to situations in which there was no fit. This set of analyses compares the second and third sets of data in Table 4. The t-tests reported in this table show that fit improved decision quality (t(53)=2.81, p=.003) and number of ideas (t(18)=2.13, p=.024), but
had no effect on time or satisfaction. H3a and H3b were supported; H3c, H3d and H3e were not.

**H4: Effect of Appropriation Support in the Absence of a Task-Technology Fit**

H4 argued that providing appropriation support in cases where there was not a task-technology fit should have no effect on performance. This set of analyses compares the last two sets of data at the bottom of Table 4. The t-tests reported in this table show that appropriation support improves process satisfaction ($t(18)=3.61, p=.001$) but has no other effects. H4a, H4b, H4c, and H4e were supported; H4d was not.

**Discussion**

Over time as the number of studies has grown, researchers have analyzed the findings of GSS studies in different ways. McLeod's (1992) first meta-analysis summarized the general overall effects of GSS, offering researchers and practitioners a general understanding of GSS and broad trends by which individual studies could be compared. Benbasat and Lim's (1993) meta-analysis extended this by including more studies and focusing on moderating factors. Fjermestad and Hiltz (1999) probed more deeply into the then much larger pool of studies in search of additional moderators. Over this same time period, non-statistical reviews of GSS identified inconsistencies between general tendencies and individual studies and proposed that mediating factors, such as the group, task, and the way in which the GSS was used could make a difference (Dennis and Gallupe 1993; Dennis et al. 1991; Kline and McGrath 1999).

At one level, our results in Table 3 should again summarize the research to date in terms of the general effects of GSS use. However, the fact that there is significant heterogeneity of variances across all five dependent variables suggests that there are important moderating variables that may cause pronounced departures from the overall mean. While the mean effects sizes in Table 3 do offer a *statistically appropriate* interpretation of the general body of research, the presence or absence of fit and appropriation support have such a significant effect on
performance as to make the general overall results potentially misleading.

Thus, this study’s principal contribution comes from using a Fit-Appropriation Model (FAM) to organize and investigate the effects of GSS use in different situations. FAM proposes a set of two overarching concepts (fit and appropriation) by which to design and evaluate research, and to design and implement organizational interventions. By considering both concepts, researchers may be better able to understand past research and design new and more precise future studies that offer fewer inconsistent results. Likewise, practitioners may be able to understand and develop boundary conditions and “rules” for GSS use in organizations.

By understanding GSS effects with and without fit and appropriation support, we are better able to understand and predict effects of GSS use. Our meta-analysis suggests that if GSS are used appropriately (with fit and appropriation support), then we would expect them to increase the number of ideas generated, reduce the amount of time required, increase participants' satisfaction with the meeting process, but have no effect on decision quality or outcome satisfaction. These conclusions are strikingly different from past interpretations of GSS research.

Our analyses included four field studies. There has been some concern that field study research has produced different conclusions than laboratory experiments (e.g., see Dennis et al. 1991). In general, field studies have found more positive effects attributed to GSS use than have laboratory experiments. While this may be due to the research methods used, Dennis et al. (1991) argue that the major differences are due instead to the groups and tasks studied, and the process by which the GSS was used; that is, that the differences are due to the situations studied and not to any inherent difference in research methods.

Thus, from a theoretical perspective, we would argue that the inclusion of the field studies should have no significant impact, beyond providing a larger sample of studies with a richer and more diverse set of situations in which to draw conclusions about GSS effects (see Hunter and
Schmidt 1990). Nonetheless, we repeated our analyses omitting the four field studies. There were no changes to the statistical conclusions for any of the hypotheses, except for H1a. H1a, which hypothesized that decision quality should improve with fit and appropriation support. Although this was not supported in our original analysis, it was supported by excluding the field studies \((t(48)=1.68, p=.050)\); that is, omitting the field studies, marginally improved decision quality in the fit with appropriation condition (by removing studies with lower decision quality), so that the differences became significant. However, it removed enough studies so that our test of decision time becomes problematic; although it remains statistically significant, there are too few studies for a reliable statistical test. We believe that this provides support for the retention of field studies.

Our analyses (H3) suggest some conclusions about the relative impacts of fit and appropriation support. Simply, ensuring a good fit between the GSS and the task (without providing appropriation support) improves effectiveness in terms of decision quality and the number of ideas produced. However, groups take longer (mean effect size 1.28) and may be less satisfied with the process (mean effect size \(-.50^{11}\)) than groups working without a GSS. We speculate that without appropriation support, GSS groups encounter appropriation challenges in revising their work processes to incorporate the GSS structures. This need to figure out how to appropriate the GSS means that they spend more time on the task and become more frustrated with the process than groups without the GSS who are able to use their well-established habitual routines for working in meetings\(^{12}\). Of course, such habitual routines may not be effective in

---

11. The mean effect size is not statistically significantly different from zero \((p=.051)\) but is virtually identical to the mean effect size in the no fit set \((-0.49)\) which is significantly different from zero. We believe, but cannot prove, that the lack of significance is due to the small sample size, not to a true lack of difference from zero.

12. We should note that student groups formed for use in laboratory experiments typically have not worked together before as a group, and thus may not have habitual group routines per se. However, habitual routines are structures that pervade organizations and cultures so that most members of the culture (i.e., students, in this case) are likely to
producing ideas or good decisions, but they are comfortable and quick to employ (Gersick and Hackman, 1990).

As an aside, we note that work by McGrath and his colleagues (Hollingshead et al. 1993; McGrath et al. 1993) provides some empirical evidence of initial performance differences due to the appropriation of GSS. They examined the use of GSS over a series of 10 experimental sessions, and found that performance of GSS groups improved only after they gained experience with the GSS. “Surprisingly, the relationship between technology and task performance appeared to be more dependent on experience with the technology and with group membership than on the type of task on which the group was working.” (Hollingshead et al. 1993, pp. 328-329). “A technology new to the participants appears to impede fast and effective performance on group tasks initially ... one that can be overcome with experience, but that may again become manifest when further changes are made in the group as a system.” (McGrath et al. 1993, p. 412).

Once there is a task-technology fit, the provision of appropriation support to aid the incorporation of the GSS into the work processes improves efficiency (reduces the time needed to perform the task) and improves participant's satisfaction with the process. This is true relative both to GSS groups working with a fit but without appropriation support, as well as relative to groups working without the GSS. We interpret these results as suggesting that while task-technology fit improves outcome effectiveness, appropriation support acts to improve the process by improving efficiency and process satisfaction.

One other interesting question is what happens when appropriation support is provided but there is no fit between the GSS and the task (H4). That is, when the meeting is designed to use GSS capabilities in way that does not fit the ideal profile for the task (e.g., using solely electronic have similar assumptions and structures and enact those structures with ease even if they have not had direct personal experience in working together before.
communication for a choice task) can appropriation support in the form of facilitation, training, and/or restrictiveness overcome the problems of a poor fit? FAM suggests that there should be few differences in performance, because helping a group to more effectively adopt a poor process should not improve performance. Our analyses found that appropriation support improves satisfaction with the process, but has no other effects. We conclude that providing appropriation support in cases of poor fit cannot improve a group's effectiveness or efficiency, but may reduce the dissatisfaction that members experience with the poor process. But, it is important to note that even with appropriation support, group members are not more satisfied than groups without the GSS (i.e., the mean effect size is not greater than zero). As an aside, we also should note that groups in the fit without appropriation support condition were less satisfied than those in the no fit with appropriation support condition (t(16)=2.06, p=.028) and more satisfied than those in the no fit without appropriation support condition (t(18)=1.81, p=.043).

If we compare the pattern of effects of providing appropriation support in situations where there is and is not a fit, we see an improvement of process satisfaction in both cases. Thus we conclude that in general appropriation support does help groups feel more satisfied with the meeting process, because they do not have to figure out the process for themselves. However, while appropriation support reduces time when there is a fit, it does not when there is not a fit, and the difference in means -- while not significant -- is in the opposite direction: more time with appropriation support than without. We speculate that the lack of appropriation support when groups are faced with using a GSS in a way that does not fit the task may better enable them to adopt the GSS unfaithfully and thus take less time than if they had used the GSS faithfully.

**Implications for Future Research**

The results of our meta-analysis have several implications for future research. This paper used a theoretical model to organize past research to better understand how GSS use affects
group performance. Our analysis using meta-analytic methods to interpret previous research suggests that FAM can be a useful model for understanding some of the factors that affect group performance and can help pinpoint the differential effects of task-technology fit and appropriation support. More research is obviously needed to refine the model and incorporate other factors such as those pertaining to the group and organization. Nonetheless, we believe it is an appropriate first step in developing more integrative models of group performance.

Another important issue for future research is investigating the extent to which FAM may apply to other research contexts. We believe that the integration of the concepts of task-technology fit and the appropriation of technology into the habitual routines of individuals, groups and organizations has the potential to improve our understanding of performance in other research areas. While the specific instantiation of FAM for this study will likely not extend beyond the bounds of GSS research, we believe that the general model (in which task-technology fit affects individual, group or organizational effectiveness, while appropriation support and habitual routines affect appropriation, which affects the efficiency by which the task is performed and the participants' satisfaction with the work processes) may be useful in a wide variety of areas, but this remains an empirical question.

What does FAM and the results of our analyses suggest as the next step for GSS research? From an empirical viewpoint, there is an obvious lack of published research in several areas. Table 4 contains cells in which the n values are very small (e.g., only two studies measure outcome satisfaction when both task technology fit and appropriation support exist). Very little research has examined time or satisfaction in situations where there was a fit between task and technology, either with or without appropriation support. While many researchers may believe that sufficient “simple” studies comparing the performance of groups using GSS to those without have been done, most of these studies have examined cases for which there was no fit. While we
have enough of these studies with no fit, we clearly need more under the arguably more
important conditions for understanding ongoing organizational use of GSS: fit with appropriation
support. We encourage researchers to investigate these holes in empirical research.

From a theoretical viewpoint, there are also holes that need to be filled. The meta-analysis
on the set of studies with fit and appropriation support give homogeneity of variance statistics
that may be useful in pinpointing areas in need of additional theoretical development. The
variance statistics for decision quality (homogeneity Q=4.51, p=.952, percent explained=100%)
and time (homogeneity Q=2.92, p=.233, percent explained=81%) suggest that the studies in this
set have homogeneous variance; that is, among studies with fit and appropriation support it is
unlikely that there are important moderating factors that greatly affect decision quality and time
in a systematic fashion. The opposite is true for the number of ideas (homogeneity Q=59.93,
p=.001, percent explained=36%), outcome satisfaction (homogeneity Q=5.98, p=.015, percent
explained=18%) and process satisfaction (homogeneity Q=24.20, p=.004, percent
explained=47%). That is, the variance statistics suggest that there likely are important and
systematic moderating factors that affect the number of ideas and satisfaction.

Given the relatively small sample sizes, we need to interpret these statistics cautiously. One
interpretation would suggest that the richest starting point for the development of new theories to
predict performance lies in understanding and explaining differences in idea generation
performance and participant satisfaction. The effect sizes for the number of ideas generated in
individual studies ranged from -.34 to 3.60, an incredible range. Recent research suggests that
even something as simple as the way in which the idea generation question is posed to the group
can affect performance by 40% or more (Dennis et al. 1999a). Clearly we need to better
understand the moderating factors that can so dramatically affect idea generation performance.

Participant satisfaction has important implications for the successful organizational adoption
of GSS (DeSanctis and Gallupe 1987). Our results suggest that process satisfaction is closely linked with appropriation support; when participants lack the support they need in integrating the GSS into their work processes -- perhaps becoming frustrated in learning to apply an unfamiliar technology -- process satisfaction suffers. Our results show few general patterns in outcome satisfaction, which remained almost unchanged from condition to condition or even in the presence or absence of GSS. Briggs and his colleagues argue that outcome satisfaction is affected by the extent to which the participants' individual vested interests have been accommodated (Briggs and de Vreede 1997; de Vreede et al. 2000). Laboratory experiments were the majority of the studies in our meta-analysis, and it is difficult to imagine situations in laboratory settings in which participants' individual vested interests can be accommodated; because tasks are artificial, participants have no vested interests. Thus we believe that our results showing few effects on outcome satisfaction offer some support for Briggs et al.'s line of reasoning. At the same time, this argument challenges the ability to generalize these findings to organizations in which vested interests are more common.

Our analyses suggest that the presence or absence of appropriation support has little impact on effectiveness, such as the number of ideas generated or decision quality. Thus in designing laboratory experiments (or generalizing results from them) we conclude that providing high levels of appropriation support is less important for understanding impacts on effectiveness. However, if the goal is to understand and/or generalize effects on time or satisfaction, appropriation support becomes important. If there is a lack of appropriation support, participants may require more time to fit the new GSS structures into their processes. We conclude that generalizing the time or satisfaction results of a study that does not provide appropriation support when it is needed to the on-going post-appropriation use of GSS is problematic.

Interestingly, one longstanding discrepancy between lab and field research has been the
issue of participant satisfaction and the time needed to complete the task. Lab research has often found GSS use to decrease satisfaction and increase the time required (Benbasat and Lim 1993; McLeod 1992), while field studies have often reported improved satisfaction and decreased time (Dennis et al. 1991). We conclude that these effects are not due to the GSS per se, but instead are due to the appropriation process. When appropriation support is provided (e.g., most field studies), the GSS is more easily incorporated into the group's habitual routines and the participants spend less time and are more satisfied.

Facilitation is one moderator in particular that needs to be better understood. Facilitation has received little attention in laboratory research (e.g., Anson et al. 1995; Griffith et al. 1998), but is an important part of GSS use in the field (e.g., Tyran, et al., 1992). Bostrom and colleagues have studied facilitator behaviors to better understand what facilitators do in GSS meetings and what behaviors are seen as being important (Bostrom et al. 1996; Clawson et al. 1993). In a study of 45 highly experienced GSS facilitators, they found that planning and designing meeting agendas was the single most important behavior, statistically significantly different from all other facilitator behaviors (Bostrom et al. 1996). The next most significantly important behaviors were a set of five behaviors: matching GSS tools to the task; adapting the meeting agenda as needed; clarifying meeting goals, agenda items, and terms; remaining focused on the outcome; and creating an open environment for participation. Other behaviors, such as managing the meeting, asking questions, building rapport, managing conflict, and creating comfort with the technology were ranked as being significantly less important.

There are many ways to interpret these rankings, but to us they suggest an interesting pattern. We conclude that the most important facilitator behaviors were those that focused on fitting the GSS capabilities to the task and helping participants to successfully appropriate the GSS into their meeting processes (e.g., identifying meeting goals, creating and changing the
agenda, selecting GSS tools, clarifying agenda items and terms). In contrast, traditional facilitation behaviors that focus on group dynamics such as building rapport, managing conflict, managing the meeting, and creating comfort were seen as being less important. If the primary contribution of a GSS facilitator is indeed providing a good task-technology fit and helping the appropriation process (not managing group dynamics), then the provision of training, standardized templates, or expert consultation (e.g., a help line) may provide many of the same benefits without facilitator to being present while the group works. Thus one important issue for future research is designing and testing GSS interventions that do not require outside facilitators.

The analyses suggest one general issue for future research: research design and reporting. We encourage researchers and editorial boards to adopt minimum reporting standards for the data included in published manuscripts. The American Psychological Association recommends as minimum reporting standards: (a) sample sizes; (b) reliability information on the instruments; (c) means and standard deviations of each cell in experimental situations; (d) t-scores, F-scores or statistical values; and (e) detailed information regarding the research design (e.g., task, GSS used). We had to discard seven studies (10%) that otherwise met our criteria for inclusion (i.e., a control group and a comparison group) because they failed to report means, standard deviations, or statistics that would enable us to calculate a $d$ statistic (i.e., effect size) for the study.

One final issue is the need for comparative field studies. Field studies play an important role in the research process (Cook and Campbell 1979). Laboratory studies (the majority of our sample) primarily focus on internal validity. It is equally important to address external validity, how well those theoretical processes generalize beyond the artificial constraints of the laboratory. If the processes of interest are still observed in an uncontrolled environment with all of its extraneous influences, this bespeaks of the power of those processes. Even though the effect sizes from field studies may not be the same as those in a laboratory, the effect sizes over a
body of field studies should be consistent with the theoretical expectations. We strongly encourage researchers to undertake more field studies that include comparison groups interacting without a GSS. Although this is challenging because most organizations do not want to mix the use of GSS and non-GSS approaches, Cook and Campbell (1979) note that it is possible to create such "natural experiments" when it is impossible to provide a new treatment (i.e., GSS) to all potential users due to capacity issues. In rationing the new resource (i.e., GSS), one can create a natural design.

**Implications for Practice**

For practitioners, the key question has been, “does GSS use improve or impair group performance?” In 1993, the answer was that in general, GSS use could be expected to improve decision quality, increase the number of ideas/alternatives, increase the time taken, have no effect on satisfaction with the process, and reduce satisfaction with the outcomes (Benbasat and Lim, 1993). Today, our answer is more complex (see Table 4). Perhaps a better question now is “in what situations does GSS use improve or impair group performance?”

When used appropriately (i.e., there is a task-technology fit and the group receives appropriation support if they lack extensive experience in applying the GSS to the task), then GSS use should increase the number of ideas generated, take less time, and lead to more satisfied participants than if the group works without the GSS. However, if the GSS structures do not fit the task or no appropriation support is provided, then GSS use may still result in more ideas, but will likely take longer and result in less satisfied participants. The need for ideas and corporate creativity is again being recognized by many large companies such as Shell, Nortel, and Proctor and Gamble, which have created GSS-based "idea factories" to spark innovation and creativity in the hopes of generating radical, paradigm-breaking ideas that can transform current products and services (Stepanek, 1999). From a practitioner’s standpoint, one should not focus on whether or
not to use GSS for a task, but instead on selecting appropriate GSS structures for the task, and then providing the right appropriation support.

For example, FAM has distinct implications for the use of GSS distributed in time and space. FAM argues that for generating ideas or discussing decision alternatives, electronic communication with information processing support is a good fit; verbal communication is not necessary. Thus these tasks (or sub-tasks) should be able to be accomplished in virtual teams whose members work distributed in space and time. However, for the actual decision making process, verbal communication should improve performance. In this case, a face-to-face meeting, a video-conference, an audio conference, or a series of telephone calls may be called for.

The performance that results from GSS use is a critical issue, but managers also have investment decisions that must be made; sometimes trade-offs have to occur, making “ideal” circumstances infeasible. FAM offers managers some guidance regarding what trade-offs can be made and what effects may result. For example, fit improves effectiveness, even in cases without appropriation support. Thus it may be best for managers to devote attention to selecting GSS tools that best fit the tasks as a first priority for GSS use. Although processes may initially take longer and receive lower satisfaction ratings, outcomes (i.e., decision quality, number of ideas) likely will improve using only the “fitted” GSS.

The most expensive component in the organizational use of GSS is the facilitator. Facilitators are a large expense, particularly when outside experts or consultants are used, but also when an internal group of facilitators needs to be maintained. If the primary contributions of a GSS facilitator are as Bostrom et al.'s (1996) data suggest -- to provide a task-technology fit and appropriation support in the form of meeting design, rather than managing group dynamics, or
helping to perform the task by providing content suggestions (see Griffith et al. 1998)\textsuperscript{13} -- then the need for facilitators may decline after GSS is fully appropriated and habitual routines develop. Even prior to this full appropriation, it may be possible to provide the major benefits associated with facilitation through other less expensive appropriation support mechanisms, such as appropriation training, restrictive software, or use templates that prescribe organizational norms.

It has been argued that the only consistent pattern in GSS research has been the inconsistency of findings. Our meta-analysis again shows the lack of consistency in the overall pattern of results (see Table 3). However, by using FAM to organize the studies, a sharply different pattern emerges. We conclude that, in general, the findings of prior research are \textit{not} inconsistent when using a theoretical lens such as FAM.

\textsuperscript{13} One might also argue that if the primary contribution of the facilitator is group dynamics or task contributions, is there really a need for the GSS if a facilitator is present (see Weick and Meader 1993)?
**Table 1**

The Evolution of GSS Structures

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1 Support</strong></td>
<td>Process Support</td>
<td>Communication Support</td>
<td>Communication Support</td>
<td>Supports or enhances communication among participants</td>
<td>Anonymity, Parallelism</td>
<td></td>
</tr>
<tr>
<td><strong>Level 2 Support</strong></td>
<td>Task Structure</td>
<td>Information Processing Support</td>
<td>Information Processing Support</td>
<td>Supports the evaluation, gathering, aggregation, structuring, and analysis of information</td>
<td>Voting, Modeling</td>
<td></td>
</tr>
<tr>
<td><strong>Process Structure</strong></td>
<td>Process Structure</td>
<td>Appropriation Support</td>
<td></td>
<td>Defines the process by which the group works on the task to help the group resolve the task in the most effective and efficient way(s) possible</td>
<td>Agenda, Facilitation, Training, Restrictive Software</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2
Hypotheses

<table>
<thead>
<tr>
<th></th>
<th>Fit</th>
<th>No Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appropriation Support</td>
<td>I</td>
<td>III</td>
</tr>
<tr>
<td>No Appropriation Support</td>
<td>II</td>
<td>IV</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>I</td>
<td>&gt;</td>
</tr>
<tr>
<td>H2</td>
<td>I</td>
<td>&gt;</td>
</tr>
<tr>
<td>H3</td>
<td>II</td>
<td>&gt;</td>
</tr>
<tr>
<td>H4</td>
<td>III</td>
<td>=</td>
</tr>
</tbody>
</table>
### Table 3

**Overall Effect Sizes**

<table>
<thead>
<tr>
<th>Measure</th>
<th>n</th>
<th>Effect Size</th>
<th>Prior Effect Sizes</th>
<th>Fail-Safe N</th>
<th>Homogeneity Q</th>
<th>Percent Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Quality</td>
<td>67</td>
<td>.07</td>
<td>.32***</td>
<td>.29**</td>
<td>150.97***</td>
<td>55 %</td>
</tr>
<tr>
<td>Number of Ideas</td>
<td>42</td>
<td>.86***</td>
<td>1.03**</td>
<td>138</td>
<td>111.28***</td>
<td>41</td>
</tr>
<tr>
<td>Time</td>
<td>26</td>
<td>.67***</td>
<td>.54***</td>
<td>.94**</td>
<td>60</td>
<td>32</td>
</tr>
<tr>
<td>Outcome Satisfaction</td>
<td>23</td>
<td>-.08</td>
<td>-.43**</td>
<td>98.23***</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>Process Satisfaction</td>
<td>39</td>
<td>-.25*</td>
<td>-.22***</td>
<td>.05</td>
<td>391.61***</td>
<td>13</td>
</tr>
</tbody>
</table>

* p<.05, **p<.01, ***p<.001