How Feedback Can Improve Managerial Evaluations of Model-based Marketing Decision Support Systems

Ujwal Kayande, Arnaud de Bruyn, Gary Lilien,
Arvind Rangaswamy and Gerrit Van Bruggen

ERIM Report Series *Research in Management*	
ERIM Report Series reference number	ERS-2006-039-MKT
Publication	July 2006
Number of pages	39
Persistent paper URL	
Email address corresponding author	gbruggen@rsm.nl
Address	Erasmus Research Institute of Management (ERIM) RSM Erasmus University / Erasmus School of Economics Erasmus Universiteit Rotterdam P.O.Box 1738 3000 DR Rotterdam, The Netherlands Phone: + 31 10 408 1182 Fax: + 31 10 408 9640 Email: info@erim.eur.nl Internet: www.erim.eur.nl

Bibliographic data and classifications of all the ERIM reports are also available on the ERIM website:
www.erim.eur.nl

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT

REPORT SERIES
RESEARCH IN MANAGEMENT

ABSTRACT AND KEYWORDS	
Abstract	Marketing managers often provide much poorer evaluations of model-based marketing decision support systems (MDSSs) than are warranted by the objective performance of those systems. We show that a reason for this discrepant evaluation may be that MDSSs are often not designed to help users understand and internalize the underlying factors driving the MDSS results and related recommendations. Thus, there is likely to be a gap between a marketing manager's mental model and the decision model embedded in the MDSS. We suggest that this gap is an important reason for the poor subjective evaluations of MDSSs, even when the MDSSs are of high objective quality, ultimately resulting in unreasonably low levels of MDSS adoption and use. We propose that to have impact, an MDSS should not only be of high objective quality, but should also help reduce any mental model-MDSS model gap. We evaluate two design characteristics that together lead model-users to update their mental models and reduce the mental model-MDSS gap, resulting in better MDSS evaluations: providing feedback on the upside potential for performance improvement and providing specific suggestions for corrective actions to better align the user's mental model with the MDSS. We hypothesize that, in tandem, these two types of MDSS feedback induce marketing managers to update their mental models, a process we call deep learning, whereas individually, these two types of feedback will have much smaller effects on deep learning. We validate our framework in an experimental setting, using a realistic MDSS in the context of a direct marketing decision problem. We then discuss how our findings can lead to design improvements and better returns on investments in MDSSs such as CRM systems, Revenue Management systems, pricing decision support systems, and the like.
Free Keywords	Marketing Decision Support Systems, Marketing Decision Models, Marketing Information Systems, Feedback, Learning
Availability	
Classifications	

How Feedback Can Improve Managerial Evaluations of Model-based Marketing Decision Support Systems

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

*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 H. 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.

How Feedback Can Improve Managerial Evaluations of Model-based Marketing Decision Support Systems

Abstract

Marketing managers often provide much poorer evaluations of model-based marketing decision support systems (MDSSs) than are warranted by the objective performance of those systems. We show that a reason for this discrepant evaluation may be that MDSSs are often not designed to help users understand and internalize the underlying factors driving the MDSS results and related recommendations. Thus, there is likely to be a gap between a marketing manager's mental model and the decision model embedded in the MDSS. We suggest that this gap is an important reason for the poor subjective evaluations of MDSSs, even when the MDSSs are of high objective quality, ultimately resulting in unreasonably low levels of MDSS adoption and use. We propose that to have impact, an MDSS should not only be of high objective quality, but should also help reduce any mental model-MDSS model gap. We evaluate two design characteristics that together lead model-users to update their mental models and reduce the mental model-MDSS gap, resulting in better MDSS evaluations: providing feedback on the *upside potential* for performance improvement and providing specific suggestions for *corrective actions* to better align the user's mental model with the MDSS. We hypothesize that, in tandem, these two types of MDSS feedback induce marketing managers to update their mental models, a process we call deep learning, whereas individually, these two types of feedback will have much smaller effects on deep learning. We validate our framework in an experimental setting, using a realistic MDSS in the context of a direct marketing decision problem. We then discuss how our findings can lead to design improvements and better returns on investments in MDSSs such as CRM systems, Revenue Management systems, pricing decision support systems, and the like.

Key words: Marketing Decision Support Systems, Marketing Decision Models, Marketing Information Systems, Feedback, Learning

1.0 Introduction

Technological and modeling advances have dramatically increased the availability and quality of model-based marketing decision support systems (Wierenga, Van Bruggen, and Staelin 1999). Many such systems (e.g., database marketing systems, customer relationship management systems, marketing dashboards, pricing decision support systems, sales territory alignment systems) are based on decision models, which have been shown to improve the *objective* quality of marketing decision making (e.g., McIntyre 1982, Lodish et al. 1988, Hoch and Schkade 1996, Silva-Risso, Bucklin, and Morrison 1999, Eliashberg et al. 2000, Zoltners and Sinha 2005, Divakar, Ratchford, and Shankar 2005), thus improving managerial and organizational performance. However, such decision models and systems are not adopted by marketers to the extent that would be expected based on their objective quality (CMO Council Report, June 2006). For example, retail industry analysts report that retailers have been slow to adopt pricing decision models that are known to improve retail performance (Reda 2002, 2003). Sullivan (2005) reports that only 5 to 6% of retailers use price-optimization models, while most prefer to use their gut-feel for making pricing decisions. As a consequence, "actual retail prices observed over time may differ greatly from model-recommended courses of action" (Nijs, Srinivasan, and Pauwels 2006). While there are likely many reasons for the slow and limited adoption of such model-based systems (interface complexity, high costs, etc.), the likelihood of adoption of any such system is negatively affected when those affected do not recognize its value. And researchers have indeed found that marketing managers often do not recognize the value of decision models or support systems, leading to increased resistance and impeding system adoption (e.g., McIntyre 1982, Davis 1989, Van Bruggen et al. 1996).

We investigate why subjective evaluations of model-based marketing decision support systems (hereafter called MDSSs) are often at odds with their known objective quality, and how one can improve those evaluations so that MDSSs are more likely to be adopted and used. We propose that marketing managers recognize the value of an MDSS when the MDSS is able to bring the marketing manager's mental model of the decision environment closer to the decision model embedded in the MDSS (hereafter referred to as the *MDSS model*), i.e., by helping marketing managers update their mental models. By mental models we mean an individual's mental representation of the relative importance of the various determinants of market or individual behavior (see, for example, Day and Nedungadi 1994). We empirically demonstrate the effectiveness of two feedback mechanisms in reducing the incongruence between mental models and decision models, leading to better managerial evaluations of MDSSs.

Our 3-Gap framework (Figure 1) summarizes our perspective on the MDSS evaluation problem (discussed in detail in Section 2.1). We hypothesize that the magnitude of the gaps between three models of the decision environment – the marketing manager's mental model, the MDSS model, and the unknown true model (which generates data in the real world, but is only partially observed ex-post) – determines the marketing managers' decisions, the consequent outcomes, and MDSS evaluations. As an example, consider the Recency-Frequency-Monetary Value model, a popular decision model used to select customers for direct marketing campaigns. This model can be parameterized by managers *a priori* (i.e., subjectively, based on their experience) or by using a variety of statistical techniques to determine the weights to be placed on each of the three factors in a customer selection task (Bult and Wansbeek 1995). To provide high quality decision support, the gap between the MDSS model and the true model must be small (Gap 2 in Figure 1), which is the aim of researchers who develop better statistical techniques in marketing. A large marketing and management science literature is focused on reducing Gap 2.

However, marketing managers' actions--reflecting their subjective weights noted above--often differ significantly from the recommendations of decision models (Van Bruggen et al. 1996). Such differences induce a form of dissonance that can lead to discomfort with the MDSS and its recommendations. For an MDSS to be favorably evaluated (and hence, ceteris paribus, more likely to be adopted and used), we hypothesize that the MDSS should help marketing managers *internalize* the reasoning of the high-quality MDSS model, because that internalization will make users feel comfortable with the MDSS recommendations. This internalization will reduce the gap between the MDSS model and the mental model (Gap 1 in Figure 1). By definition, if Gap 2 is small, reducing Gap 1 helps users form a better understanding of the mechanism that drives the real world observations (reducing Gap 3). Therefore, what is required for improving subjective evaluations of MDSSs is a mechanism to help marketing managers update their mental models to be more closely aligned with the MDSS model.

In this paper, we show that a dual-feedback MDSS, one that incorporates feedback *both* about upside potential (i.e., how much more can be gained by internalizing the MDSS model) *and* feedback on corrective actions (i.e., guidance on how the marketing manager's mental model should be corrected) produces significant mental model updating while single feedback MDSSs produce only marginal mental model updating. That mental model updating, in turn, leads to better subjective MDSS evaluations than when little or no mental model updating occurs, i.e. for MDSSs with only one type or no feedback.

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 MDSS model influences MDSS evaluation. We also review previous research on how to improve the quality of decision support systems, and show that past research does not address the fundamental source of the MDSS evaluation problem. Then we develop a model of how dual feedback on upside potential and corrective actions helps in updating 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, MDSS Evaluation, and MDSS Design

2.1 The Effects of Mental Model Changes on MDSS Evaluation

Users of high quality MDSSs often perform better by simply implementing MDSS recommendations in a mechanistic fashion, but without necessarily understanding the rationale behind the MDSS's recommendations. We propose that this lack of understanding reflects a fundamental gap between the MDSS model and the user's mental model of the decision environment. If the gap between the MDSS model and the mental model (Gap 1 in Figure 1) is large, the MDSS model's recommended course of action and that implied by the user's mental model are likely to vary, resulting in conflicting information about the appropriate decision. Such conflicting information results in high uncertainty (Einhorn and Hogarth 1980). In line with research on risk-adjusted preference (Keeney and Raiffa 1976, Roberts and Urban 1988, Rust et al. 1999), the objective quality of the MDSS is then likely to be discounted by a risk-averse marketing manager to account for the high uncertainty, leading to poor subjective evaluations. Therefore, one potential source of the MDSS evaluation problem lies in the inability of current MDSS designs to change decision makers' mental models, i.e., to close the gap between the user's mental model and the MDSS model (Gap 1 in Figure 1). As a consequence, we suggest that the greater the change of the mental model in the direction of the MDSS model, the better is the evaluation of the MDSS that is used to effect the change (formalized later as Hypothesis H1 in Section 2.3). Note that this expected effect assumes that the MDSS 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). This assumption about a small Gap 2, consistent with much of the marketing science literature (see Eisenstein and Lodish, 2002, for example) is one we make throughout and control for explicitly in our empirical work below.

As noted earlier, much of the work in marketing science and operations research 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 decision support systems have explored factors that lead to greater system usage and/or better performance. These factors include "task-technology" fit (Lim and Benbasat 2000), tabular vs. graphical presentation format and color (e.g., Benbasat and Dexter 1985), fit between cognitive style and presentation format (Ramaprasad 1987), accessibility (e.g., Mawhinney and Lederer 1990), adaptability/flexibility (e.g., Udo and Davis 1992), perceived ease of use and usefulness (Kim and Malhotra 2005), information quality and systems quality (e.g., Delone and McLean 1992), and restrictiveness of the system guidance (Silver 1990).

Although the DSS design literature helps identify the drivers of MDSS usage, there is much evidence that users do not appear to value MDSSs (Wierenga, Van Bruggen, and Staelin 1999; Lilien et al. 2004) and that users are slow to adopt high quality MDSSs (Montgomery 2005). Pauwels (2004) suggests that feedback loops are essential to improve the effectiveness of MDSSs. However, it appears that researchers have paid little attention to the different types of feedback and the central role that those types of feedback play in reducing the gap between mental model and the MDSS model, therefore enhancing MDSS acceptability.

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

To be recognized as valuable and providing an (expected) return on investment, MDSS design must incorporate characteristics that induce learning, i.e., a change in the user's mental model that improves his/her performance. The change in mental models could be of at least two types – (i) a relatively permanent deep change, or (ii) a transient change that disappears when the MDSS is unavailable. We formally define these changes as follows:

> 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, pg. 224).
>
> Shallow 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" (Goodman 1998 pg. 224; also see Kluger and Denisi 1996, pg. 278).

A permanent change will tangibly reduce the uncertainty about the MDSS recommendations, i.e., reduce (cognitive) dissonance, resulting in improved MDSS evaluations, while a transient change, i.e., shallow learning, will not influence a marketing manager's uncertainty about the MDSS recommendations. Our main interest, therefore, is in how deep learning comes about because we expect deep learning to affect MDSS evaluation. Goodman, Wood, and Hendrickx (2004) suggest that deep learning, i.e., a relatively permanent change in mental models, is most likely to occur when individuals (i) are motivated to, and actually exert effort, to change their mental model, and (ii) are provided guidance on how to modify that mental model. We formalize the joint effect of effort and guidance 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 effort and guidance for users to significantly update their mental models. Next we describe how each type of feedback individually influences effort and guidance, and therefore affects learning.

2.2.1 Effects of Upside Potential Feedback on Effort and Learning. Information about upside potential addresses the following question – how much better can the marketing manager perform relative to current performance? For example, the sales module of the Siebel CRM system provides a salesperson with information such as the sales achieved by the best performing salesperson, or the sales levels in the best performing sales territory, providing a proxy for upside potential. Upside potential feedback helps managers set specific (and challenging) goals, resulting in increased individual and collective effort to achieve those goals (Locke et al. 1981, Bandura 1997). Chenoweth et al. (2004) show that users of decision support systems are willing to put in more effort to learn complex models when they know the upside potential of using complex models.

However, several researchers have shown that while effort increases with more challenging goals, that increased effort does not necessarily lead to deep learning because such goal-oriented behavior can direct the individuals' attention on the self, rather than on the task (Wood et al. 1990). As a result, task-learning processes are not activated (Kluger and Denisi 1996), leading to only shallow learning and poorer out-of-task performance. Upside potential feedback helps the marketing manager set specific and challenging goals (for e.g., match the best salesperson's performance), but does not provide the feedback necessary to learn how to perform better. Earley et al. (1990) found that the relationship of goal-setting to learning and performance is greatly enhanced when individuals are provided with feedback about how to correct their strategies.

In summary, upside potential feedback will induce increased effort, but may direct attention away from task-learning processes, resulting in *increased effort without appropriate guidance*. So if upside potential feedback were to be combined with feedback that focuses attention on the task, we would expect marketing managers to exert the increased levels of effort and obtain the guidance necessary to obtain a significant level of deep learning. The effect of upside potential feedback on learning is summarized in Figure 2a.

2.2.2 Effects of Corrective Feedback on Guidance and Learning. Corrective feedback, also called process feedback (Earley et al. 1990), can improve decision making, particularly in complex tasks, by increasing attention to task-learning processes and improving the 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 (Goodman 1998, Denisis and Kluger 2000, Goodman, Wood, and Hendrickx 2004). Thus, corrective feedback might only lead to shallow learning, not deep learning, because individuals adjust behavior by using the feedback rather than by focusing on understanding the task. Atkins et al. (2002) find that if feedback is presented such that it is very easy for decision-makers to derive guidelines for action, that feedback may cause decision-makers to derive these guidelines from the information presented 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. Goodman, Wood, and Hendrickx (2004) state that, "Essentially, feedback does the work for the performers, making it seemingly unnecessary for them to engage in the exploration, information-processing, and recall activities essential for learning." (pg. 249).

In summary, corrective feedback directs attention to the task and task-learning process but also leads to less exploration and less effort, which are both 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 The combined effects of upside potential and corrective feedback are summarized in Figure 2c.

2.3 Theoretical Model and Hypotheses

Figure 3 provides a summary of our theoretical framework, relating the two types of feedback in an MDSS to effort and guidance, deep learning, and MDSS evaluation, as well as a summary of the key results obtained from an empirical test, developed later. Our process model is as follows:

(M1) $MDSS\ Evaluation = \beta_{01} + \beta_{11} \cdot DeepLearning + \varepsilon_1$

(M2) $DeepLearning = \beta_{02} + \beta_{12} \cdot Effort + \beta_{22} \cdot Guidance + \beta_{32} \cdot \left(Effort \times Guidance\right) + \varepsilon_2$

(M3) $Effort = \beta_{03} + \beta_{13} \cdot \text{UPSIDE POTENTIAL FEEDBACK} + \beta_{23} \cdot \text{CORRECTIVE FEEDBACK} + \varepsilon_3$

(M4) $Guidance = \beta_{04} + \beta_{14} \cdot \text{UPSIDE POTENTIAL FEEDBACK} + \beta_{24} \cdot \text{CORRECTIVE FEEDBACK} + \varepsilon_4$

Our first hypothesis relates MDSS evaluation to the relatively permanent change in mental models:

H1 An increase in deep learning leads users to provide more favorable evaluations of the MDSS. Therefore, we expect $\beta_{11} > 0$.

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

H2 The interaction of effort and guidance will have a positive effect on deep learning. Therefore, we expect $\beta_{32} > 0$.

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. Therefore, we expect $\beta_{12} = 0$.

H2.2 An increase in guidance without effort does not lead to deep learning. Therefore, we expect $\beta_{22} = 0$.

Per our discussion in Section 2.2.1 and 2.2.2, effort increases when a marketing manager is provided with feedback on upside potential, but effort does not increase with corrective feedback. Accordingly, we expect $\beta_{13} > 0$, but $\beta_{23} = 0$. On the other hand, guidance is influenced by the presence of corrective feedback, but not by feedback on upside potential. Accordingly, we expect β_{24} to be > 0, but $\beta_{14} = 0$. Our expectations about these parameters serve as manipulation checks in our empirical analysis. Models M1-M4 comprise a test of the process model proposed in Figure 3.

3.0 Empirical Study

We designed our empirical study to incorporate a challenging set of criteria. We sought:

(i) a decision environment comparable in realism to those faced by marketing managers,
(ii) a decision environment sufficiently complex so that marketing managers would benefit from using a MDSS, but not so complex as to be outside the skill range of our research participants,
(iii) an MDSS 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 MDSS with each of the types of feedback (upside potential feedback and corrective feedback), both individually and jointly,
(vi) a task that would allow us to measure deep and shallow learning unobtrusively, and
(vii) a task that would allow us to measure the process variables of interest (effort and guidance).

Criteria (i)-(iv) relate to design of the overall context of the study, while criteria (v)-(vii) relate to the 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 MDSS users about the relationship between the decisions they make and the corresponding market response. Tversky and Kahnemann (1987, p. 90) suggest that such feedback is often lacking in real world situations because, (1) outcomes are commonly delayed and not easily attributable to a particular decision/action; (2) variability in the environment degrades the reliability of the feedback, especially where outcomes of low probability are involved, (3) there is often no information about what the outcome would have been if another decision had been taken; and (4) important decisions are often unique and provide little opportunity for learning. 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 very difficult to effect 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 marketing 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. MDSSs 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 MDSS 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. This context satisfies the 'realism' criterion (i).

3.1.1 Decision Environment: To satisfy criterion (ii), we sought a decision environment that would be sufficiently complex, but not outside the skill range of our participants. A direct marketing decision environment that is complex enough to require the use of an MDSS to select customers involves a large database (200,000 in our case) of (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*. The first three characteristics are often used by direct marketing firms in targeting models, typically referred to as RFM (Recency-Frequency-Monetary Value) models. We added age to the model to increase the complexity of the decision environment. In line with the most common models of purchase probability (Agresti, 2002) we modeled the probability that a particular donor would make a donation, if solicited, by a logit function as follows:

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

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

(2) $$X = \beta_0 + \left(\beta_1 \times recency\right) + \left(\beta_2 \times frequency\right) + \left(\beta_3 \times amount\right) + \left(\beta_4 \times age\right)$$

The parameters of the "true" data generating model were $\beta = \{20, -20, 40, 10, 30\}$. In line with these parameters, we informed participants at the start that donors were more likely to donate if they had donated more recently, donated more frequently in the past five years, donated greater amounts, and if they were older. (To easily operationalize the weights for β above, we rescaled the variables to account for differences in units of measurement.).

We generated a database of customers to satisfy two criteria: first, the probabilities of donation in our database should be similar to those observed in actual not-for-profit databases and second, the characteristics should be generated so that the each donor could be described on a 0 to 100 attractiveness scale with an average at the midpoint, so that we could subsequently ask participants to rate each donor on the same scale. To satisfy these criteria, we generated donor characteristics from uniform distributions between 0 and 1, (accounting for the rescaling noted above) independent of one another, and incorporated these values within the functional form of the logit function 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 of donation varied between 0.67% and 50% with an average of 7.6%. Although average response rates, in practice, vary widely for different charities, 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) for the design of our empirical study, we wanted an MDSS model that would be sufficiently close to the true data-generating model. Therefore, we designed Gap 2 to be small by constructing an MDSS 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 random error to the true model in equation (1) so that the actual donations from donors could be predicted only approximately. In expectation, the true model and MDSS model were identical.

3.1.2 Calibrating Participants' Mental Models: To satisfy criterion (iv), we devised an unobtrusive and unbiased mechanism to calibrate the participant's mental model. We designed our study so that participants made a sufficient number of decisions at one time to allow us to unobtrusively calibrate their mental models, similar to Kunreuther's (1969) work on estimating managerial decision coefficients. Such a process allowed us to infer their mental model without having to actually ask the participants to reflect on their mental processes. The latter method suffers from significant biases (Rowe and Cook 1995) that 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 0 to 100-point sliding scale. These 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 4. This rating process (after rescaling and sorting) corresponds to the typical scoring mechanism that emerges in most direct marketing DSSs (see David Shephard Associates, 1999) who sort a prospect list for an offer from most to least attractive.

To obtain an unobtrusive measure of the participant's mental model, we related their donor ratings statistically 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 for independent estimation of each weight. While a fractional factorial design is typically the design choice in such cases, the number of profiles that our participants would have to rate made that approach infeasible. Therefore, we generated donors' characteristics (recency, frequency, etc.) 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 participant's rating sample, until no inter-item correlation was higher than 0.15.

Once a participant submitted his or her ratings, we estimated a linear regression model to determine the implicit weights (β_0', β_1'… β_4') that 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 solicit. We told participants that each solicitation costs $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'$), mental model parameters would be equal to true parameters ($\beta = \beta'$), and the solicitation strategy would be optimal.

To assist the participants in their decision-making, we provided them an MDSS to select attractive customers from a database. We addressed the issue of 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 Task 1 and Task 2 (described more fully in the next section), in addition to a $15 participation payment. .

3.2 Design of the experiment (manipulations and measurements)

We summarize the sequence of steps in our experiment in Box A of Figure 5. 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 MDSS. In Part 1 of the study, we asked participants to rate the same 20 donors in each of ten simulations. The participants had access to an MDSS to help them determine the best possible ratings. In each simulation, participants rated the 20 donors, submitted those ratings to the MDSS simulator, and obtained the MDSS's prediction 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 MDSS 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 MDSS to reflect the two types of feedback under study. We varied upside potential feedback at two levels (present or absent) and corrective feedback at two levels (present or absent), for a design with four cells. Both types of feedback were absent in the control condition, in which we only provided participants with information about the expected performance of the donor ratings. Our four conditions were:

1. "CONTROL CONDITION": The participant was only informed of the expected performance of the donor ratings. For example:

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

2. "UPSIDE POTENTIAL 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 MDSS predicts that a marketing campaign based on your ratings would generate $76,654 in revenue.
   The MDSS 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 MDSS model. For example:

The MDSS 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 potential, and corrective actions (conditions 1-3 above), in that order. For example:

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

(Note that all feedback was dynamically generated and customized, based on the actual ratings provided by each participant).

Participants were randomly assigned to one of the four feedback 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 their Task 1 ratings.

3.2.2 Part 2 of the Study: Measuring Learning. Our goal in Part 2 is to measure deep and shallow learning. Per our definition of deep learning, we sought a measure of mental model change that survives when feedback is removed and the task is changed. 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 an MDSS, i.e., there was no feedback for this set of ratings.

We then constructed a measure of mental model accuracy – the distance between the true model (which in our study is identical, on average, to the MDSS model) and the mental model. We sought a measure of mental model accuracy that reflects the participant's ability to judge the relative importance of those factors. A measure that satisfies this criterion is as follows:

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

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 MDSS model, then, on average, Gap 3 is identical to Gap 1 and $\omega_j = \beta_j$. (We obtained substantially equivalent empirical results with an equally weighted, simple Euclidean distance measure, suggesting that our results are not sensitive to our choice of this metric).

We provide a graphical explanation of our learning measures in Box B of Figure 5, 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_1 is a measure of the change in mental model accuracy that is due to the participant's use of the MDSS. Part of this change is a result of an internalization of the MDSS model, i.e., deep learning, and part is a transient change that will disappear with the removal of the feedback, i.e., shallow learning. Asking participants to complete Task 2 gives us the ability to independently calibrate each part, 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) $DL = \left(WED_0 - WED_2\right)$.

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 participant 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 shallow learning (SL) as:

(5) $SL = \left(WED_2 - WED_1\right)$.

While Parts 1 and 2 of our study allowed us to manipulate feedback provided by the MDSS and measure participants' mental models, we also needed to measure process variables and MDSS evaluation, which are described next.

3.2.3 Part 3 of the Study: Subjective Construct Assessment. The main process variables of interest here were 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 effort (four items adopted from Lilien et al. 2004: "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"), and guidance ("The MDSS gave clear guidance on how I could do better"). As a proxy measure of effort, we also recorded the time spent by each participant on the simulations and the two tasks. After being informed of their results, participants evaluated the MDSS ("I would definitely recommend an MDSS like the one I had available to direct marketers"). All items were 5-point Likert scale questions, with 1=completely disagree, and 5=completely agree. Additionally, participants were asked to respond to multi-item scales on perceived usefulness of the MDSS, perceived ease of use of the MDSS, perceived enjoyment of the task, decision confidence, and decision style.

3.3 Sample and Experimental Procedure

We sought participants who would be appropriate surrogates for direct marketing managers. Such participants had to have well-formed mental models of the drivers of donor attractiveness prior to the experimental study. Hence, we recruited 81 MBA students at a large northeastern U.S. university, who had been exposed to RFM models in at least one of their courses, to assume the role of the direct marketing fundraising manager. We randomly assigned participants to one of the four conditions. We sought a measure to determine whether our participants' mental models were well-formed (i.e., not random). One such measure is the R–Square associated with the calibration of each participant's initial mental model (simulation 1). The R-Square measure captures the consistency of the model used by the participant to rate each donor, indicating whether the initial mental model was well-formed (high R-Square) or poorly-formed (low R-Square). The R-Square for our participants ranged from 0.13 to 0.99, with a tri-modal distribution. A group of three participants had R-Squares between 0.13 and 0.20, a second group (6 participants) between 0.34 and 0.49, and a final group (72 participants) between 0.58 and 0.99. Based on the structure of the empirical distribution, we classified those participants with R-Squares under 0.58 as having ill formed mental models and excluded them, leaving us with 72 participants for analysis. The final sample resulted in 16-20 participants per condition. Participants earned between \$23 and \$46, with the average payment being \$37. They took an average of 35 minutes to complete the two tasks, with times ranging from 14 to 77 minutes.

3.4 Analysis

Figure 3 and Models (M1)-(M4) sketch the process model of the effect of MDSS design characteristics on deep learning and MDSS evaluation. There are several sub-models embedded within this framework, with errors that are likely to be correlated. For the empirical analysis, we included two additional terms in Model M1 to control for shallow learning and their individual financial compensation, as follows:

(6) $MDSS\ Evaluation = \beta_{01} + \beta_{11} \cdot DeepLearning + \beta_{21} \cdot ShallowLearning + \beta_{31} \cdot Financial\ compensation + \varepsilon_1$.

We estimated the parameters of the four models simultaneously with a full information maximum likelihood (FIML) routine in SAS. This procedure simultaneously estimates all models in a system, assuming correlated errors across those models.

3.5 Results

Results of estimating models (M1)-(M4) are shown in Table 1. We hypothesized that users' evaluations of the MDSS depend on the extent to which they had internalized the MDSS model (Hypothesis H1; Part A of Table 1). We find strong support for H1 (β_{11}= 0.045, p<0.01). We also find that shallow learning is not a significant driver of MDSS evaluation (β_{21}= -0.001, *ns*), further supporting our theory that MDSS evaluation depends on a *significant* updating of mental models. We also find that financial compensation is not a significant driver of evaluation (β_{31}= 0.01, *ns*). These results show that deep learning is a significant driver of MDSS evaluation, after controlling for shallow learning and financial 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 (β_{32}= 3.39, p<0.05)[1]. 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 (i.e., manipulation checks) are shown in Part C of Table 1. Effort was significantly increased by the presence of upside potential feedback (β_{13}= 0.29, p<0.01). In contrast, effort decreased with corrective feedback (supporting H4.1; β_{24}= -0.23, p<0.05). While we did not expect this significant negative effect, it is not entirely surprising considering the findings in the literature that corrective feedback leads to less inclination to exert effort (Atkins et al. 2002). Guidance, the other process variable, significantly increased in the presence of corrective feedback (β_{24}= 0.88, p<0.01). As expected, guidance was not affected by the presence of upside potential feedback (β_{14}= 0.21, *ns*). (We also estimated Models (M1)-(M4) using "time taken to complete the simulations and tasks" as a proxy measure of effort, with similar results).

Our Model (M2) results show that the interaction of effort and guidance produces deep learning. In other words, effort and guidance combine in complementary ways to help marketing managers update their mental models. But what is not clear is whether the combination of effort and guidance (i.e.,

[1] We controlled for the time taken by the participant in Model (M2), obtaining substantively similar results.

Effort×Guidance) is significantly greater in the ALL condition than in the other conditions. In Table 2, we show that the combination was significantly greater in the ALL condition than in the UPSIDE POTENTIAL FEEDBACK (*difference* =2.98, p<0.01, two-tailed) or CORRECTIVE FEEDBACK (*difference* = 1.76, p<0.05, two-tailed) conditions. These aggregate results are consistent with our process model (see Figure 3).

Table 3 (Column A) summarizes the 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 = 12.45, p<0.01, one -tailed). There is evidence that deep learning in the UPSIDE POTENTIAL FEEDBACK condition (mean = 5.34, p<0.05, one-tailed) and the CORRECTIVE FEEDBACK condition (mean = 5.40, p<0.05, one-tailed) are both significantly different from zero. However, deep learning in both these conditions is significantly less than that in the ALL condition, consistent with our hypotheses (results shown under Column B of Table 2).

Table 4 presents the analysis of whether significant shallow learning occurred in each of the conditions. We find that a significant level of shallow learning occurred in all the feedback conditions, but that it was significantly greater in the UPSIDE POTENTIAL FEEDBACK condition (mean = 4.97) and the CORRECTIVE FEEDBACK condition (mean = 6.33) than in the ALL condition (mean = 2.41). The presence of shallow learning and deep learning in each of the conditions indicates 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.

In Table 5, we report the financial performance and compensation of participants in each condition. Participants in the ALL condition performed significantly better than those in the other conditions. This result shows that participants provided with both types of feedback are more likely to perform better as well.

To validate our theoretical model, we sought to test it against two alternatives: a model specifying both direct and indirect effects of feedback on deep learning and an alternate model specification without process variables. We show the results of our validation in Table 6. To assess the first alternative, we compared the fit of our model system (AIC=1087.53) with that of a system with additional direct effects of the two types of feedback on deep learning (called the full model in Table 6; AIC=1090.73). To assess the second alternative, we compared the fit of our equation system (M1-M4) (AIC=1087.53 again) against an alternate model system without process variables (called the restricted model in Table 6; AIC=1100.70). Our proposed model passed both these validity tests.

4.0 Discussion, Future Research, and MDSS Design Implications

4.1 Discussion and Contributions

We proposed and empirically demonstrated that deep learning is crucial for marketing managers to form a favorable MDSS evaluation. Our study shows that an MDSS that provides upside potential feedback can motivate managers to perform better, resulting in greater effort. However, increased effort alone is not sufficient to generate deep learning; the MDSS must also provide clear guidance about how and why a modification of a mental model leads to a superior outcome. Our results also show that mere shallow learning does not lead to better evaluations of the MDSS, implying that MDSSs that offer no opportunity to understand their recommendations are likely to be poorly evaluated by users and hence, used less frequently. We found, hence, that a dual-feedback approach, combining upside potential and specific guidance is required to help managers internalize the MDSS model.

Our results show that MDSS feedback influences users' underlying learning process, which in turn, helps users internalize the relationship between decisions and outcomes. While much prior research has examined decision outcomes, our study enriches the story by showing why objectively superior MDSSs are often not evaluated more positively.

In summary, the main contributions of our research are: (1) the development of the "3-Gap framework to understand MDSS Evaluation" (Figure 1); (2) the specification of the role of deep learning (i.e., mental model changes) on managerial evaluations of MDSSs (Figure 3); and (3) the assessment of the individual and joint effects of two types of feedback, corrective and upside, on deep learning in the MDSS context. We have also conceptualized and employed an unobtrusive mechanism to assess MDSS users' mental models and their changes, an approach we hope other researchers will find useful.

4.2 Limitations and future research

Our work suggests several avenues for fruitful future research. In situations more realistic than our study, participants are likely to have much stronger *a priori* beliefs, which may be harder to overcome through MDSS-based learning. This is an empirical issue that is worth exploring, especially for understanding the role of MDSSs in bringing about enduring changes to how decisions are made in organizations. We also focused more on the evaluation process and did not include a direct measure of adoption in our study, a measure best obtained in a field study.

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 MDSS (Wierenga and Van Bruggen, 1997), and to explore how MDSSs influence different types of learning (e.g. learning about the MDSS 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 marketing manager might ignore MDSS recommendations and not learn because implementing MDSS recommendations might require an increased budget. Also, if a senior manager has very strongly held beliefs that are contrary to the MDSS model, a subordinate manager making decisions might simply ignore MDSS recommendations to be consistent with what the senior manager believes. An interesting research question is whether and how a MDSS can overcome such organizational constraints to learning. Finally, we assumed the functional form (linear additive) of our participants' mental models to be the same as that of the decision model. Functional forms might be non-linear and might vary across managers.

4.3 MDSS Design and Managerial Implications

Our research offers managerial insights about design aspects of MDSSs that increase the likelihood of their adoption and use, thereby increasing the return on investments in such systems. Specifically, for firms that develop and market decision support systems (e.g., Siebel, Salesforce.com, DemandTec), our results reinforce the importance of incorporating two types of feedback mechanisms that boost the value that managers find in an MDSS, thereby likely increasing its adoption and use. Similarly, organizations that invest in MDSSs can benefit from this research by understanding the importance of specific design characteristics they should ask their MDSS vendors to incorporate.

Many commercial MDSSs do not embed any feedback, offering recommendations without providing substantiating reasons. For example, many of the MDSS 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 MDSS may promote better decisions, and those decisions may lead to improved performance, users are unlikely to recognize the role the MDSS played, hampering future use. In the context of retail pricing, Montgomery (2005) suggests that retailers are likely to reject MDSSs that do not provide some intuition of the underlying decision model.

Many sophisticated commercial MDSSs do provide upside potential feedback. For example, salesforce.com (http://www.salesforce.com/products/analytics.jsp) provides a salesperson with information about the sales achieved by the top-performing salesperson. 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.

By designing an MDSS 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 MDSS adoption (because of its effect on MDSS evaluation). For example, in a 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 MDSS 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."

We note here that marketing managers who have internalized the MDSS model might not necessarily be able to articulate the weights or functional form of the MDSS model. For example, the foremen in the timber-cutting company do not need to "know" the dynamic programming algorithm; rather what is required is that they internalize the underlying rationale of the model in such a way that their actions are consistent with those recommended by the decision model.

Our research has the potential to help obtain a better return on investment for firms that have been heavily investing in MDSSs such as CRM (Customer Relationship Management) and marketing dashboards. 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). However, it is not clear that these investments have been successful. For example, Kale (2004) suggests that 60-80% of CRM investments produce returns much below expectations. Industry analysts suggest that line managers do not necessarily recognize the benefits of systems, leading to increased resistance to adopt, which eventually proves costly for the firm (Petouhoff 2006). Roach (2002, as quoted in the IDC report) suggests that sustained enhancement in performance results less from technological breakthroughs and more from substantial changes in the "*cerebral production function*" of the knowledge worker. Our results provide direct evidence that investments in MDSSs that help transform the mental models (i.e., cerebral production functions) of marketing knowledge workers are more likely to have substantial payoffs.

On net, we hope we have taken a significant step toward better understanding the mental model barriers to MDSS acceptance and how those barriers might be overcome.

REFERENCES

Agresti, Alan (2002), *Categorical Data Analysis*. John Wiley, 2nd Edition.

Atkins, Paul W.B, Robert E. Wood, Paul J. Rutgers (2002), "The Effects of Feedback Format on Dynamic Decision Making," *Organizational Behavior and Human Decision Processes*, 88(2) 587-604.

Balzer, William K., Lorne M. Sulsky, Leslie B. Hammer, K.E. Sumner (1992), "Task Information, Cognitive Information, or Functional Validity Information: Which Components of Cognitive Feedback Affect Performance?" *Organizational Behavior and Human Decision Processes,* 53(1) 35-54.

Bandura, Albert (1997), *Self-efficacy: The exercise of control.* New York: W.H. Freeman.

Benbasat, Izak, Albert S. Dexter (1985), "An Empirical Evaluation of Graphical and Color-enhanced Information Presentation," *Management Science,* 31(11) 1348–1364.

Bult, Jan R., Tom J. Wansbeek (1995), "Optimal Selection for Direct Mail," *Marketing Science*, 14(4), 378-394.

Chenoweth, Tim, Karen L. Dowling, R.obert D. St. Louis (2004), "Convincing DSS Users that Complex Models are Worth the Effort," *Decision Support Systems.* 37(1) 71-82.

CMO Council Report (2006), *Market Vigilance, Product Diligence*. June (www.cmocouncil.org).

David Shephard Associates (1999), *The New Direct Marketing*, 3d Edition New York: McGraw Hill

Davis, Fred D. (1989), "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly,* 13(3) 319-340.

Day, George, Prakash Nedungadi (1994), "Managerial Representations of Competitive Advantage," *Journal of Marketing* 58 (April): 31-44.

DeLone, William H., Ephraim R. McLean (1992), "Information Systems Success: The Quest for the Dependent Variable," *Information Systems Research,* 3(1) 60–95.

DeNisi, Angelo S., Avi N. Kluger (2000), "Feedback Effectiveness: Can 360-degree Appraisals be Improved?," *Academy of Management Executive, 14*(1), 129-139

Divakar, Suresh, Brian T. Ratchford, and Venkatesh Shankar (2005), "*CHAN4CAST*: A Multichannel, Multiregion Sales Forecasting Model and Decision Support System for Consumer Packaged Goods," *Marketing Science*, 24 (3), 334-350.

Earley, Christopher P., G.B. Northcraft; C. Lee; T.R. Lituchy (1990), "Impact of Process and Outcome Feedback on the Relation of Goal Setting to Task Performance," *The Academy of Management Journal*, 33 (1), 87-105.

Einhorn, Hillel J., Robin M. Hogarth, (1985) "Ambiguity and uncertainty in probabilistic inference," *Psychological Review,* 92(4) 433-461.

Eisenstein, Eric M., Leonard Lodish (2002), "Marketing Decision Support and Intelligent Systems: Precisely Worthwhile or Vaguely Worthless?" In *Handbook of Marketing*, edited by B. Weitz and R. Wensley, Sage Publications, Thousand Oaks, CA

Eliashberg, Josh, Jedid-Jah Jonker, Mohanbir S. Sawhney, Berend Wierenga (2000), "MOVIEMOD: An Implementable Decision-Support System for Prerelease Market Evaluation of Motion Pictures," *Marketing Science*, 19(3), pages 226-243

Goodman, Jodi S. (1998), "The Interactive Effects of Task and External Feedback on Practice Performance and Learning," *Organizational Behavior and Human Decision Processes*, 76(3) 223-252.

Goodman, Jodi S., Robert E. Wood and Margaretha Hendrickx (2004), "Feedback Specificity, Exploration, and Learning," *Journal of Applied Psychology,* 89(2), 248-262.

Hoch, Stephen J., David A. Schkade (1996), "A Psychological Approach to Decision Support Systems," *Management Science,* 42(1) 51–64.

IDC Report (2002), *The Financial Impact of Business Analytics* (www.idc.com).

Kale, Sudhir H. (2004), "CRM Failure and the Seven Deadly Sins," *Marketing Management*. 13 (5). 42-46.

Keeney, Ralph L., Howard Raiffa (1976), *Decision Making with Multiple Objectives: Preference and Value Tradeoffs*. John Wiley and Sons, New York.

Kim, Sung S., Naresh K. Malhotra (2005), "A Longitudinal Model of Continued IS Use: An Integrative View of Four Mechanisms Underlying Post-Adoption Phenomena," *Management Science,* 51(5) pp. 741-755.

Kluger, Avi N., Angelo Denisi (1996), "The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-analysis and a Preliminary Feedback Intervention Theory," *Psychological Bulletin,* 119, 254-284.

Kunreuther, Howard (1969), "Extensions of Bowman's theory on managerial decision-making," *Management Science*. 15(8) 415-439.

Lembersky, Mark R., Uli H. Chi (1986), "Weyerhaeuser Decision Simulator Improves Timber Profits," *Interfaces* 16(1) 6-15.

Lilien, Gary L., Arvind Rangaswamy, Gerrit H. Van Bruggen, Katrin Starke (2004), "DSS Effectiveness in Marketing Resource Allocation Decisions: Reality vs Perception," *Information Systems Research*. 15(3) 216-235.

Lim, Kai, Izak Benbasat (2002), "The Influence of Multimedia in Improving the Comprehension of Organizational Information," *Journal of Management Information Systems*, 19(1) 99-127.

Lister, Gwyneth J. 2001. *Building Your Direct Mail Program*, Jossey-Bass.

Locke, Edwin A., K.N. Shaw, L.M. Saari, Gary P. Latham (1981), "Goal Setting and Task Performance: 1969-1980," *Psychological Bulletin,* 90(1) 125-152.

Lodish, Leonard M., Ellen Curtis, Michael Ness, M. Kerry Simpson (1988), "Sales Force Sizing and Deployment using a Decision Calculus Model at Syntex Laboratories," *Interfaces* 18(1) 5-20.

Mawhinney, Charles H., Albert L. Lederer (1990), "A Study of Personal Computer Utilization by Managers," *Information Management,* 18(5) 243-253.

McIntyre, Shelby H. (1982), "An Experimental Study of the Impact of Judgment-based Marketing Models," *Management Science,* 28(1) 17-33.

Montgomery, Alan (2005), The implementation challenge of pricing decision support systems for retail managers, *Applied Stochastic Models in Business and Industry*, 27 (4-5), 367-378.

Nijs, Vincent, Shubha Srinivasan, and Koen Pauwels (2006), "Retail-Price Drivers and Retailer Profits," *Marketing Science*, forthcoming.

OLAP Report (2005), http://www.olapreport.com/.

Pauwels, Koen (2004), "How Dynamic Consumer Response, Competitor Response, Company Support and Company Inertia Shape Long-term Marketing Effectiveness," *Marketing Science,* 23 (4), 596-610.

Petouhoff, Natalie (2006), "The Scientific Basis for CRM Failures," *Customer Relationship Management Magazine*, 10(3), 48.

Ramaprasad, Arkalgud (1987), "Cognitive Process as a Basis for MIS and DSS Design," *Management Science,* 33(2) 139-148.

Reda, Susan (2002), "Retailers Slow to Adopt Analytics Software," *Stores*, 84 (6), 22.

Reda, Susan (2002), "Despite Early Positive Results, Retailers Haven't Jumped on Analytics Bandwagon," *Stores*, 85 (3), 34.

Roberts, John H., Glen L. Urban (1988), "Modeling Multiattribute Utility, Risk and Belief Dynamics for New Consumer Durable Brand Choice," *Management Science,* 34(2) 167-185.

Rowe, Anna L., Nancy J. Cook (1995), "Measuring Mental Models: Choosing the Right Tools for the Job," *Human Resource Development Quarterly,* 6(3) 243-255.

Rust, Roland T., J. Jeffrey Inman, Jiamin Jia, and Anthony Zahorik (1999), "What You Don't Know About Customer-Perceived Quality: The Role of Customer Expectation Distributions," *Marketing Science*, 18 (1), 77-92.

Silva-Risso, Jorge, Randolph E. Bucklin and Donald G. Morrison (1999), "A Decision Support System for Planning Manufacturers' Sales Promotion Calendars," *Marketing Science*, (Vol 18, No 3), 274-300.

Silver, Mark S. (1990), "Decision Support Systems: Directed and Nondirected Change," *Information Systems Research,* 1(1) 47-70.

Sullivan, Laurie (2005), "Fine-tuned pricing," *Information Week*, August. (www.informationweek.com)

Tversky, Amos, Daniel 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.Godwin, Steve J. Davis (1992), "A Comparative Analysis of DSS User-Friendliness and Performance," *International Journal Of Information Management,* 12(3) 209-223.

Van Bruggen, Gerrit H., Ale Smidts, Berend Wierenga (1996), "The Impact of the Quality of a Marketing Decision Support System: An Experimental Study," *International Journal of Research in Marketing,* 13(4) 331-343.

Wierenga, Berend, Gerrit H. Van Bruggen (1997), "The Integration of Marketing Problem-Solving Modes and Marketing Management Support Systems," *Journal of Marketing*, 16(3) 21-37.

Wierenga, Berend, Gerrit H. Van Bruggen and Rick Staelin (1999), "The Success of Marketing Management Support Systems," *Marketing Science*, 18(3), 196-207.

Wood, Robert E., Albert Bandura, Trevor Bailey (1990), "Mechanisms Governing Organizational Performance in Complex Decision-making Environments," *Organizational Behavior and Human Decision Processes*, 46(2) 181-201.

Zoltners, Andris A. and Prabhakant Sinha (2005), "Sales Territory Design: Thirty Years of Modeling and Implementation," *Marketing Science*. 24 (3), 313-331.

Figure 1: The 3-Gap Framework: The Effect of Gaps between Mental Model, MDSS model, and True Model

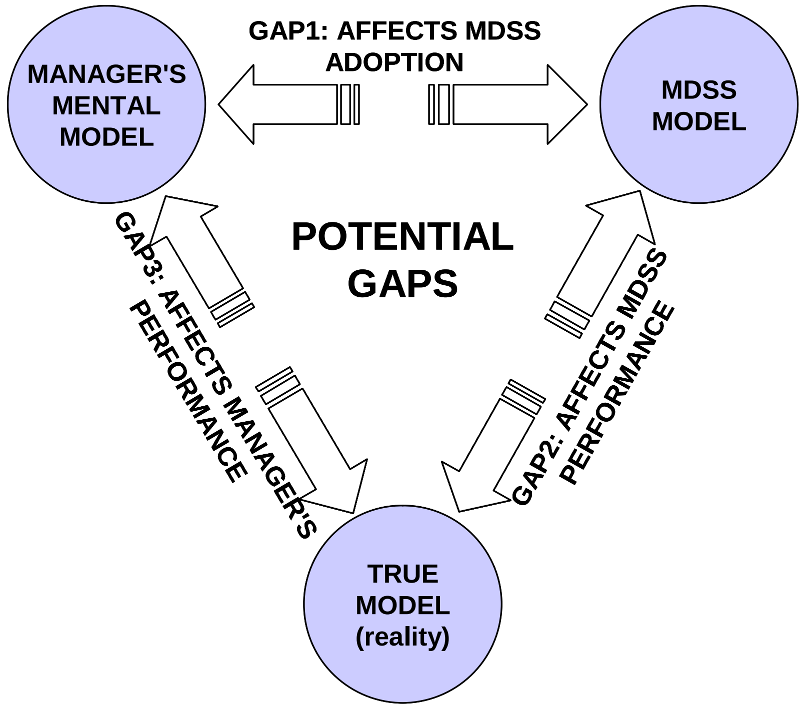
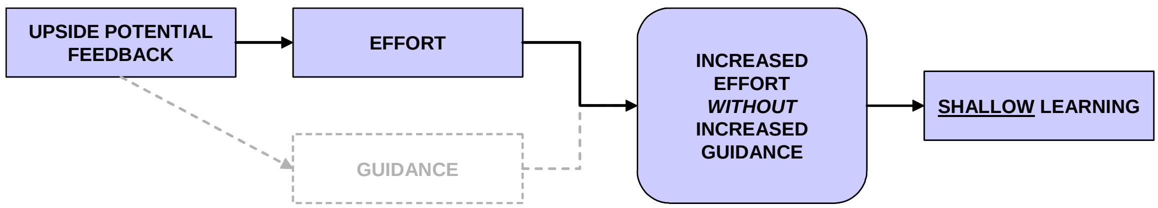
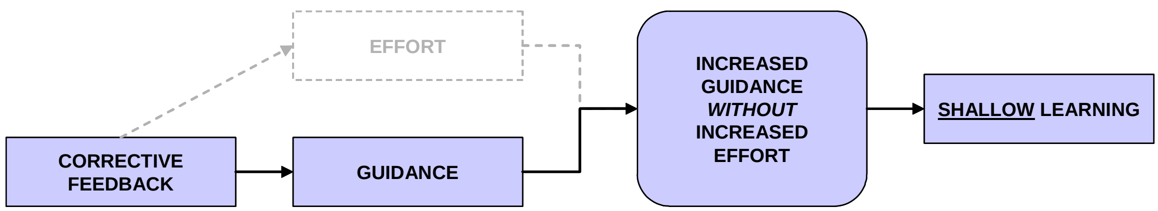


Figure 2: Theoretical Framework Relating Feedback to Learning and Evaluation (dotted lines indicate expectations of non-significant links)

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



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



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

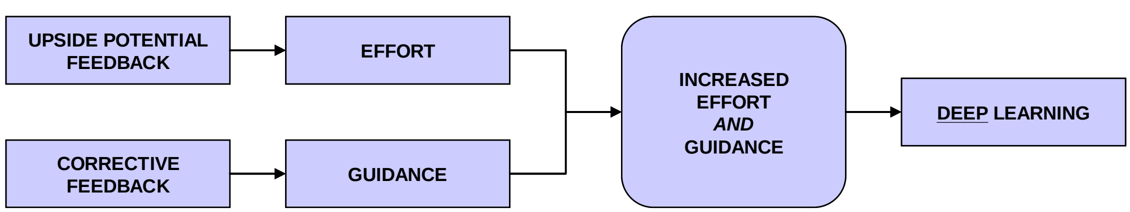
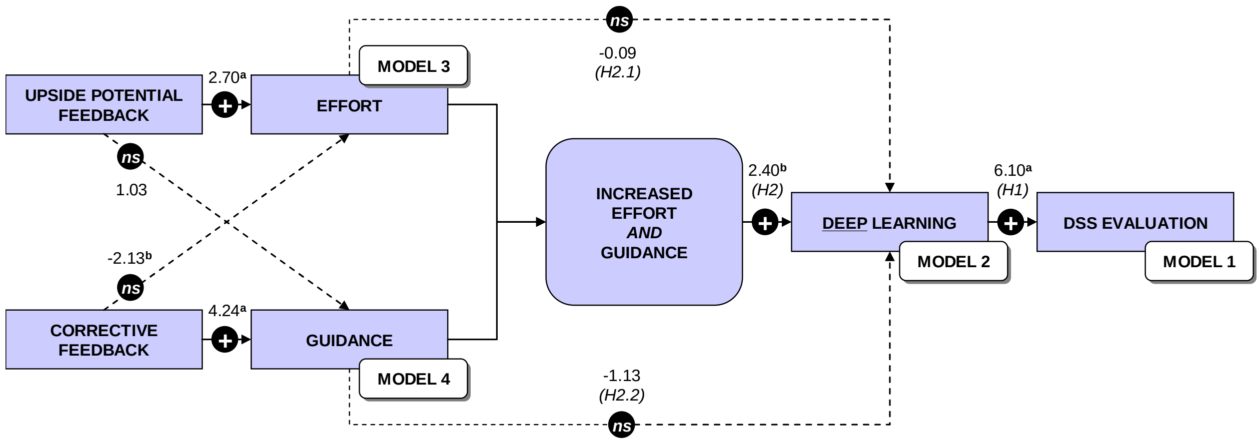


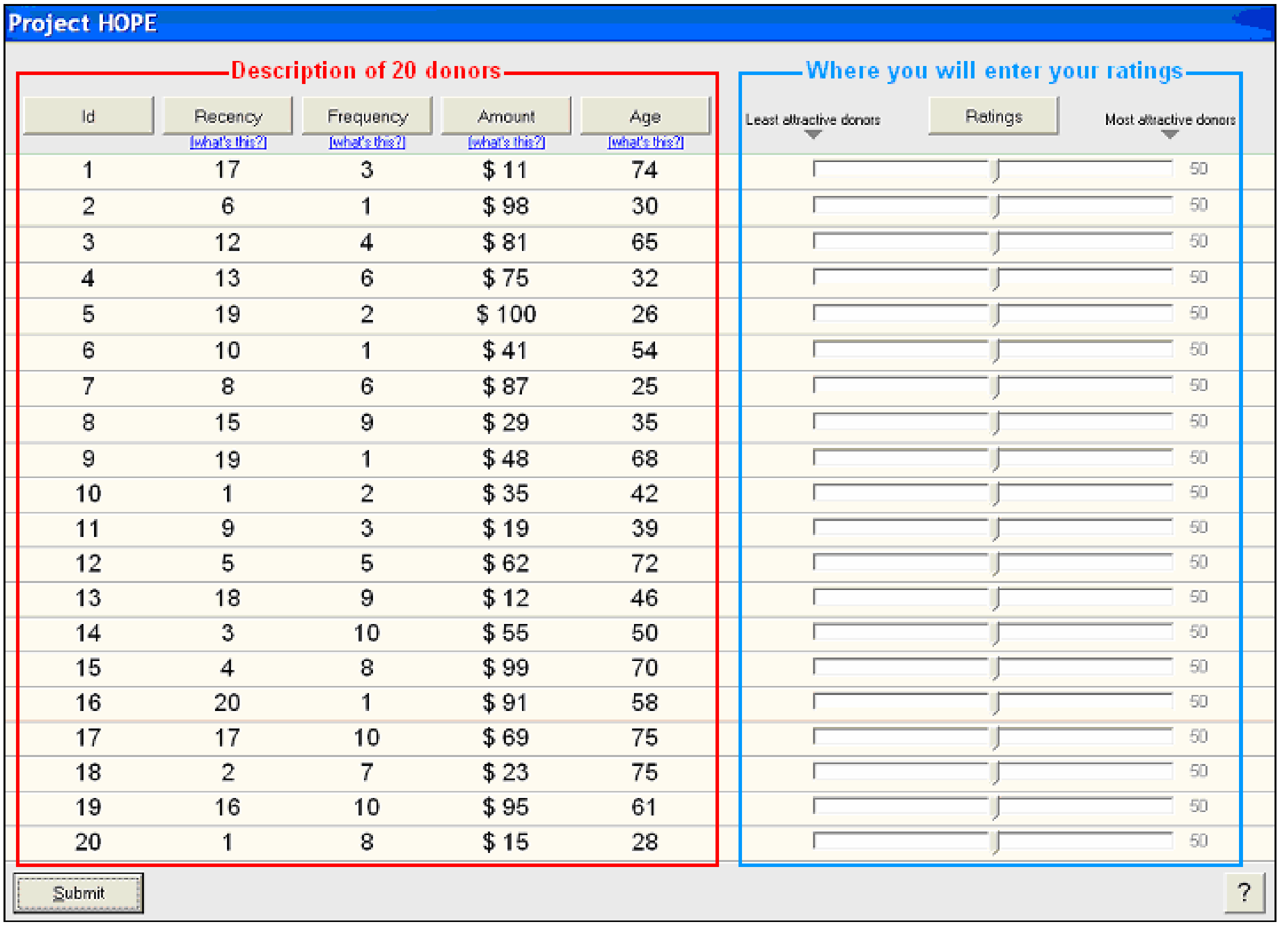
Figure 3: Connecting MDSS Design Characteristics, Deep Learning, and MDSS Evaluation (models M1-M4; dotted lines indicate expectations of non-significant links)



Notes:

1. ***ns*** stands for *no significant effect expected.*
2. + stands for *positive effect expected.*
3. We report t-statistics and statistical significance ([a]p<0.01, [b]p<0.05, two-tailed), as well as hypothesis numbering when appropriate.

Figure 4: MDSS Interface, Illustrating the Respondent Task



Id	Recency	Frequency	Amount	Age	Ratings
1	17	3	$ 11	74	50
2	6	1	$ 98	30	50
3	12	4	$ 81	65	50
4	13	6	$ 75	32	50
5	19	2	$ 100	26	50
6	10	1	$ 41	54	50
7	8	6	$ 87	25	50
8	15	9	$ 29	35	50
9	19	1	$ 48	68	50
10	1	2	$ 35	42	50
11	9	3	$ 19	39	50
12	5	5	$ 62	72	50
13	18	9	$ 12	46	50
14	3	10	$ 55	50	50
15	4	8	$ 99	70	50
16	20	1	$ 91	58	50
17	17	10	$ 69	75	50
18	2	7	$ 23	75	50
19	16	10	$ 95	61	50
20	1	8	$ 15	28	50

Figure 5: Experimental Sequence (Box A) and Measures of Learning (Box B)
(dotted arrows indicate the mapping of tasks to measures)

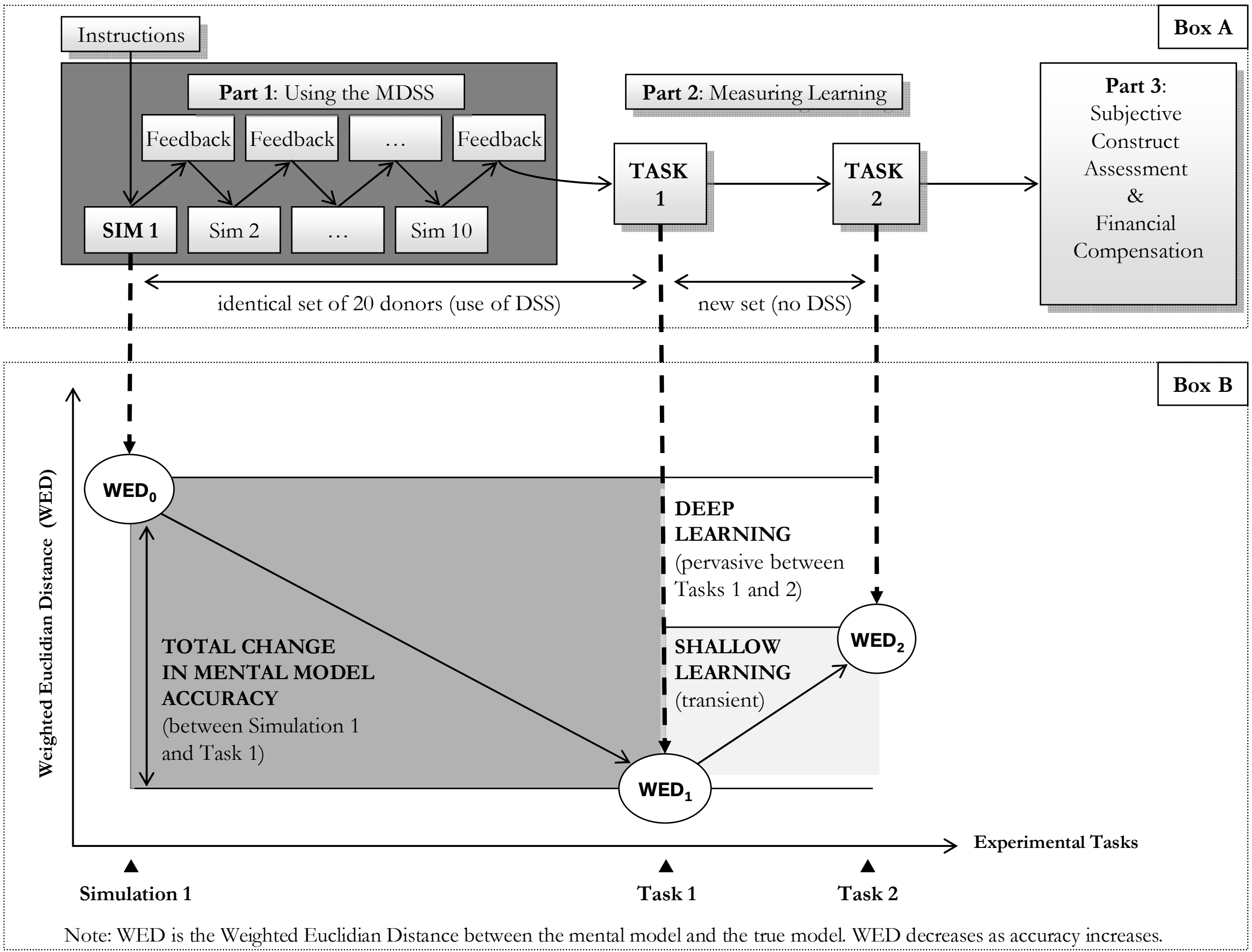


Table 1: Relationship between MDSS Evaluation, Deep Learning, Effort, and Guidance

Part A: Effect of Deep Learning on MDSS Evaluation (Model M1)

Variable	Coefficient	Beta	t-stat	Hypothesis
Intercept	β_{01}	3.351	4.01[a]	
Deep learning	β_{11}	0.045	6.10[a]	H1
Shallow Learning	β_{21}	-0.001	-0.01	
Financial compensation	β_{31}	0.010	0.43	

Part B: Effect of Effort and Guidance on Deep Learning (Model M2)

Variable	Coefficient	Beta	t-stat	Hypotheses
Intercept	β_{02}	-19.79	-0.93	
Effort	β_{12}	-0.45	-0.09	H2.1
Guidance	β_{22}	-6.67	-1.13	H2.2
Effort×Guidance	β_{32}	3.39	2.40[b]	H2

Part C: Effect of Feedback Type on Effort and Guidance

1. Effort (Model M3)

Variable	Coefficient	Beta	t-stat
Intercept	β_{03}	4.43	39.86[a]
UPSIDE POTENTIAL FEEDBACK	β_{13}	0.29	2.70[a]
CORRECTIVE FEEDBACK	β_{23}	-0.23	-2.13[b]

2. Guidance (Model M4)

Variable	Coefficient	Beta	t-stat
Intercept	β_{04}	2.94	15.04[a]
UPSIDE POTENTIAL FEEDBACK	β_{14}	0.21	1.03
CORRECTIVE FEEDBACK	β_{24}	0.88	4.24[a]

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

Table 2: Effect of Upside Potential Feedback and Corrective Feedback on the Combination of Effort and Guidance[1]

Condition	Mean	Difference from ALL condition	t-stat
CONTROL condition	12.84[2]	-4.48	-3.75[a]
UPSIDE POTENTIAL FEEDBACK condition	14.34[2]	-2.98	-3.44[a]
CORRECTIVE FEEDBACK condition	15.56[3]	-1.76	-2.13[b]
ALL condition	17.32		

Notes:

1. *We test here whether (*Effort×Guidance) *is greater in the ALL condition than in each of the other conditions.*
2. *Mean in this condition is significantly less that for the ALL condition at* [a]$p<0.01$ *(two-tailed).*
3. *Mean in this condition is significantly less that for the ALL condition at* [b]$p<0.05$ *(two-tailed).*

Table 3: Effect of Upside Potential Feedback and Corrective Feedback on Deep Learning[1]

		A. Is deep learning significantly different from 0?		B. Is deep learning significantly less than that in ALL condition?	
Condition	N	Average Deep Learning	t-stat	Difference	t-stat
CONTROL condition	16	0.25	0.09	-12.20	-3.03[a]
UPSIDE POTENTIAL FEEDBACK condition	18	5.34	2.00[b]	-7.11	-1.90[b]
CORRECTIVE FEEDBACK condition	20	5.40	1.92[b]	-7.05	-1.84[b]
ALL condition	18	12.45	4.42[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 POTENTIAL FEEDBACK condition is significantly different from 0 at [b]$p<0.05$, *and is significantly less than that in the ALL condition at:* [b]$p<0.05$ *(both are one tailed tests).*
3. Deep learning in CORRECTIVE FEEDBACK condition is significantly different from 0 at [b]$p<0.05$, *and is significantly less than that in the ALL condition at:* [b]$p<0.05$ *(both are one tailed tests).*
4. Deep learning in ALL condition is significantly different from 0 at: [a]$p<0.01$ *(one tailed).*
5. Indicates that deep learning in ALL condition is significantly greater than that in other conditions.

Table 4: Effect of Upside Potential Feedback and Corrective Feedback on Shallow Learning[1]

		A. Is shallow learning significantly different from 0?		B. Is shallow learning significantly more than that in ALL condition?	
Condition	N	Average Shallow Learning	t-stat	Difference	t-stat
CONTROL condition	16	4.52	3.27[a]	2.11	1.21
UPSIDE POTENTIAL FEEDBACK condition	18	4.97	3.99[a]	2.56	1.59[c]
CORRECTIVE FEEDBACK condition	20	6.33	4.86[a]	3.92	2.38[b]
ALL condition	18	2.41	1.84[b]		

Notes:
1. We test here whether (A) shallow learning is significantly different from 0 in each of the conditions, and (B) shallow learning is significantly more than that in the ALL condition.
2. Shallow learning in UPSIDE POTENTIAL FEEDBACK condition is significantly different from 0 at [a]$p<0.01$, *and is significantly more than that in the ALL condition at:* [c]$p<0.10$ *(both are one tailed tests).*
3. Shallow learning in CORRECTIVE FEEDBACK condition is significantly different from 0 at [a]$p<0.01$, *and is significantly more than that in the ALL condition at:* [b]$p<0.05$ *(both are one tailed tests).*

Table 5: Effect of Upside Potential Feedback and Corrective Feedback on Financial Performance and Compensation[1]

Condition	Average financial performance in Task 1	Average financial performance in Task 2	Average financial compensation	Difference from ALL condition	t-stat
CONTROL	$ 76,328	$ 61,583	35.98[2]	-3.46	-2.50[a]
UPSIDE POTENTIAL FEEDBACK	$ 77,860	$ 62,301	36.18[2]	-3.26	-2.60[a]
CORRECTIVE FEEDBACK	$ 84,226	$ 62,778	36.67[3]	-2.77	-2.16[b]
ALL	$ 85,118	$ 76,402	39.44		

Notes:

1. We test here whether financial compensation of participants is greater in the ALL condition than in each of the other conditions.

2. Mean compensation in this condition is significantly less that for the ALL condition at [a]$p<0.01$ *(two-tailed).*

3. Mean compensation in this condition is significantly less that for the ALL condition at [b]$p<0.05$ *(two-tailed).*

4. We note that financial compensation is directly proportional to financial performance (participants were paid 0.015% of financial performance, plus $15 participation fee). The maximum financial performance possible in each task is about $100,000.

Table 6: Validation results

MODEL	SPECIFICATION	LL	AIC
Full model (with direct and indirect links)	(M1) $MDSS\ Evaluation = \beta_{01} + \beta_{11} \cdot DeepLearning + \beta_{21} \cdot ShallowLearning + \beta_{31} \cdot Financial\ compensation + \varepsilon_1$ (M2) $DeepLearning = \beta_{02} + \beta_{12} \cdot Effort + \beta_{22} \cdot Guidance + \beta_{32} \cdot (Effort \times Guidance) +$ $+ \beta_{42} \cdot \text{UPSIDE POTENTIAL FEEDBACK} + \beta_{52} \cdot \text{CORRECTIVE FEEDBACK} + \varepsilon_2$ (M3) $Effort = \beta_{03} + \beta_{13} \cdot \text{UPSIDE POTENTIAL FEEDBACK} + \beta_{23} \cdot \text{CORRECTIVE FEEDBACK} + \varepsilon_3$ (M4) $Guidance = \beta_{04} + \beta_{14} \cdot \text{UPSIDE POTENTIAL FEEDBACK} + \beta_{24} \cdot \text{CORRECTIVE FEEDBACK} + \varepsilon_4$	-529.36	1090.73
Our Model	(M1) $MDSS\ Evaluation = \beta_{01} + \beta_{11} \cdot DeepLearning + \beta_{21} \cdot ShallowLearning + \beta_{31} \cdot Financial\ compensation + \varepsilon_1$ (M2) $DeepLearning = \beta_{02} + \beta_{12} \cdot Effort + \beta_{22} \cdot Guidance + \beta_{32} \cdot (Effort \times Guidance) + \varepsilon_2$ (M3) $Effort = \beta_{03} + \beta_{13} \cdot \text{UPSIDE POTENTIAL FEEDBACK} + \beta_{23} \cdot \text{CORRECTIVE FEEDBACK} + \varepsilon_3$ (M4) $Guidance = \beta_{04} + \beta_{14} \cdot \text{UPSIDE POTENTIAL FEEDBACK} + \beta_{24} \cdot \text{CORRECTIVE FEEDBACK} + \varepsilon_4$	-529.76	1087.53
Restricted Model (without process variables)	(M1) $MDSS\ Evaluation = \beta_{01} + \beta_{11} \cdot DeepLearning + \beta_{21} \cdot ShallowLearning + \beta_{31} \cdot Financial\ compensation + \varepsilon_1$ (M2) $DeepLearning = \beta_{02} + \beta_{42} \cdot \text{UPSIDE POTENTIAL FEEDBACK} + \beta_{52} \cdot \text{CORRECTIVE FEEDBACK} + \varepsilon_2$ (M3) $Effort = \beta_{03} + \varepsilon_3$ (M4) $Guidance = \beta_{04} + \varepsilon_4$	-545.35	1110.70

Notes: Based on the AIC criterion, our model has a better fit than the competing alternatives.

Publications in the Report Series Research* in Management

ERIM Research Program: "Marketing"

2006

The Effect of Feedback and Learning on DSS Evaluations
Ujwal Kayande, Arnaud de Bruyn, Gary Lilien, Arvind Rangaswamy and Gerrit Van Bruggen
ERS-2006-001-MKT
http://hdl.handle.net/1765/7302

Interaction Between Shelf Layout and Marketing Effectiveness and Its Impact On Optimizing Shelf Arrangements
Erjen van Nierop, Dennis Fok and Philip Hans Franses
ERS-2006-013-MKT
http://hdl.handle.net/1765/7632

Institutional Forecasting: The Performance of Thin Virtual Stock Markets
Gerrit H. Van Bruggen, Martin Spann, Gary L. Lilien and Bernd Skiera
ERS-2006-028-MKT
http://hdl.handle.net/1765/7840

Irritation Due to Direct Mailings from Charities
Merel van Diepen, Bas Donkers and Philip Hans Franses
ERS-2006-029-MKT
http://hdl.handle.net/1765/7832

When Should Nintendo Launch its Wii? Insights From a Bivariate Successive Generation Model
Philip Hans Franses and Carlos Hernández-Mireles
ERS-2006-032-MKT
http://hdl.handle.net/1765/7867

How Feedback Can Improve Managerial Evaluations of Model-based Marketing Decision Support Systems
Ujwal Kayande, Arnaud de Bruyn, Gary Lilien, Arvind Rangaswamy and Gerrit van Bruggen
ERS-2006-039-MKT

---

* A complete overview of the ERIM Report Series Research in Management:
https://ep.eur.nl/handle/1765/1

ERIM Research Programs:
LIS Business Processes, Logistics and Information Systems
ORG Organizing for Performance
MKT Marketing
F&A Finance and Accounting
STR Strategy and Entrepreneurship