

The Effect of Feedback and Learning on DSS Evaluations

By

Ujwal Kayande

The Pennsylvania State University

Arnaud De Bruyn

ESSEC Business School

Gary Lilien

The Pennsylvania State University

Arvind Rangaswamy

The Pennsylvania State University

Gerrit Van Bruggen

Rotterdam School of Management
Erasmus University

ISBM Report 1-2006

Institute for the Study of Business Markets
The Pennsylvania State University
484 Business Building
University Park, PA 16802-3603
(814) 863-2782 or (814) 863-0413 Fax
www.isbm.org, isbm@psu.edu

The Effect of Feedback and Learning on DSS Evaluations

Ujwal Kayande*
Arnaud De Bruyn**
Gary Lilien*
Arvind Rangaswamy*
Gerrit Van Bruggen***

January 10, 2006

*Ujwal Kayande is Assistant Professor of Marketing, Gary L. Lilien is Distinguished Research Professor of Management Science, and Arvind Rangaswamy is the Jonas H. Anchel Professor of Marketing in the Smeal College of Business at Pennsylvania State University.

**Arnaud De Bruyn is Assistant Professor of Marketing at ESSEC Business School.

***Gerrit Van Bruggen is Professor of Marketing at Rotterdam School of Management, Erasmus University.

Contact information: Ujwal Kayande, 479, Business Building, Smeal College of Business, Pennsylvania State University, University Park, PA 16802, USA. Phone: 814 863 4250, email: ukayande@psu.edu.

The Effect of Feedback and Learning on DSS Evaluations

Abstract

Model-based decision support systems (DSSs), designed to help decision-makers make better decisions, often do not help decision makers understand either how or why they work. As a result, there is likely to be a large gap between a manager's mental model and the decision model embedded in the DSS. We suggest that this gap is an important reason for the poor subjective evaluations of DSSs, even when the DSSs have been shown to be of high objective quality, ultimately resulting in unexpectedly poor DSS adoption and usage. In this paper, we hypothesize that to increase its effectiveness, a DSS should not only be of high quality, but must also help reduce any mental model-DSS model gap. We evaluate two design characteristics that together lead users to update their mental models, resulting in better DSS evaluations: providing feedback on *upside potential* and providing suggestions for *corrective actions*. We hypothesize that, in tandem, these two types of feedback induce managers to update their mental models, a process we call deep learning, whereas individually, these two types of feedback will only have a small or negligible effect on deep learning. We validate our framework in an experimental setting, using a realistic DSS in a direct marketing context. We conclude with a discussion of both the theoretical and practical implications of our findings.

1.0 Introduction

Firms invest heavily in implementing such IT (Information Technology) systems as CRM (Customer Relationship Management) and ERP (Enterprise Resource Planning), aimed at helping managers make better decisions and making their organizations more effective. For example, the global market for on-line analytical processing (OLAP) software, which assists in active decision-support applications such as marketing and sales analysis, direct marketing, and profitability analysis, was estimated to be worth about \$4.3bn in 2004 (OLAP Report 2005). But there is continuing uncertainty about the value of such investments in driving productivity improvements, resulting in the belief that such IT investments by themselves do not enhance productivity (see, for example, Carr 2004; IDC Reports 2002, 2004). Only when knowledge workers are able to transform their work through IT do such investments have big payoffs: “In the end, sustained white-collar productivity enhancement is less about breakthrough technologies and more about newfound efficiencies in the *cerebral production function* of the high value-added knowledge worker” (Roach 2002, quoted in IDC report, italics added). This quote captures the essence of the critical IT productivity issue that we address in this paper: that while good technology might raise expectations of the return on technology investments, firms will realize that expected return only if the technology is able to change the mental models (i.e., the cerebral production function) of knowledge workers.

The challenge then is to design IT systems, for marketing and other functions of the firm, that are not only of good objective quality, but also capable of changing the mental models of knowledge workers. This concern is particularly relevant for the design of a model-based decision support system (hereafter referred to as a DSS), such as a CRM system, because of the conflicting evidence in the literature about the value of decision models that are embedded in such DSSs. Focusing only on marketing DSSs, several researchers have shown, in field and laboratory settings, that well-designed decision models improve the *objective* quality of decision making (e.g., McIntyre 1982, Lodish et al. 1988, Hoch and Schkade 1996, and Lilien et al. 2004), thus improving performance. However, other studies have shown that decision makers have difficulty in recognizing the value of the decision models they are using (e.g., McIntyre 1982, Davis 1989, Van Bruggen et al. 1996). We propose that the value expected from using a model-based DSS is realized by the user only when the DSS is able to bring the mental models of decision-makers closer to the decision

model embedded in the DSS (hereafter referred to as the DSS model), i.e., by helping managers update their mental models; the objective quality of the DSS model is not sufficient to ensure that users will realize the full value of the DSS.

In Figure 1, our 3-Gap Model summarizes our perspective on the DSS evaluation problem (discussed in detail in Section 2.1). We hypothesize that the sizes of the gaps between three models of the decision environment – the manager’s mental model, the DSS model, and the unknown true model (which generates data in the real world, but is only partially observed ex-post) – determines managers’ decisions, the consequent outcomes, and DSS evaluations. To provide high quality support, the gap between the DSS model and the true model must be small (Gap 2 in Figure 1). A large management science and IS literature is focused on closing this gap. However, for a DSS to be adopted and used, we hypothesize that it should actively contribute towards helping users *internalize* the high-quality DSS model, because that will make users more comfortable with the recommendations given by the DSS. This internalization will reduce the gap between the DSS model and the mental model (Gap 1 in Figure 1). By definition, if Gap 2 is small, reducing Gap 1 leads to users forming a better understanding of the real world data generating process (reducing Gap 3). Our goal in this paper, then, is to identify DSS design characteristics that lead to a reduction in Gap 1.

The key question then is how to design a DSS that can help in reducing Gap 1. We propose and show that a DSS that provides feedback about upside potential (i.e., how much more can be gained by internalizing the DSS model) *and* feedback on corrective actions (i.e., guidance on how the mental model can be “corrected”) results in significant mental model updating. By incorporating both these types of feedback (i.e., upside potential and corrective actions), a DSS can help users to significantly update their mental models, while providing only one feedback type results in a small or negligible update of the mental model. We also find that, after controlling for differences in objective performance, the magnitude of change in the users’ mental model significantly affects DSS evaluation. Thus, we find strong support for our argument that the ability of a DSS to change mental models is an important driver of DSS evaluation, an antecedent of DSS adoption (Davis 1989). We also make a methodological contribution by developing an unobtrusive way to measure a user’s mental model, avoiding the biases associated with more obtrusive approaches (Rowe and Cooke 1995).

The paper is organized as follows. We first present a conceptual framework that identifies why the gap between the user's mental model and the DSS model influences DSS evaluation. We also review previous research that has explored how to improve the quality of DSSs, and show that that research does not address the fundamental source of the DSS evaluation problem. Then we develop a model of how feedback on upside potential information and corrective actions combine to change users' mental models. We develop specific hypotheses and then test them in a realistic experimental setting. We conclude by discussing the theoretical and practical implications of our work.

2.0 Mental Model Changes, DSS Evaluation, and DSS Design: Theory Development

2.1 The Effects of Mental Model Changes on DSS Evaluation

Users' subjective evaluations of DSSs are often misaligned with the objective quality of DSS models. While there could be many reasons for such misalignment, we identify the inability of DSSs to close the gap between the DSS model and users' mental models as a potential source of the problem.

High-quality DSS models provide decision makers with useful information and insights about the optimal values of, and/or weights, to be placed on decision variables. Using this information from a DSS, decision makers are more likely to make decisions that lead them to obtain higher market shares, revenues and/or profits (Lilien and Rangaswamy 2004). Users of high quality DSSs often perform better by simply implementing DSS recommendations in a mechanistic fashion, but without necessarily understanding the rationale behind the DSS's recommendations. This lack of understanding reflects a fundamental gap between the DSS model and the user's model of the decision environment. If the gap between the DSS model and the mental model is large, the DSS model's recommended course of action and that implied by the user's mental model are likely to vary, which could result in high uncertainty about the value of the DSS (Einhorn and Hogarth 1980). In line with research on risk-adjusted preference (Keeney and Raiffa 1976, Roberts and Urban 1988), we argue that the objective quality of the DSS is then likely to be discounted by a risk-averse decision-maker to account for the high uncertainty, leading to poor evaluations of the DSS. Therefore, we suggest that one potential source of the DSS evaluation problem lies in the relative inability of current DSS designs to change the mental models of decision makers, i.e., to close the gap between the user's mental model and the DSS model (Gap 1 in Figure 1). As a consequence, we suggest that the greater the change in the mental

model, the better is the evaluation of the DSS that is used to effect the change (formalized later as Hypothesis H1 in Section 2.3)¹.

As noted earlier, much of the work in operations research and the management sciences focuses on developing decision models that accurately represent a real world data-generating process (small Gap 2). On the other hand, researchers interested in the design of DSSs have explored factors that lead to greater DSS usage and/or better performance. These factors include “task-technology” fit (Goodhue and Thompson 1995, Lim and Benbasat 2000), tabular vs. graphical presentation format and color (e.g., Lusk and Kersnick 1979; Benbasat and Dexter 1985), fit between cognitive style and presentation format (Dickson et al. 1977, Ramaprasad 1987), accessibility (e.g., Mawhinney and Lederer 1990), adaptability/flexibility (e.g., Udo and Davis 1992), perceived ease of use and usefulness (Davis 1989; Kim and Malhotra 2005), information quality and systems quality (e.g., Delone and McLean 1992), and restrictiveness of the DSS guidance (Silver 1990).

Although this stream of research is valuable in understanding the drivers of DSS usage, there is continuing evidence that users do not appear to value DSSs (Lilien et al. 2004). To our knowledge, little attention has been paid to feedback as a DSS design characteristic that can reduce the gap between mental model and DSS model, an issue we address next.

2.2 Effects of Feedback on Mental Model Changes (reducing Gap 1 in Figure 1)

Mental models represent the knowledge base accumulated in an individual’s long-term memory (Sweller et al. 1990), which serves as a framework for filtering and interpreting information within a decision environment and determining appropriate courses of action (Gary and Wood 2005). Mental models are known to be sticky (Diehl and Serman 1995; Sweller 1988), so that changing them is difficult. But to be adopted and viewed as providing an expected return on investment, DSS design must incorporate characteristics that induce learning, i.e., a positive change in the user’s mental model. The changes in mental models could be of at least two types – relatively permanent deep changes or transient changes that are the result of a mechanistic approach to the task. We formally define these changes as follows:

¹ This effect assumes that the DSS model is of high objective quality (small Gap 2) and that it is much better in quality than the user’s mental model (large Gap 1). For the remainder of the paper, including the empirical study, we assume that Gap 2 is very small. We operationalize this assumption in our experiments by only using a DSS in which we directly control, and ensure, it is of high quality.

Deep learning is a change in an individual's mental model that endures over time and/or over changes in conditions – in other words, a change that concerns the relatively permanent acquisition of skills, understanding, and knowledge (Goodman 1998).

Mechanistic learning, in contrast, is a change in an individual's mental model that occurs only in the presence of external feedback or other conditions of practice, but disappears over time or when the supportive conditions are eliminated (also called shallow learning; see Kluger and Denisi 1996, pg. 278).

A transient change in a mental model, i.e., mechanistic learning, will not influence a user's uncertainty about the DSS recommendations, while a permanent change will tangibly reduce the uncertainty about the DSS recommendations, resulting in improved DSS evaluations. Our main interest, therefore, is in deep learning and how it affects DSS evaluation. Sprinkle (2000), drawing on Agency theory (Eisenhardt 1989), suggests that for deep learning to occur, individuals must (i) have knowledge of why they should exert the effort to learn, i.e., they must have an incentive to exert the effort to learn, and (ii) individuals must be provided guidance on how they can perform better. Put differently, deep learning occurs when there is *guided effort*, defined as follows:

Guided effort is the combination of high levels of effort from the individual and high levels of guidance to the individual (say, from a DSS) on how to perform better.

We formalize the effect of guided effort on deep learning as Hypothesis H2 in Section 2.3. We propose further that feedback on upside potential and corrective actions together provide the necessary levels of guided effort for users to significantly update their mental models. Next we describe how each type of feedback individually influences guided effort, and therefore affects learning.

2.2.1 Effects of Upside Potential Feedback on Learning. Information about upside potential answers the following question – what increase in financial performance could the user gain from internalizing the DSS model? In this sense, upside potential feedback serves as an incentive to exert the effort required to learn the DSS model. According to Agency theory (Eisenhardt 1989), rational individuals are motivated by incentives to perform better, and therefore, work harder, as long as the relationship between effort, performance, and incentives is transparent and salient. Goal-setting theory (Locke et al. 1981, Bandura 1997) offers another explanation of why upside potential feedback might lead to increased effort – feedback on upside potential helps decision-makers not only in setting specific and challenging goals, but also in committing to achieve those goals by exerting more effort. Chenoweth et al. (2004) provide empirical support

that DSS users are willing to put in more effort to learn complex models if the link between the complex model, performance, and incentives is made more salient.

However, several studies have reported that incentives, although associated with increased effort, are frequently not empirically associated with better performance. Wood et al. (1990) and Hogarth et al. (1991) argue that incentives serve as extrinsic rewards which focus attention on the self, rather than on the task. As a result, task-learning processes are not activated (Kluger and Denisi 1996), leading to only mechanistic learning and poorer out-of-task performance.

In summary, the benefit of providing an incentive is that it obtains increased motivation and effort from the user. However, incentives alone direct attention away from the task, resulting in *increased effort without appropriate guidance*. So if upside potential feedback were to be combined with a mechanism that focuses attention on the task, we would expect users to exert the increased levels of guided effort necessary to obtain a significant level of deep learning. However, in the absence of such a guidance mechanism, the behavioral outcome of providing only upside potential feedback should be insignificant deep learning, because of an associated low level of guided effort. The effect of upside potential feedback on learning is summarized in Figure 2a.

2.2.2 Effects of Corrective Feedback on Learning. Corrective feedback can improve decision making, particularly in complex tasks, through increased attention to task-learning processes and better quality of decision making (Balzer et al. 1992, Kluger and DeNisi 1996). This attention to task-learning processes improves performance. However, research also suggests that such feedback effects might only be transient – removal of the feedback can bring performance back to where it originally was (Kluger and DeNisi 1996). Thus, corrective feedback might only lead to mechanistic learning, not deep learning, because individuals simply adjust behavior by using the feedback rather than by focusing on understanding the task. This heuristic processing does not lead individuals to understand the complexities of the decision environment, and therefore does not lead to deep learning. Goodman (1998) reports that external feedback, i.e., feedback provided by a source external to the task environment, is detrimental to deep learning and improved out-of-task performance, even if it improves within-task performance. Atkins et al. (2002) find that if feedback is presented such that it is very easy for decision-makers to derive guidelines for action, it may

cause decision-makers to only derive these guidelines from the information presented to them and not exert the effort to understand the rationale underlying these guidelines. They argue further that learning should benefit from feedback to the extent that feedback promotes the exploration and information-processing activities, both of which require effort from individuals.

In summary, the benefit of providing corrective feedback is that it directs attention to the task and task-learning process. However, corrective feedback is also known to lead to less exploration and therefore no increase in effort, which is necessary to obtain deep learning. The result then is *increased guidance without increased effort*, resulting in low levels of deep learning. The effect of corrective feedback on learning is summarized in Figure 2b.

2.2.3 Effects of Combining Upside Potential Feedback and Corrective Feedback. The arguments above suggest that these two types of feedback should be viewed as complementary mechanisms, each providing what is missing in the other; if the two feedback mechanisms are combined, the result should be an increase in guidance *and* effort, leading to deep learning. In a study of how monetary incentives interact with management accounting information, Sprinkle (2000) showed that the combination of monetary incentives and belief-revision feedback leads to a greater magnitude of learning, supporting our arguments. The combined effects of upside potential and corrective feedback are summarized in Figure 2c.

2.3 Hypotheses

Figure 2d provides a summary of our theoretical framework, relating DSS design characteristics to effort and guidance, deep learning, and DSS evaluation. We propose that a user's evaluation of a DSS increases with deep learning, after controlling for mechanistic learning and objective performance.

Our first hypothesis relates DSS evaluation to the change in mental models:

H1: An increase in deep learning leads users to provide more favorable evaluations of the DSS, after controlling for mechanistic learning and objective performance.

Mechanistic learning captures the extent to which the user has simply used some mechanistic method to perform the task without significantly updating his or her mental model. Because mechanistic learning involves no significant updating of the mental model, we expect that it would not have a significant impact on evaluation of the DSS.

We then hypothesize that it is the combination of increased effort and guidance – the extent of guided effort – that leads to deep learning.

H2: An increase in guided effort leads to an increase in deep learning.

We also hypothesize that neither effort nor guidance alone leads to deep learning:

H2.1: An increase in effort without guidance does not lead to deep learning.

H2.2: An increase in guidance without effort does not lead to deep learning.

Next, we examine the process by which DSS design characteristics affect guided effort. We hypothesize that effort is induced by the motivation to do better, which in turn, is influenced by the presence of feedback on upside potential, but not by corrective feedback. On the other hand, guidance is influenced by the presence of corrective feedback, but not by feedback on upside potential. Therefore, we hypothesize:

H3: More motivated users will exert more effort.

H4: Users will be more motivated when provided with upside potential feedback than when they are not provided such feedback.

H4.1: There will be no difference in levels of motivation between users who are provided with corrective feedback and those who are not provided with such feedback.

H5: Users will perceive a DSS with corrective feedback to be better at guiding them than one that does not have corrective feedback.

H5.1: There will be no difference between the levels of guidance perceived by users who are provided with upside potential feedback and those who are not provided with that feedback.

These hypotheses together comprise a test of the process model proposed in Figure 2d. In addition, we investigate the aggregate consequences of our process model by examining whether deep learning and mechanistic learning occur in each of the conditions, and whether guided effort is different across feedback conditions. We expect deep learning to occur when both types of feedback are provided and mechanistic learning to occur when only one is provided, because of the higher level of guided effort expected when both types of feedback are provided than when only one is provided.

3.0 Empirical Study

Our empirical study was designed to incorporate a challenging set of criteria; we sought:

- (i) a decision environment comparable in realism to those faced by managers,
- (ii) a decision environment sufficiently complex so that managers would benefit from using a DSS, but not so complex as to be outside the skill range of our research participants,
- (iii) a DSS whose underlying model sufficiently captures the real-world phenomenon (i.e., a small Gap 2 in Figure 1),
- (iv) a context that would allow us to measure the user's mental model unobtrusively,

- (v) a task in which we would be able to embed the DSS with each of the feedback characteristics (upside potential feedback and corrective feedback), both individually and jointly,
- (vi) a task that would allow us to measure deep and mechanistic learning unobtrusively, and
- (vii) a task that would allow us to measure the process variables of interest (effort, guidance, and motivation).

Criteria (i)-(iv) relate to design of the overall context of the study, while criteria (v)-(vii) relate to design of the specific experiment to test our hypotheses. With respect to criterion (i), we note that to assess mental model changes (i.e., learning), we must provide immediate and accurate feedback to DSS users about the relationship between the decisions they make and the corresponding market response. Such feedback is often lacking in real world situations because, (i) outcomes are commonly delayed and not easily attributable to a particular decision/action; (ii) variability in the environment degrades the reliability of the feedback, especially where outcomes of low probability are involved, (iii) there is often no information about what the outcome would have been if another decision had been taken; and (iv) important decisions are often unique and provide little opportunity for learning (Tversky and Kahneman 1987, p. 90). Therefore, to obtain both realism and control, we tested our hypotheses under controlled experimental conditions using a frequently occurring and realistic decision problem for which we could offer immediate feedback with known reliability. Such control is not possible in a field study.

We now describe the design of the study context and the experimental procedure.

3.1 Design of the context

The solicitation of donations through direct mail for non-profit or charitable organizations provides a context that corresponds to the above criteria. In the US alone, direct mail is the medium that accounts for between \$20 billion and \$25 billion of the charitable educational and social change dollars contributed annually (Lister 2001, p.2). Direct marketing managers in charitable organizations typically solicit donations using large databases of potential donors. Each solicitation has a cost attached to it, and donation amount is donor-specific, so that it is critical for the direct marketing manager to identify the most likely (and high value) donors. This situation, in turn, requires the manager to have an understanding of the factors that influence the donor's likelihood of donation – that is, a mental model of the drivers of donation. DSSs are often used by direct marketing manager to assist them in selecting high potential donors, and “it is not uncommon for self-serve direct marketers to realize a 20% to 350% increase in response rates simply by using [direct

marketing decision support] software” (MarketMiner Analyst™ website, www.modelingautomation.com). In our study, we asked participants to assume the role of a direct marketing manager of a large nonprofit charity, and we provided them with a DSS to assist in their decision making. Their main task was to identify the most attractive donors from a database of past donors for solicitation in a direct marketing campaign.

The context above satisfies the ‘realism’ criterion (i); to satisfy criteria (ii) and (iii), we sought a decision environment that would be sufficiently complex and a DSS model that would be sufficiently close to the true data-generating model. Second, to satisfy criterion (iv), we sought a method to calibrate the participant’s mental model.

3.1.1 Decision Environment: We constructed a database of 200,000 hypothetical donors, described on four characteristics – their *recency* of donation (the number of quarters since last donation), their *frequency* of donation (the number of donations the donor has made in the past 5 years), their *amount* of past donations (the average donation amount, in dollars, observed in the past for this particular donor), and the donor’s *age*. These characteristics are often used by direct marketing firms in targeting models, typically referred to as RFM (Recency-Frequency-Monetary Value) models. In line with the most common models of purchase probability (Lilien et al. 1993), we modeled the probability that a particular donor would make a donation, if solicited, by a Logit function as follows:

$$(1) \quad p = 1 / (1 + \exp(5 - (X / 20)))$$

where X is called the donor's “attractiveness”, and is given by:

$$(2) \quad X = \beta_0 + (\beta_1 \times \text{recency}) + (\beta_2 \times \text{frequency}) + (\beta_3 \times \text{amount}) + (\beta_4 \times \text{age})$$

The parameters of the 'true' data generating model were $\beta = \{20, -20, 40, 10, 30\}$. In line with these parameters, we informed participants that donors were more likely to donate if they had donated more recently, more frequently in the past five years, in greater amounts, and if they were older.

The four donor characteristics vary greatly in terms of the units of measurement (for e.g., donors’ ages is measured in years, but recency is measured by number of quarters since last donated), which makes their coefficients non-comparable. Therefore, we scaled each characteristic to be between 0 and 1 in the database, even though we used realistic units of measurement to describe the characteristics to study

participants. This scaling ensured that the relative importance of a characteristic did not depend on its unit of measurement.

Our database of 200,000 hypothetical donors had to satisfy two criteria: first, we wanted probabilities of donation in our database to be similar to those observed in not-for-profit databases; second, we wanted the characteristics to be generated so that we could then describe the attractiveness of each donor on a reasonable scale (say, 0 to 100) with an average at the midpoint, so that we could subsequently ask participants to rate each donor on the same scale. Thus, we generated donor characteristics from uniform distributions between 0 and 1, independent of one another, with the functional form of the logit function as described in equations (1) and (2). As a result, a donor's true attractiveness varied between 0 and 100 with an average of 50, and donors' probability to donate varied between 0.67% and 50% with an average of 7.6%. Although in practice, average response rates vary widely, depending on the charity, the cause, the targets and the creative proposition, an average response rate of 7.6% falls within industry averages for 'warm' donors (see <http://www.fundraising.co.uk/forum/thread.php?id=500>)

As per criterion (iii), we designed Gap 2 to be very small by constructing a DSS model that was identical to the true model in terms of the weights of each factor. To obtain a sufficient level of realism, we added a small random error to the true model in equation (1) so that the actual donations from donors could be predicted only approximately. On average, the true model and DSS model were identical.

3.1.2 Calibrating Participants' Mental Models: In line with the work on estimating managerial decision coefficients (Kunruether 1969), we needed participants to make a sufficient number of decisions at one time to calibrate their mental models. Such a process would allow us to infer their mental model without having to actually ask the participant to reflect on their mental processes. The latter method is known to suffer from significant biases (Rowe and Cook 1995), which our unobtrusive method avoids. We asked each participant to rate 20 donors from the database on how attractive each of the donors was for selection in a marketing campaign, using a 100-point sliding scale, where 0 meant not attractive at all for a solicitation, and 100 most attractive. The 20 donors were described along the four drivers of donation behavior – recency, frequency, donation amount, and age. We provide a screen shot of the participant's task in Figure 3.

To obtain the unobtrusive measure of the participant's mental model, we statistically related their donor ratings to the descriptions of the 20 donors and thus inferred the implicit weights participants placed on the four factors. To estimate this relationship, there must be sufficient variation in the description of the donors on each of the four factors and the factors must not be multicollinear, allowing independent estimation of each weight. While a fractional factorial design is typically the design choice in such cases, the number of possible profiles made that approach infeasible. Therefore, we generated donors' characteristics (recency, frequency, etc.) using cosine functions, so that extreme values were represented more often in the sample than in the population, while spanning the entire parameter space. To avoid multicollinearity, we randomly permuted donors' characteristics in the sample until no inter-item correlation was higher than .15.

Once participants submitted their ratings, we estimated a linear regression model to estimate the implicit weights ($\beta_0', \beta_1' \dots \beta_4'$) the participant placed on recency, frequency, donation amount and age. We then applied this calibrated mental model to the larger database of 200,000 donors to determine who to solicit and who not to. We told participants that each solicitation would cost \$2 and, if successful, would generate a constant \$20 donation (to keep the task within participant skill range, criteria ii), yielding a profitability threshold of 10% probability of donation. We applied the estimated mental model to the entire database, computing X' and p' for each of the 200,000 donors, and soliciting those donors with $p' > 0.1$. In addition to the marginal costs of solicitation, the fundraising campaign was subject to fixed costs of \$10,000. To determine whether a solicited donor actually made a donation, we drew a random number z from a uniform distribution [0,1] for each donor, and each solicited donor made a donation of \$20 if z was less than the true probability of donation (p). Note that if participants provided perfectly accurate scores ($X = X'$), mental parameters would be equal to true parameters ($\beta = \beta'$), and the solicitation strategy would be optimal.

To assist the participants in their decision-making, we provided them a DSS to select attractive customers from a database. To enhance the external validity of our study, we ensured incentive alignment by informing participants in all conditions that the amount of money they earned would be directly proportional to their financial performance. Participants were paid 0.015% of their financial performance on two tasks, Task 1 and Task 2 (described fully in the next section), in addition to a \$15 participation fee.

3.2 Design of the experiment (manipulations and measurements)

We summarize the sequence of steps in our experiment in Box A of Figure 4. Our experiment consisted of three main parts, addressing study design criteria (v), (vi), and (vii) respectively.

3.2.1 Part 1 of the Study: Using the DSS. In Part 1 of the study, we asked participants to rate the 20 donors in each of ten simulations. The participants had access to a DSS to help them determine the best possible ratings. In each simulation, participants rated the 20 donors, submitted those ratings to the DSS simulator, and obtained the DSS's expectation of the performance of the campaign based on the participants' donor ratings. In the background, we calibrated the regression relationship between the participants' ratings and the description of the 20 donors on the four factors. The DSS was therefore both a support tool for users to make decisions, and also a research tool to measure users' mental models.

We varied the feedback provided by the DSS to reflect four different feedback conditions:

1. "OUTCOME FEEDBACK": The participant was only informed of the expected performance of the donor ratings. This condition serves as a control in our experiment. For example:

The DSS predicts that a marketing campaign based on your ratings would generate \$76,654 in revenue.

2. "UPSIDE FEEDBACK": In addition to information about the expected performance, the participants in this condition were also informed of the maximum financial performance they could have achieved if they had been able to uncover the 'true' attractiveness scores of the 20 donors. For example:

The DSS predicts that a marketing campaign based on your ratings would generate \$76,654 in revenue.

The DSS predicts that it would be possible to generate up to \$99,934 in revenue from this database.

3. "CORRECTIVE FEEDBACK": In addition to information about the expected performance of the donor rating strategy, participants in this condition were given feedback on whether they were placing too much or too little weight on each of the four factors. To operationalize this feedback, we compared the participants' mental model parameters to the parameters of the DSS model. For example:

The DSS predicts that a marketing campaign based on your ratings would generate \$76,654 in revenue.

Here is some corrective feedback that will help you improve your ratings. In developing your ratings for these donors:

- you assume a relationship between recency and donating behavior that is opposite to what is known.
- you are greatly overestimating the importance of frequency.
- you are underestimating the importance of age.

4. “ALL”: In this condition, we provided participants with feedback on expected outcome, upside, and corrective actions (conditions 1-3 above), in that order. For example:

The DSS predicts that a marketing campaign based on your ratings would generate \$76,654 in revenues.

The DSS predicts that it would be possible to generate up to \$99,934 in revenue from this database.

Here is some corrective feedback that will help you improve your ratings. In developing your ratings for these donors:

- you assume a relationship between recency and donating behavior that is opposite to what is known.
- you are greatly overestimating the importance of frequency.
- you are underestimating the importance of age.

Participants were randomly assigned to one of the four conditions. To provide an incentive for participants to focus on the task during the simulations, we informed them that they would be required, after completing the ten simulations, to rate the same donors for a real direct mail campaign that we refer to as Task 1. We calibrated their mental model based on Task 1 ratings.

3.2.2 Part 2 of the Study: Measuring Learning. Our goal in Part 2 is to measure deep and mechanistic learning. Per our definition of deep learning, we sought a measure of mental model change that survives even when feedback is removed. Therefore, in Task 2, we asked participants to rate 20 donors who were different from those in Part 1. To ensure they applied their (updated) mental model of donor behavior to this task, we told them that the 20 new donors were from the same database used in Task 1, so the extent to which each factor impacted donor behavior was the same for these new donors as it was for the donors in Task 1. Because we were only interested in measuring their mental model at this stage, we did not provide the participants access to a DSS.

A change in an individual’s mental model can be of two types – the mental model becomes better than before (Gap 3 reduces) or it becomes worse than before (Gap 3 increases). To ensure that we were measuring the former and not the latter, we first constructed a measure of mental model accuracy – the

distance between the true model (which in our study is identical, on average, to the DSS model) and the mental model. Such a measure then benchmarks the mental model against the true model, providing a cleaner measure of positive change in the mental model. We sought a measure of mental model accuracy that reflects the participant's ability to judge which factors were important and the relative importance of those factors. A measure that satisfies these criteria is as follows:

$$(3) \quad WED_t = \left[\sum_j \omega_j \cdot (\beta'_{jt} - \beta_j)^2 \right]^{0.5},$$

where t is the task ($t=1, 2$), WED_t is the Weighted Euclidian Distance between the mental model and the true model in task t , ω_j is the true importance of the j^{th} ($j=1 - 4$ in our study) driver of donation behavior, β'_{jt} is the mental model parameter associated with the j^{th} driver of donation behavior in task t , and β_j is the true parameter associated with the j^{th} driver of donation behavior. WED_t is a measure of mental model accuracy (Gap 3) in our study (we note here that accuracy *increases* as WED_t decreases). Weighting the distance between coefficients by ω_j implies that a mental model that is close to the true model on the most important drivers is better than a mental model that is close to the true model on the less important drivers. Note that per our study design, if the mental model were to converge to the DSS model, then, on average, Gap 3 is identical to Gap 1 and $\omega_j = \beta_j$. (We obtained substantively equivalent empirical results with an equally weighted, simple Euclidean distance measure, suggesting that our results do not hinge on our choice of metric).

We provide a graphical explanation of our learning measures in Box B of Figure 4, relating those measures to each step of the experiment. The participant's initial mental model accuracy is measured by WED_0 , calibrated using the mental model parameters from the first simulation in the same manner as WED_t . The difference between WED_0 and WED_t is a measure of the change in mental model accuracy that is due to the participant's use of the DSS. Part of this change is a result of an internalization of the DSS model, i.e., deep learning, and part is a transient change that will disappear with the removal of the feedback, i.e., mechanistic learning. We could independently calibrate each part because of Task 2, explained next.

A rather persistent change in the mental model would be reflected in the extent to which the mental model in Task 2 is more accurate than that in the initial simulation. Therefore, we construct our measure of deep learning by taking the difference between WED_2 , mental model accuracy in Task 2, and WED_0 , the accuracy of the initial mental model. We define deep learning (DL) as,

$$(4) \quad DL = (WED_0 - WED_2).$$

In contrast to deep learning in equation (4), if mental model accuracy in Task 1 is much greater than that in Task 2, it indicates that the accuracy of the mental model in Task 1 was a result of the user being mechanistic in their approach to the task – an approach that would lead to decision quality deterioration if conditions were changed, as in Task 2 with a new set of donors. We define mechanistic learning (ML) as:

$$(5) \quad ML = (WED_2 - WED_1).$$

While Parts 1 and 2 allowed us to manipulate feedback provided by the DSS and measure participants' mental models, we also need to measure process variables and DSS evaluation, described next.

3.2.3 Part 3 of the Study: Subjective Construct Assessment. The main process variables of interest in this study were motivation, effort, and guidance. After participants completed Tasks 1 and 2, but before they were informed of the financial performance results, we asked them to complete a questionnaire measuring motivation (“The DSS motivated me to do better”), effort (four items: “I was totally immersed in addressing this problem”, “I took this task seriously”, “I put in a lot of effort”, and “I wanted to do as good a job as possible no matter how much effort it took”), guidance (“The DSS gave clear guidance on how I could do better”). After being informed of their results, participants evaluated the DSS (“I would definitely recommend a DSS like the one I had available to direct marketers”). All items were 5-point Likert questions, with 1=completely disagree, and 5=completely agree.

3.3 Sample and Experimental Procedure

We recruited 61 MBA students at a large northeastern U.S. university to assume the role of the direct marketing fundraising manager. We randomly assigned participants to one of four conditions, resulting in 14-16 participants per condition. Participants earned between \$23 and \$44, with the average payment being \$38. They took an average of 42 minutes to complete the two tasks, with times ranging from 14 to 80 minutes.

3.4 Analysis

Figure 2d sketches the process model of the effect of DSS design characteristics on deep learning and DSS evaluation. There are several sub-models embedded within this framework, with errors that are likely to be correlated. The following specification with correlated errors across all five models addresses this issue:

$$(6.1) \text{ DSS Evaluation} = \beta_{01} + \beta_{11} \cdot \text{DeepLearning} + \beta_{21} \cdot \text{MechanisticLearning} + \beta_{31} \cdot \text{MonetaryCompensation} + \varepsilon_1$$

$$(6.2) \text{ DeepLearning} = \beta_{02} + \beta_{12} \cdot \text{Effort} + \beta_{22} \cdot \text{Guidance} + \beta_{32} \cdot (\text{Effort} \times \text{Guidance}) + \varepsilon_2$$

$$(6.3) \text{ Effort} = \beta_{03} + \beta_{13} \cdot \text{Motivation} + \varepsilon_3$$

$$(6.4) \text{ Motivation} = \beta_{04} + \beta_{14} \cdot \text{UPSIDE FEEDBACK CONDITION} + \beta_{24} \cdot \text{CORRECTIVE FEEDBACK CONDITION} + \varepsilon_4$$

$$(6.5) \text{ Guidance} = \beta_{05} + \beta_{15} \cdot \text{UPSIDE FEEDBACK CONDITION} + \beta_{25} \cdot \text{CORRECTIVE FEEDBACK CONDITION} + \varepsilon_5 .$$

We estimated the parameters of all five models simultaneously using the full information maximum likelihood (FIML) routine in SAS. Hypothesis H1 (deep learning affects evaluation) can be tested by examining the significance of the coefficient β_{11} (Model 6.1). Hypothesis H2 (the combination of guidance and effort affects deep learning, but each individually does not) can be tested by examining the significance of the coefficient β_{32} and hypotheses H2.1 and H2.2 by the non-significance of the coefficients β_{12} and β_{22} (Model 6.2). Hypothesis H3 relates to the significance of coefficient β_{13} (Model 6.3). Hypotheses H4 and H4.1 relate to the significance of the coefficient β_{14} and the non-significance of coefficient β_{24} respectively (Model 6.4). Hypotheses H5 and H5.1 relate to the significance of the coefficient β_{25} and the non-significance of coefficient β_{15} respectively (Model 6.5).

In addition, we tested whether guided effort (the combination of effort and guidance) was significantly greater in the ALL condition than UPSIDE FEEDBACK condition or CORRECTIVE FEEDBACK condition. To do so, we estimated the following model using the correlated-errors modeling framework:

$$(7) \text{ GuidedEffort} = \eta_0 + \eta_1 \cdot \text{OUTCOME FEEDBACK CONDITION} + \eta_2 \cdot \text{UPSIDE FEEDBACK CONDITION} \\ + \eta_3 \cdot \text{CORRECTIVE FEEDBACK CONDITION} + \varepsilon_6 ,$$

where *GuidedEffort* is defined as,

$$(8) \text{ GuidedEffort} = \text{Effort} \times \text{Guidance} , \text{ and}$$

each of the feedback conditions (except ALL) was dummy-coded. The mean level of *GuidedEffort* in the ALL condition is indicated by the intercept term in equation (7). The significance of the coefficients indicates

whether *GuidedEffort* in each of the conditions is significantly different (smaller, if the coefficients are negative) from *GuidedEffort* in the ALL condition. Additionally, we tested whether deep and mechanistic learning were different from zero in each of the feedback conditions, and whether deep learning was significantly greater in the ALL condition than in other conditions.

3.6 Results

We hypothesized that users' evaluations of a DSS depend on the extent to which they internalized the DSS model (Hypothesis H1; Part A of Table 1). We find strong support for H1 ($\beta_{11} = 0.051, p < 0.01$). We also find that mechanistic learning is not a significant driver of DSS evaluation ($\beta_{21} = -0.019, ns$), further supporting our theory that DSS evaluation depends on a significant updating of mental models, not just on mechanistic performance improvement. We also find that monetary compensation, which reflects the objective performance of participants, is not a significant driver of evaluation ($\beta_{31} = 0.037, ns$). This result shows that deep learning is a significant driver of DSS evaluation, after controlling for mechanistic learning and monetary compensation, supporting H1.

Next, we tested Hypothesis H2, i.e., whether deep learning is affected by the combination of effort and guidance. The results (Part B of Table 1), support H2 ($\beta_{32} = 3.14, p < 0.05$). This result shows that the combination of effort and guidance is a significant driver of deep learning. The intercept term is not significant, indicating that deep learning does not occur without effort and guidance. The coefficients for effort and guidance are not significant, indicating that neither of these process variables alone is capable of obtaining deep learning. These results support hypotheses H2.1 and H2.2.

Process effects are shown in Part C of Table 1. Effort was significantly affected by a participant's motivation to do better (supporting H3; $\beta_{13} = 0.57, p < 0.01$), while a participant's motivation to do better was significantly increased by the presence of upside potential feedback (supporting H4; $\beta_{14} = 0.59, p < 0.01$). As expected, motivation to do better was not affected by corrective feedback (supporting H4.1; $\beta_{24} = -0.27, ns$). Guidance, the other component of guided effort, significantly increased with the presence of corrective feedback (supporting H5; $\beta_{25} = 0.69, p < 0.01$). Surprisingly, guidance also increased significantly when participants were provided upside potential feedback (H5.1 not supported; $\beta_{15} = 0.62, p < 0.05$).

Table 2 presents analyses related to whether the level of guided effort is different across feedback conditions. *GuidedEffort* was significantly higher in the ALL condition than in the UPSIDE FEEDBACK ($\eta_2 = -2.63, p < 0.05$, one-tailed) or CORRECTIVE FEEDBACK ($\eta_3 = -2.98, p < 0.05$, one-tailed) conditions. These aggregate results on guided effort are consistent with our process model.

Table 3 (Column A) reports tests of whether there was significant deep learning in each of the four conditions. We find that deep learning is significantly different from zero in the ALL condition (mean = 13.71, $p < 0.01$, one-tailed). There is evidence that deep learning in the UPSIDE FEEDBACK condition is also significantly different from zero (mean = 6.65, $p < 0.05$, one-tailed). This result is consistent with the significance of the effect of upside potential feedback on both effort and guidance, the two conditions required for deep learning to occur. Deep learning is not significantly different from 0 in either the outcome feedback condition or the corrective condition. In addition, we find that the level of deep learning is significantly greater in the ALL condition compared to other conditions (results shown under Column B of Table 2).

Table 4 presents the analysis of whether significant mechanistic learning occurred in each of the conditions. We find that mechanistic learning does not occur in the ALL condition ($\alpha_{42} = 2.19, ns$). A very high and significant magnitude of mechanistic learning occurred in the CORRECTIVE FEEDBACK condition ($\alpha_{32} = 7.57, p < 0.01$). A significant magnitude of mechanistic learning also occurred in the UPSIDE FEEDBACK condition ($\alpha_{22} = 3.70, p < 0.05$), implying that the observed mental model accuracy in Task 1 is partly based on a mechanistic approach specific to the 20 donors in the simulations and partly based on a real change in their mental model (as indicated by the significant deep learning found in this condition). On the other hand, the increase in mental model accuracy observed in Task 1 for the CORRECTIVE FEEDBACK condition is almost entirely a result of mechanistic learning, given the significant drop in mental model accuracy in Task 2 and the lack of significance of deep learning for this condition.

4.0 Discussion, DSS Design Implications, and Contributions

4.1 Discussion

We proposed and empirically demonstrated that deep learning is crucial for users to form a favorable DSS evaluation, a pre-cursor to adoption. Our study shows that a DSS with upside potential feedback can motivate users to perform better, resulting in greater user effort. However, increased effort alone is not sufficient to generate deep learning; the DSS must also provide clear guidance about how and why a modification of a decision process leads to a superior outcome. Our results also show that mere mechanistic learning does not lead to better evaluations of the DSS, implying that systems that offer no opportunity to understand their recommendations are likely to be poorly evaluated by users and hence, used less frequently. Although we found, unexpectedly, that upside potential provides guidance in addition to increasing effort, the combination of upside potential and specific guidance is clearly a superior way to help users internalize the DSS model.

In sum, our results suggest that DSS feedback influences users' underlying learning process, which in turn, helps users internalize the relationship between decisions and outcomes. Most prior research has only examined decision outcomes, and not the underlying processes by which DSS influences outcomes, revealing only part of the picture. Our study helps enrich the story by providing insight about why objectively superior DSSs are often not evaluated more positively by users.

4.2 Implications for DSS Design

In designing DSSs, the emphasis in the management science literature has most often been on optimizing the quality of the embedded DSS model (i.e., minimizing Gap 2). A DSS with a high-quality DSS model is likely to lead to improved performance on the tasks for which the system was developed. However, this improvement might result from users “blindly” following DSS recommendations. DSSs that do not facilitate deep learning remain a “black-box” to their users, making it difficult for them to attribute positive outcomes of better decisions to the DSS.

Many commercial DSSs do not embed any feedback, offering recommendations without providing substantiating reasons. For example, many of the DSS templates for a “best practice” marketing plan simply tell users the sequence of steps to follow in developing the plan, without indicating how or why that plan will impact performance in the situation at hand. Although such a DSS may promote better decisions, and those

decisions may lead to improved performance, users are unlikely to recognize the role the DSS played, hampering future use.

Many sophisticated commercial DSSs do provide upside potential feedback. For example, the sales module of the Siebel CRM system provides a salesperson with such information as the sales achieved by the best performing salesperson, or the sales levels in the best performing sales territory. Such feedback might motivate salespeople to try to do better, but does not suggest what they should do. Our research suggests that just providing feedback in the form of upside potential is not sufficient to obtain the maximum improvement in performance, especially in out-of-task situations (e.g., for selling a new product introduced by the firm). On the other hand, providing only guidance through corrective feedback is also insufficient. As seen in our study, an increase in effort is critical to obtaining increased accuracy.

By designing a DSS that combines upside potential feedback with corrective feedback, we can induce deep learning that both improves out-of-task performance and increases the likelihood of DSS adoption. For example, in the case of CRM systems, upside potential feedback could be information on the performance of top-performing salespeople, and corrective feedback might involve generating context-specific guidelines on what those salespeople would do in the situation at hand and why. In an Edelman prize-winning application, Lembersky and Chi (1986) present an example of how combining upside potential and corrective feedback can change users' mental models, leading to overwhelming support for a DSS in a firm's activities. In their application, a decision simulator based on a sophisticated dynamic programming model provided guidance and upside potential feedback to users (foremen) on how to cut timber stems to maximize profit. Our paper provides a theoretical explanation for the success of their application. We quote from their paper (p. 12):

“We had to gain the understanding and acceptance of the woods foremen. The foremen used VISION to see the results of cutting and allocating a sample of stems from their region using their old instructions. Then, they were given the new instructions and asked to re-cut the same set of stems. They were also encouraged to experiment with any other cutting patterns of their own invention. The experienced field expert sponsored and participated in these foreman activities. With their value demonstrated, the foremen readily embraced the new instructions and saw the implications of the dynamic programming algorithm.”

4.3 Contributions, limitations and future research

Our primary contributions are: (1) the development of the “3-Gap model to understand DSS Evaluation” (Figure 1); (2) the specification of the role of deep learning (i.e., mental model changes) on user-evaluation of

DSSs (Figure 2d); and (3) the assessment of the individual and joint effects of two types of feedback, corrective and upside, on deep learning. We have also employed an unobtrusive mechanism to assess DSS users' mental models and their changes, an approach we hope other researchers will find useful.

Our work suggests several avenues for fruitful future research. In our study, participants had weak prior beliefs about the markets in which they were operating. In more realistic situations, participants are likely to have much stronger *a priori* beliefs, which may be harder to overcome through DSS-based learning. This is an empirical issue that is worth exploring, especially for understanding the role of IT systems in bringing about enduring changes to how decisions are made in organizations. Our study focused only on different types of feedback. It may also be worthwhile to explore how different types of problem-solving modes of users (e.g., optimizing, reasoning, analogizing, and creating) influence deep learning and evaluation of DSS (Wierenga and Van Bruggen, 1997), and to explore how DSSs influence different types of learning (e.g. learning about the DSS itself and its constraints, learning about markets, and learning about their own problem-solving modes). There may be organizational constraints to learning such as budgets and hierarchical management structures. Given a strong budget constraint, a manager might ignore DSS recommendations and not learn because implementing DSS recommendations might require an increased budget. Also, if a senior manager has very strongly held beliefs, a subordinate manager making decisions might simply ignore DSS recommendations to be consistent with what the senior manager believes. An interesting research question is whether and how a DSS can overcome such constraints to learning.

Our research had several limitations because of our choice of study criteria (i) to (vii). In particular, our use of experimental subjects with (potentially) weak mental models leaves untested, the effect of mental model strength on our results. And our experimental context involved a specific task (targeting for charitable donations) and it is unclear to what extent our results are context-specific. On net, though, we hope we have taken a significant step toward better understanding the mental model barriers to DSS adoption and how those barriers might be overcome.

REFERENCES

- Atkins, P.W.B, R. E. Wood, P.J. Rutgers. 2002. The effects of feedback format on dynamic decision making. *Organ. Behavior Human Decision Processes*, 88(2) 587-604.
- Balzer, W.K., L.M. Sulsky, L.B. Hammer, K.E. Sumner. 1992. Task information, cognitive information, or functional validity information: Which components of cognitive feedback affect performance? *Organ. Behavior Human Decision Processes*, 53(1) 35-54.
- Bandura, A. 1997. *Self-efficacy: The exercise of control*. New York: W.H. Freeman.
- Benbasat, I., A.S. Dexter. 1985. An empirical evaluation of graphical and color-enhanced information presentation. *Management Sci.*, 31(11) 1348–1364.
- Carr, N.G. 2004. *Does IT Matter? Information Technology and the Corrosion of Competitive Advantage*. Harvard Business School Publishing Corporation.
- Chenoweth, T., K.L. Dowling, R.D. St. Louis. 2004. Convincing DSS users that complex models are worth the effort. *Decision Support Systems*. 37(1) 71-82.
- Davis, F. D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quart.* 13(3) 319-340.
- DeLone, W.H., E.R. McLean. 1992. Information systems success: The quest for the dependent variable. *Inform. Systems Res.*, 3(1) 60–95.
- Dickson, G.W., J.A. Senn, N.L. Chervany. 1977. Research in management information systems: The Minnesota experiments. *Management Sci.*, 23(9) 913-923.
- Diehl, E., J. Sterman. 1995. Effects of feedback complexity on dynamic decision making. *Organ. Behavior Human Decision Processes*, 62(2) 198-215.
- Einhorn, H.J., R.M. Hogarth. 1985. Ambiguity and uncertainty in probabilistic inference. *Psych. Rev.* 92(4) 433-461.
- Eisenhardt, K.M. 1989. Agency theory: an assessment and review. *Acad. Management Rev.*, 14(1) 57-74.
- Gary, M.S., R. E. Wood. 2005. Mental models, decision making, and performance in complex tasks. working paper.
- Goodhue, D.L., R.L. Thompson. 1995. Task-technology fit and individual performance. *MIS Quart.* 19(2) 213-236.
- Goodman, J.S. 1998. The interactive effects of task and external feedback on practice performance and learning. *Organ. Behavior and Human Decision Processes*, 76(3) 223-252.
- Hoch, S.J., D.A. Schkade. 1996. A psychological approach to decision support systems. *Management Sci.* 42(1) 51–64.
- Hogarth, R.M., B.J. Gibbs, C.R.M. McKinzie, M.A. Marquis. 1991. Learning from feedback: Exactness and incentives. *J. Experimental Psych.: Learning, Memory, and Cognition*, 17(4) 734-752.
- IDC Report. 2002. *The Financial Impact of Business Analytics* (www.idc.com).

- IDC Report. 2004. *World Wide Marketing Automation Applications 2005-2009 Forecast* (www.idc.com).
- Keeney, R.L., H. Raiffa. 1976. *Decision Making with Multiple Objectives: Preference and Value Tradeoffs*. John Wiley and Sons, New York.
- Kim, S. S., N. K. Malhotra (2005), "A Longitudinal Model of Continued IS Use: An Integrative View of Four Mechanisms Underlying Post-Adoption Phenomena," *Management Sci.* 51(5) pp. 741-755.
- Kluger, A.N., A. Denisi. 1996. The effects of feedback interventions on performance: a historical review, a meta-analysis and a preliminary feedback intervention theory. *Psych. Bull.* 119, 254-284.
- Kunreuther, H. 1969. Extensions of Bowman's theory on managerial decision-making. *Management Sci.* 15(8) 415-439.
- Lembersky, M.R., U.H. Chi. 1986. Weyerhaeuser decision simulator improves timber profits. *Interfaces* 16(1) 6-15.
- Lilien, G.L., P. Kotler, S. Moorthy. 1993. *Marketing Models*, Englewood Cliffs, NJ: Prentice Hall.
- Lilien, G.L., A. Rangaswamy. 2004. *Marketing Engineering: Computer-Assisted Marketing Analysis and Planning*, Reading MA: Addison-Wesley.
- Lilien, G.L., A. Rangaswamy, G.H. Van Bruggen, K. Starke, 2004. DSS effectiveness in marketing resource allocation decisions: reality vs perception. *Inform. Systems Res.* 15(3) 216-235.
- Lim, K., I. Benbasat. 2002. The influence of multimedia in improving the comprehension of organizational information. *J. Management Inform. Sys.* 19(1) 99-127.
- Lister, G. J. 2001. *Building Your Direct Mail Program*, Jossey-Bass.
- Locke, E. A., K.N. Shaw, L.M. Saari, G.P. Latham. 1981. Goal setting and task performance: 1969-1980. *Psych. Bull.* 90(1) 125-152.
- Lodish, L. M., E. Curtis, M. Ness, M.K. Simpson. 1988. Sales force sizing and deployment using a decision calculus model at Syntex Laboratories. *Interfaces* 18(1) 5-20.
- Lusk, E.J., M. Kersnick. 1979. The effect of cognitive style and report format on task performance: The MIS design consequences. *Management Sci.* 25(8) 787-798.
- Mawhinney, C. H., A.L. Lederer. 1990. A study of personal computer utilization by managers. *Inform. Management* 18(5) 243-253.
- McIntyre, S.H. 1982. An experimental study of the impact of judgment-based marketing models. *Management Sci.* 28(1) 17-33.
- Ramaprasad, A. 1987. Cognitive process as a basis for MIS and DSS design. *Management Sci.* 33(2) 139-148.
- Roberts, J.H., G.L. Urban. 1988. Modeling multiattribute utility, risk and belief dynamics for new consumer durable brand choice. *Management Sci.* 34(2) 167-185.
- Rowe, A.L., N.J. Cook. 1995. Measuring mental models: Choosing the right tools for the job. *Human Resource Development Quart.* 6(3) 243-255.
- Silver, M.S. 1990. Decision support systems: Directed and nondirected change. *Inform. Systems Res.* 1(1) 47-70.

- Sprinkle, G. B. 2000. The effect of incentive contracts on learning and performance. *The Accounting Rev.* 75(3) 299-326.
- Sweller, J. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Sci.* 12(2), 257-285.
- Sweller, J., P. Chandler, P. Tierney, M. Cooper. 1990. Cognitive load as a factor in the structuring of technical material. *J. of Experimental Psych.: General*, 119(2), 176-192.
- Tversky, A., D. Kahneman. 1987. Rational choice and the framing of decisions. in *Rational Choice* (eds Robin M. Hogarth and Melvin W. Reder), The University of Chicago Press, Chicago, IL.
- Udo, J.G, S.J.Davis. 1992. A comparative analysis of DSS user-friendliness and performance. *Internat. J. Inform. Management.* 12(3) 209-223.
- Van Bruggen, G.H, A. Smidts, B.Wierenga. 1996. The impact of the quality of a marketing decision support system: An experimental study. *Internat. J. Res. Marketing* 13(4) 331-343.
- Wierenga, B., G.H. Van Bruggen. 1997. The integration of marketing problem-solving modes and marketing management support systems. *J. Marketing*, 16(3) 21-37.
- Wood, R.E., A. Bandura, T. Bailey. 1990. Mechanisms governing organizational performance in complex decision-making environments. *Organ. Behavior Human Decision Processes*, 46(2) 181-201.

Figure 1: The 3-Gap Model: The Effect of Gaps Between Mental Model, DSS model, and True Model

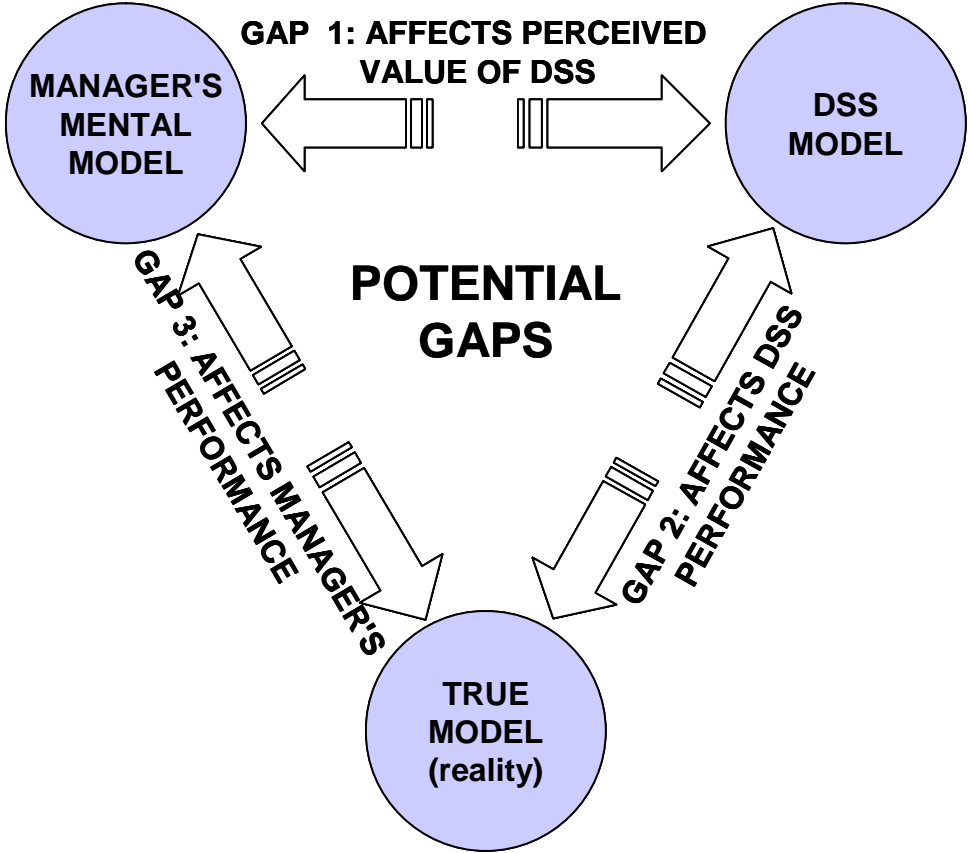
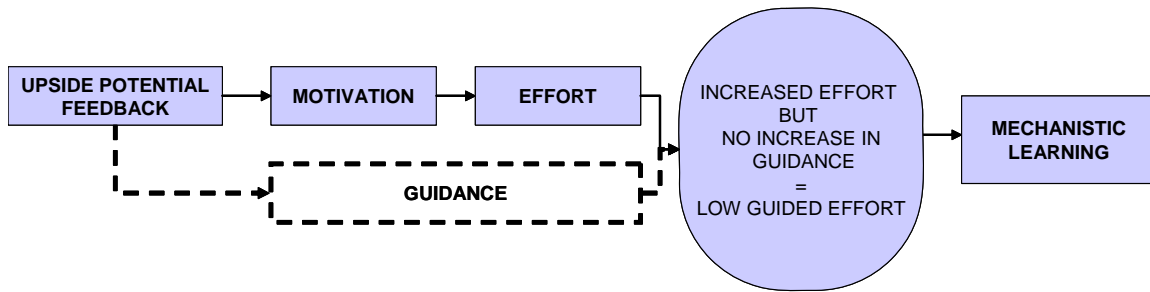
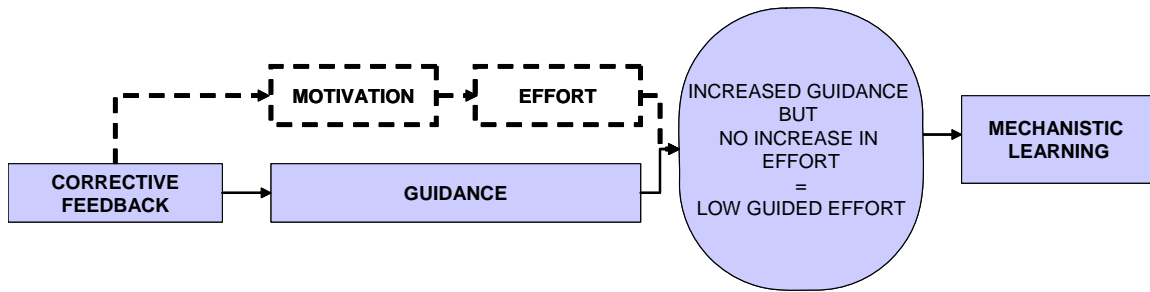


Figure 2: Theoretical Framework Relating Feedback to Learning and Evaluation (dotted lines indicate expectations of non-significant links)

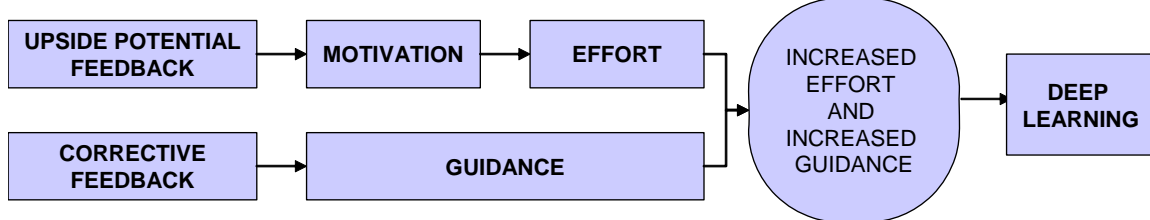
a. Effect of Upside Potential Feedback on Learning (Upside condition)



b. Effect of Corrective Feedback on Learning (Corrective condition)



c. Effect of Combining Upside Potential and Corrective Feedback on Learning (ALL condition)



d. Connecting DSS Design Characteristics, Deep Learning, and DSS Evaluation (equations 6.1-6.5)

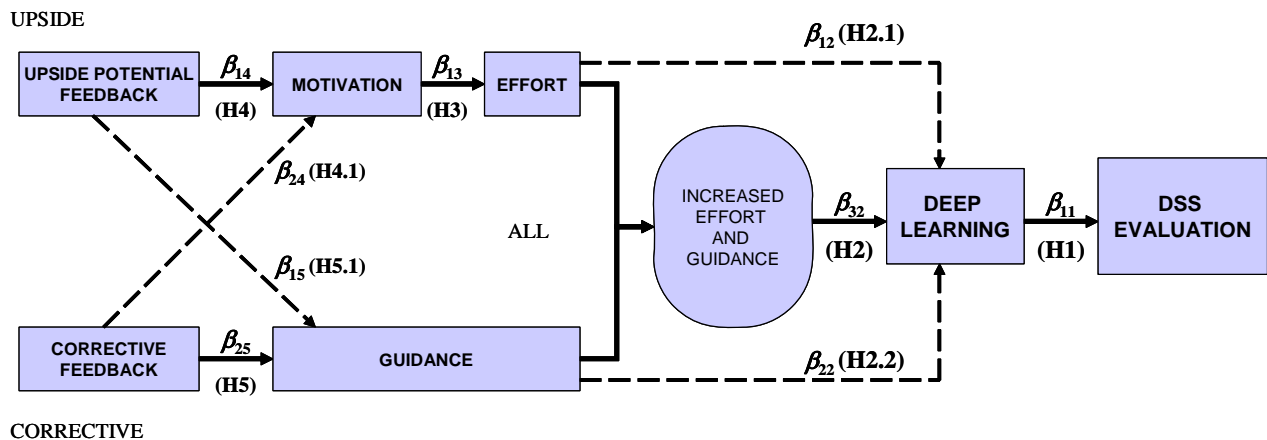


Figure 3: DSS Interface, Illustrating the Respondent Task

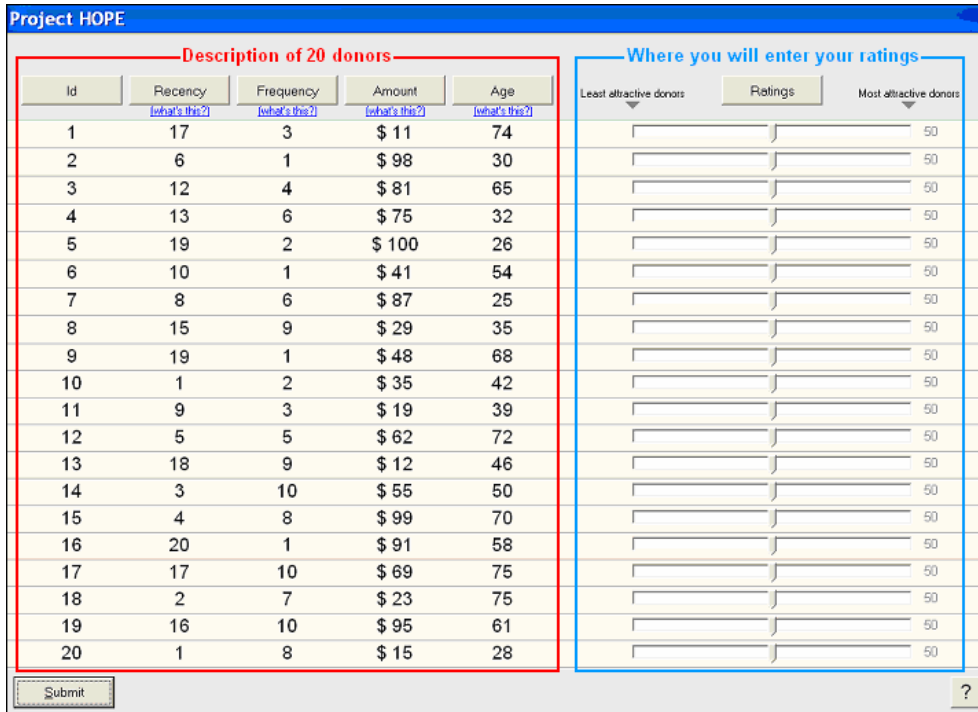
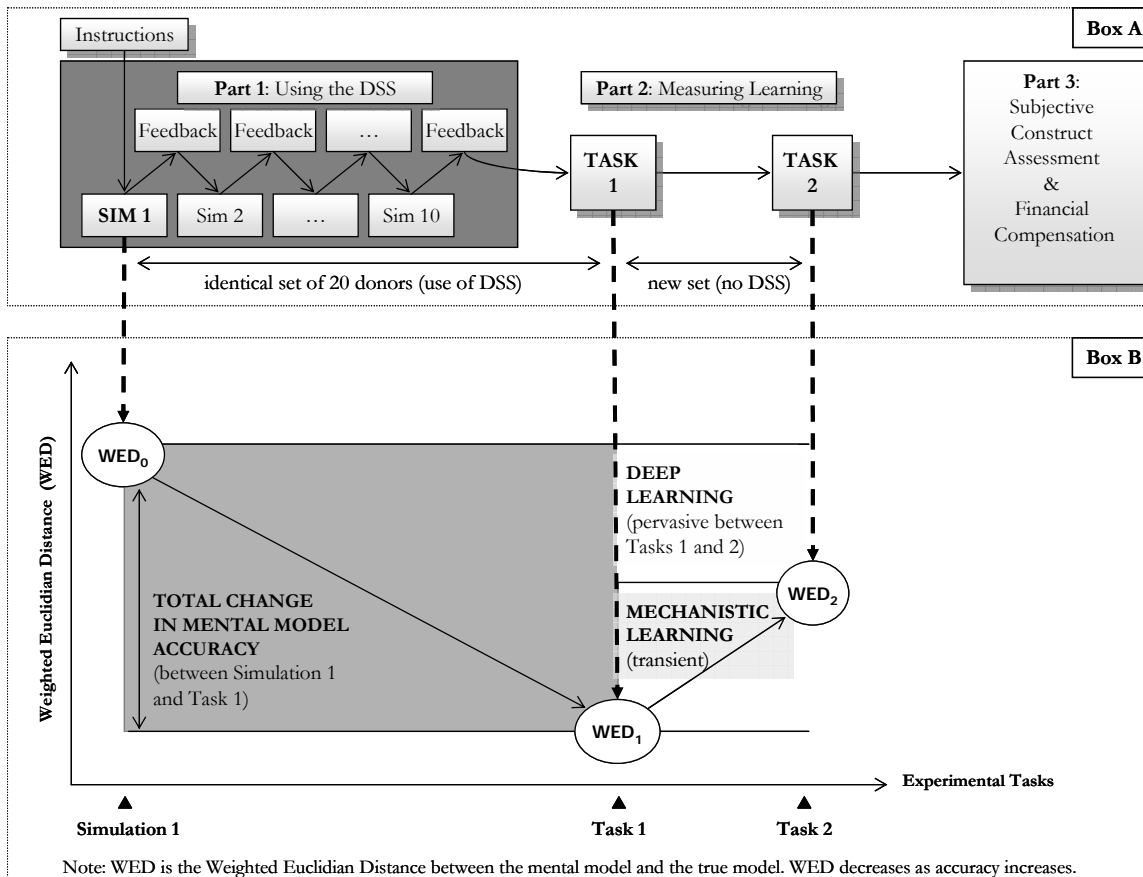


Figure 4: Experimental Sequence (Box A) and Measures of Learning (Box B) (dotted arrows indicate the mapping of tasks to measures)



Note: WED is the Weighted Euclidian Distance between the mental model and the true model. WED decreases as accuracy increases.

Table 1: Relationship between DSS Evaluation, Deep Learning, and Guided Effort

Part A: Effect of Deep Learning on DSS Evaluation

Variable	Coefficient	Beta	t-stat	Hypothesis
Intercept	β_{01}	2.383	2.71 ^a	
Deep learning	β_{11}	0.051	7.35 ^a	H1
Mechanistic Learning	β_{21}	-0.019	-1.34	
Monetary compensation	β_{31}	0.037	1.59	

Part B: Effect of Guided Effort on Deep Learning

Variable	Coefficient	Beta	t-stat	Hypotheses
Intercept	β_{02}	-3.96	-0.21	
Effort	β_{12}	-3.81	-0.80	H2.1
Guidance	β_{22}	-6.24	-1.12	H2.2
Effort×Guidance	β_{32}	3.14	2.49 ^b	H2

Part C: Effect of Feedback Type on Effort, Motivation, and Guidance

1. Effort

Variable	Coefficient	Beta	t-stat	Hypothesis
Intercept	β_{03}	2.03	6.41 ^a	
Motivation	β_{13}	0.57	7.34 ^a	H3

2. Motivation

Variable	Coefficient	Beta	t-stat	Hypotheses
Intercept	β_{04}	3.79	22.08 ^a	
UPSIDE FEEDBACK condition	β_{14}	0.59	3.17 ^a	H4
CORRECTIVE FEEDBACK condition	β_{24}	-0.27	-1.45	H4.1

3. Guidance

Variable	Coefficient	Beta	t-stat	Hypotheses
Intercept	β_{05}	2.82	12.55 ^a	
UPSIDE FEEDBACK condition	β_{15}	0.62	2.61 ^b	H5.1
CORRECTIVE FEEDBACK condition	β_{25}	0.69	2.90 ^a	H5

Notes:

1. Significant at: ^a $p < 0.01$, ^b $p < 0.05$ (two-tailed)
2. Deep learning is the most significant driver of DSS evaluation, supports H1.
3. The interaction of effort and guidance significantly affects deep learning, supports H2.
4. Effort increases with motivation, supports H3.
5. Motivation increases when upside potential feedback is provided, but not with corrective feedback, supports H4 and H4.1.
6. Guidance increases with both upside potential feedback and corrective feedback, supporting H5, but not supporting H5.1.

Table 2: Effect of Upside Potential Feedback and Corrective Feedback on Guided Effort¹

Condition	Mean		Coefficient	Beta	t-stat
	Guided Effort				
OUTCOME FEEDBACK condition	12.60 ²		η_1	-5.31	-3.03 ^a
UPSIDE FEEDBACK condition	15.28 ³		η_2	-2.63	-2.09 ^b
CORRECTIVE FEEDBACK condition	14.93 ³		η_3	-2.98	-2.37 ^b
ALL condition	17.91				

Notes:

1. We test here whether guided effort is greater in the ALL condition than in each of the other conditions.
2. Mean guided effort in this condition is significantly less than that for the ALL condition at ^a $p < 0.01$ (one-tailed).
3. Mean guided effort in this condition is significantly less than that for the ALL condition at ^b $p < 0.05$ (one-tailed).

Table 3: Effect of Upside Potential Feedback and Corrective Feedback on Deep Learning¹

Condition	N	A. Is deep learning significantly different from 0?		B. Is deep learning significantly less than that in ALL condition?	
		Mean		Difference	t-stat
		Deep Learning	t-stat		
OUTCOME FEEDBACK condition	14	-1.47	-0.43	-15.19	-3.34 ^a
UPSIDE FEEDBACK condition	15	6.65 ²	2.06 ^b	-7.06	-1.66 ^c
CORRECTIVE FEEDBACK condition	16	3.79 ³	1.19	-9.92	-2.34 ^b
ALL condition	16	13.71 ⁴	4.33 ^a		

Notes:

1. We test here whether (A) deep learning is significantly different from 0 in each of the conditions, and (B) deep learning is significantly less than that in the ALL condition.
2. Deep learning in UPSIDE FEEDBACK condition is significantly different from 0 at ^b $p < 0.05$, and is significantly less than that in the ALL condition at: ^c $p < 0.10$ (both are one tailed tests).
3. Deep learning in CORRECTIVE FEEDBACK condition is not significantly different from 0, and is significantly less than that in the ALL condition at: ^b $p < 0.05$ (both are one tailed tests).
4. Deep learning in this condition is significantly different from 0 at: ^a $p < 0.01$ (one tailed).
5. Indicates that deep learning is significantly different from zero in the ALL and the UPSIDE FEEDBACK conditions, and not significant in other conditions. Also, deep learning in the ALL condition is significantly greater than that in other conditions.

Table 4: Effect of Upside Potential Feedback and Corrective Feedback on Mechanistic Learning¹

Condition	N	Mean Mechanistic Learning	
		Learning	t-stat
OUTCOME FEEDBACK condition	14	2.85	1.47
UPSIDE FEEDBACK condition	15	3.70 ²	1.97 ^b
CORRECTIVE FEEDBACK condition	16	7.57 ³	4.16 ^a
ALL condition	16	2.19	1.21

Notes:

1. We test here whether mechanistic learning is significantly different from 0 in each of the conditions.
2. Mechanistic learning in UPSIDE FEEDBACK condition is significantly different from 0 at: ^b $p < 0.05$ (one-tailed).
3. Mechanistic learning in CORRECTIVE FEEDBACK condition is significantly different from 0 at: ^a $p < 0.01$ (one-tailed).
4. Indicates that mechanistic learning is not significant in the ALL condition, is highly significant in CORRECTIVE FEEDBACK condition, and significant in the UPSIDE FEEDBACK condition.