Social Influence *and End-User Training*

Dennis F. Galletta

Manju Ahuja

Amir Hartman

Thompson Teo

A. Graham Peace

It is a commonly held axiom that user training is a key element in MIS success [3, 16]. Included among positive outcomes afforded by user training are improved user attitudes, behavior, and performance. Although the typical focus of training programs is their technical content, many practitioners have demonstrated that social factors could be instrumental in the target system's success or failure.[1] Indeed, it is recommended that researchers examine how methods of training can enhance motivation to learn and use software [3, 16, 17]. Unfortunately, there is little or no literature that includes actual manipulation of such "soft" variables [16].

The study presented in this article provides such manipulation. More specifically, we wanted to discover the extent to which training outcomes such as attitudes, behavior, and performance are influenced by peers through informal, verbal, word-of-mouth (WOM) communication, rather than derived solely through direct experience or formal channels. This article reports on a deception experiment that employed confederates in three experimental groups.

Training Outcomes

Major streams of research have for some time investigated user attitudes, behavior, and performance, as surrogates to be used for measuring success. MIS researchers have focused on user satisfaction [8] and expectations [11] in exploring user attitudes. Olfman [16] and Bostrom et al. [3] utilize user attitudes as indicators for motivation to use a system. Most user-satisfaction studies have addressed its measurement, its conceptual bases, and its relationships with other variables.

Expectations might be important antecedents of user attitudes, especially when considering user training. Ginzberg [11] found that the contrast between what a user expects and what is delivered, was overshadowed by the consistency between expectations and resultant attitudes. That is, early attitudes are likely to persist, somewhat independent of the actual quality of the system.

User behavior has been studied from two general perspectives: acceptance and usage. Acceptance of information technology has been studied from the point of view of the user's intentions to adopt, using the Technology Acceptance Model (TAM) and Theory of Planned Behavior (TPB). These models are

---

[1] National news coverage of our study included a front page story in the *Wall Street Journal*, June 2, 1994, as well as stories in *Computerworld* (July 18, 1994, p. 86), *Computer Reseller News* (July 25, 1994, p. 101), *Information Week* (July 4, 1994, p. 8), and *PC Magazine* (August, 1994).

based on the Theory of Reasoned Action (TRA) [1] (see [14] for a review of MIS models based on this literature). Another commonly-used model is the theory of innovation-diffusion [21].

The TRA derivatives support the assertion that perceived usefulness and perceived ease of use affect intentions to use technology. The TAM model is somewhat unique because it includes "external variables," which have not been studied systematically to date. WOM communications might be a highly important external variable.

Innovation-diffusion theory has investigated acceptance of technology by focusing on many variables, studies, and methods. Rogers discovered that an innovation's rate of adoption was found to be highly dependent not only on perceived attributes of innovations themselves, but also by communication channels. Evidence of the importance of WOM messages was substantial in his study [21].

A study by Brancheau and Wetherbe [4] provided evidence that the source of greatest influence in all stages of adoption decision-making was from work colleagues. As the stages progress from initial knowledge to persuasion to the decision itself, the percentage of influence attributed to work colleagues rose steadily. Very little of the persuasion was attributed to computer specialists, consultants, vendors, mass media, teachers, and friends.

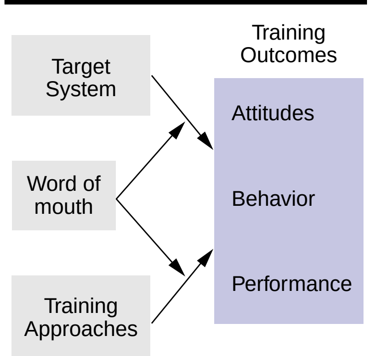
In summary, one might conclude that both acceptance of technology (measured via intentions) and usage of technology are behavioral variables that help us better understand information systems users. Further, when training is under way, WOM messages may be powerful determinants of the adoption of technology.

User performance is most often studied in the literature of human-computer interaction, a confluence of several fields such as psychology, computer science, and MIS. The literature gained substantial theoretical development in the Keystroke Model work by Card, Moran, and Newell [5]; principles set forth as a result of their research have been applied to a large portion of the experiments conducted in the field.

MIS researchers have traditionally not focused on design alternatives at or near the level of the keystroke, but have evaluated the relationships between performance and several other variables. Examples include data presentation alternatives (e.g., see [24], users' backgrounds (e.g., see [10]), and training approaches (e.g., see [7, 17].). In the MIS training literature, performance measures such as number and types of errors, and system comprehension have been used as training outcome measures [3, 22].

In this article, we explore the potential relationship between the domain of attitudes and performance, following a training exercise, working within the WOM paradigm.

Figure 1. Research model of word-of-mouth influence (adapted from [7] and [11]).



Word-of-Mouth

For many years, researchers in marketing have explored the effects of pre-purchase information received by consumers. Researchers have studied the role of WOM messages, how changed expectations affect behavior, and what factors affect the power of WOM messages.

Consumer expectations have been shown to be affected more by WOM messages than by any other factor overall [25]. Interestingly, such messages more strongly affect expectations than past personal experience, advertising, and sales promotion. One reason that consumers allow themselves to be influenced by others is to "learn about products or services by observing others and/or seeking information from others" ([2], p. 474), in particular for uncertainty reduction [21]. Other reasons are identification with or enhancement of one's image among others, or willingness to conform to others' expectations.

Consistent with the work of Ginzberg [11] in MIS described previously, researchers in marketing have found that high expectations can lead to satisfaction, even when they are disconfirmed (e.g., see [18, 23]). It is therefore important to make sure that the WOM message is highly effective.

Many researchers have attempted to understand the effectiveness of WOM messages. Several studies detect stronger effects when information about a product is unfavorable rather than favorable, and when information is verbal rather than written (e.g., see [13]). Finally, Feick and Higie [9] show that preference heterogeneity and lack of coorientation (described later) can diminish the efficacy of WOM messages.

In summary, face-to-face WOM messages have proven to be powerful influences on consumer attitudes and behavior. The results of the MIS and marketing literature suggest that there is promise in attempting to merge the perspectives of both fields. One study that has examined the impacts of interpersonal word-of-mouth communication among MIS users concluded that this impact was indeed stronger than the impact of advertising in the diffusion of end-user computing technology [4]. Our study continues the investigation into the effects of such communication on users.

Hypotheses

Based on research of Bostrom et al. [3] and Davis and Bostrom [2, 7][2] as well as the marketing literature

[2] The original models include motivation and usage, which could be classified as attitude and behavior measures, respectively. Our model is explicitly structured around the training outcomes we studied.

and the goals of our study, the model presented in Figure 1 is proposed. The model asserts that the relationships between selected training outcomes and the impacts of their antecedents are moderated by WOM peer influence. In our study, both the target system (a direct manipulation interface) and training approach (instruction) are fixed; we vary only the WOM information. In general, WOM messages have been powerful determinants of consumer expectations (e.g., see [25]), attitudes (e.g., see [13]), and intentions to purchase (e.g., see [20]).

H1: Post-training attitudes will be more favorable in the positive word-of-mouth condition than in the negative condition.
H2: Post-training intentions to purchase the software will be greater in the positive word-of-mouth condition than in the negative condition.

While the marketing literature focuses on the purchase decision, in the information system setting, *usage* is often the only outcome variable that is appropriate. The purchaser of software is often a completely different party than the user; at issue in many cases is a *psychological*, rather than an *economic*, purchase.

Studies in TAM and TPB [14], and the Theory of Reasoned Action [1] demonstrate the importance of studying behavioral intentions as well as behavior. There is evidence to suggest high correlation between the two constructs (see [1]), however the intention must refer specifically to the behavior. Because our laboratory setting could not capture meaningfully a subsequent intention and behavior, we have chosen to investigate one intention (to use the software in the future) and a slightly different behavior (optional usage at the end of the experimental session). Both variables are of great potential importance in our context, although they might not correlate very highly. We therefore express them as two separate, but related, hypotheses.

H3a: Post-training intentions to use the software again will be greater in the positive word-of-mouth condition than in the negative condition.
H3b: The amount of optional use by subjects will be greater in the positive word-of-mouth condition than in the negative condition.

The final training outcome in our model is a user's performance in using software. All other things being equal, a user with diminished intentions to use the software again is expected to be less motivated in accomplishing an experimental task, and perhaps less committed to learning it.

The importance of motivation for task performance has been discussed by [15] and [17], among others. The following exploratory hypotheses formalize the expected role of motivation and commitment to learning.

H4a: Post-training performance on a comprehension task will be higher in the positive word-of-mouth condition than in the negative condition.
H4b: Post-training performance on the experimental task will be higher in the positive word-of-mouth condition than in the negative condition.

Method

An experiment was conducted to identify the impact of peer influence on new users who learned to perform a task. There were several important guidelines in designing our study. It was important to maximize the effects of WOM communication. That is, it was necessary to base the study on face-to-face communication with peers and to provide multiple opportunities for recipients of WOM communication to become cognitively engaged in the message.

It is difficult to determine the preference heterogeneity of software [9]. The marketing literature tells us that most people will agree on evaluations of a taxi cab ride (low preference heterogeneity), while few will agree on a particular hair style (high preference heterogeneity). It is indeed possible that computer software is perceived very differently by different people, and that it is more like a hair style than a cab ride in the level of agreement likely to be generated by its users. Feick and Higie found that the WOM source's similarity to the recipient is very important when high preference heterogeneity exists, and that the source's experience is very important when there is low preference heterogeneity.

Therefore, we also made use of message sources

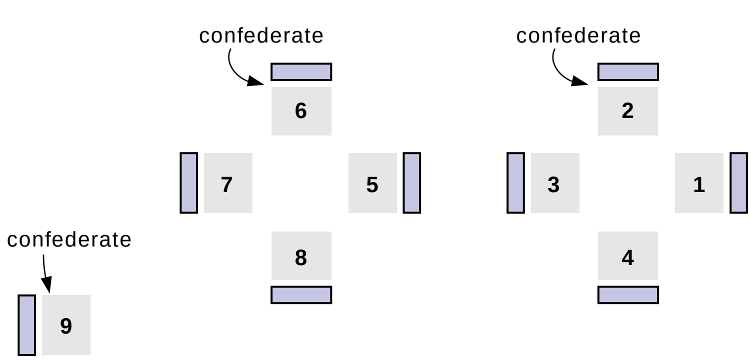


Figure 2. The temporary experimental laboratory

that were perceived as experienced while having similar values as the recipients, to mitigate the effects of high preference heterogeneity that might exist in evaluations of software.

Subjects and Incentives

Subjects were solicited from several sections of a required MBA class, presenting the opportunity to earn a $4 fee for participating, with a $3 bonus for performing above the mean, and additional prizes of between $9 and $18 if they were one of the top five performers.

All subjects in the sampling frame had received basic training in computing and spreadsheet applications nine months earlier, and since that time had received several assignments to be completed using PCs. Of the approximately 200 subjects we canvassed, 74 volunteered to participate in the study, and 53 actually attended their designated sessions.

The large number of "no-shows" caused the actual groups to be irregularly sized, with slight variation in the number of empty seats per group. Three subjects were dropped from analysis because the background questionnaire revealed they were already experienced with the somewhat rare integrated software package used in the study (described later).

Examination of all demographic measures revealed that the participating subjects mirrored the sampling frame (the single school's MBA program). The average age of the subjects was 27 (standard deviation 3.2) years, and 39 of the 50 subjects were males. The three largest majors represented were finance (42%), marketing (22%), and information systems (10%).

Experimental Design

Three separate types of experimental treatments were designed, termed as *positive, negative*, and *control*, named after the type of word-of-mouth stimuli presented to the subjects in each group. The control group subjects received no stimuli. Although no hypotheses address the control group, we included that group to provide additional data useful in explaining further the possible rejection of any hypotheses.

Each of the three treatments was administered in four separate sessions due to limitations in the size of the temporary computer lab constructed for the study (see Figure 2). Subjects signed up for a convenient session, and we chose each group's treatment type randomly to ensure that every subject had an equal chance of being assigned to each treatment. The stations were arranged to reduce the ability of

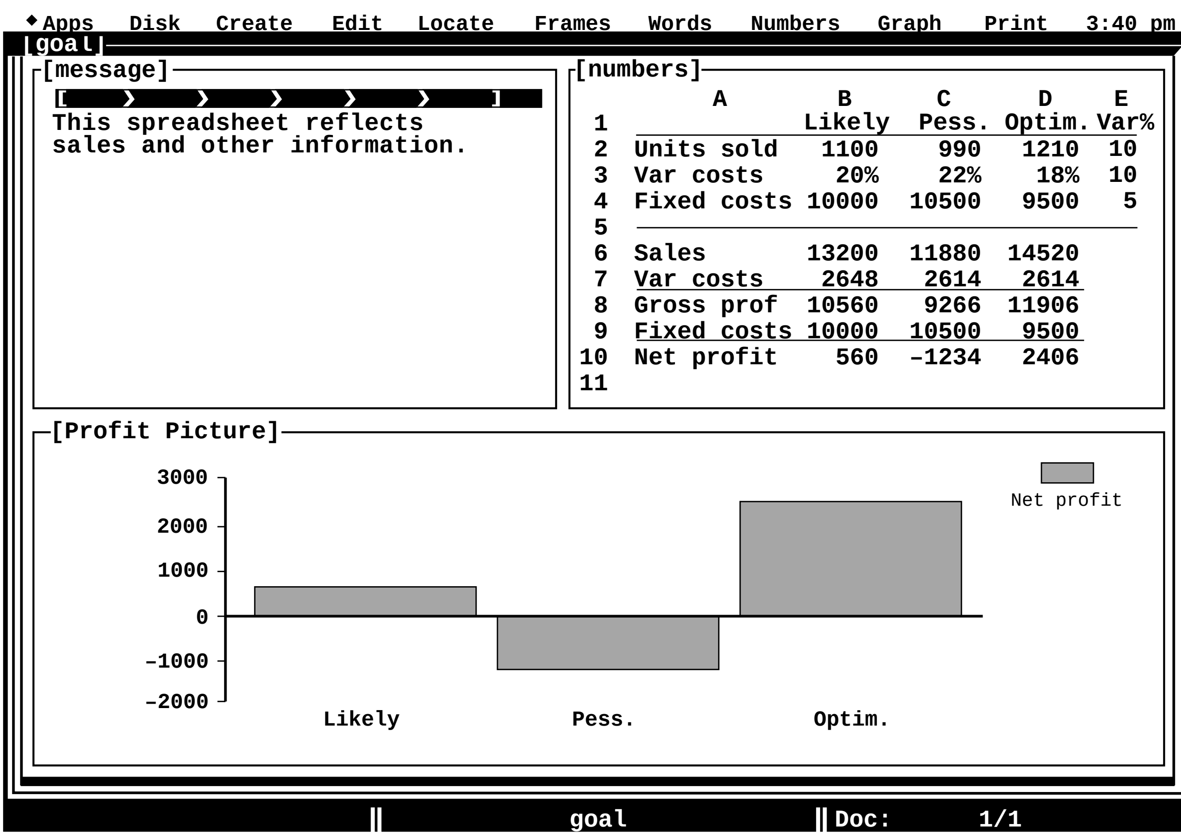


Figure 3. The goal presented to the subjects

subjects to see the work of others.

To ensure the three groups were roughly equivalent, we compared the groups on the basis of age, gender, and four other variables (extent of current use of PCs, regularity of use of word processors, regularity of use of spreadsheets, and the extent to which subjects were familiar with 21 computing devices). No differences were significant, except for the age variable. The average age of subjects in the negative group was about 25, while the average age of subjects in the other two groups was about 28. We then removed the oldest 10 and youngest 9 subjects from our testing to equalize the ages among the groups, and found most of the statistical results (described later) to be similar to those when including all subjects. Also, we computed correlations of the age variable with all of the dependent variables and found none to be significant. Thus, we found no evidence that the age difference affected our results.

Just before the second outburst ... *the experimenter walked out to the hall for a moment and returned as if he was not aware of the outburst.*

Materials

Subjects were given three packets during the course of the experiment.[3] The first was distributed to subjects while giving them general instructions, and included an informed consent form for their signature, and a pre-experiment questionnaire. The questionnaire asked about their personal characteristics and past experiences with computing (described earlier).

The second packet contained what was called a "Quick Tour" of using Borland's integrated package Framework III. The Quick Tour was inspired by, but did not closely follow, the "Minimal Manual" approach by Carroll et al. [6]. The goal was for subjects to create a "frame" that contained other frames, including text, a spreadsheet, and a graph tied to the spreadsheet (see Figure 3).

The self-paced training packet contained eight pages of keystroke-by-keystroke instructions to explain how to accomplish the task. Although the act of typing the basic spreadsheet formulas was familiar to the subjects, as were the basic typing functions involved in the text frame, these functions did not dominate the exercise. Besides the primary new feature of organizing a hierarchical document and linking its components, other novel functions were the following: building frames, navigating among frames, building the particular model, addressing ranges, selecting multiple cells, creating the graph, saving the containing frame, and formatting cell information. It is particularly realistic to use a task in which some operations are familiar and others are not, as many new systems represent "upgrades" from old ones.

At the end of the packet, subjects were asked to perform "what-if" analysis twice to determine how the graph and profit results would change by altering some of the assumptions in the spreadsheet (in the top right frame). Subjects were invited to perform this analysis on their own (up to eight times) and further observe the effects of these changes. The number of times (measured automatically using macros) each subject performed "what-if" analysis on their own gave us one measure of behavior (voluntary use, without financial incentive). The macros also saved the files for our inspection later.

The third packet contained a post-experiment questionnaire, which assessed subjects' attitudes, behavioral intentions, and amount of retention. Most measures were new because there were no existing instruments allowing assessment of the particular type of package used. Subjects' reactions to the software were assessed by summing across seven 7-point semantic differential scales (see Appendix). Behavioral intentions were assessed by using two more scales (see Appendix) that asked how likely subjects would be to reuse or purchase the software under the assumption they possessed adequate resources and needs. A quiz consisted of 10 multiple-choice questions (see Appendix), to measure comprehension task performance. Upon inspection of the saved spreadsheets, an experimental performance task score was assigned by deducting from a possible score of 10 points, 1 point for each error in following the instructions.

Pilot Tests

The materials and procedures were pilot-tested in several iterations, resulting in many changes and refinements. The Quick Tour document presented many interesting problems. First, subjects sometimes became confused and needed to backtrack when setting up the screen layout. Instructions and macros were added to allow backtracking; the material was subdivided into what became eight small, independent steps. If a subject became lost, he or she could return to the beginning of that particular step by pressing an Alt+ sequence.

Experimental Procedures

The main part of the study was intended to furnish the positive or negative cues both *before* and *during* the training. This timing was chosen because of the

[3] All packets can be obtained by contacting the first author.

potentially opposing effects of cognitive dissonance theory and contrast theory [11]. The former would predict that pre-training cues would facilitate development of attitudes consistent with those cues. However, contrast theory would predict that some subjects might then become surprised in the opposite direction, because they might raise or lower their expectations excessively. Therefore, additional outbursts during training were thought likely to reinforce the initial cues, strengthening the overall manipulation.

Upon the arrival of all subjects in a group, the experimenter described what was to be accomplished, and gave the name of the software package to be used. One of the confederates immediately raised his hand, asked for verification of the package's name, and then stated that he was already well familiar with Framework III. In the positive (negative) group, the confederate then expressed his favorable (unfavorable) perceptions of the package according to a prepared script. He also mentioned the high market share Framework holds in Europe (poor market performance in the U.S.) In the control group, the confederate expressed no opinions but merely told of his familiarity with the package. The confederate was then excused from the session and the experiment continued.

The remaining subjects were given the second packet and were asked to sit "at any computer" they wished. The remaining two confederates were always seated closest to the computers and were easily able to position themselves immediately at stations 2 and 6. The experimenter stood in the way of station 9, which was completely unable to run the software due to the lack of a hard drive.[4]

Macros were used for unobtrusive measurements. Several macros were invoked for a variety of purposes, including capturing the number of times users changed assumptions, saving the working document at various times, and presenting a title and exit screen summarizing the experimenter's verbal instructions.

On average, subjects worked on the task for 28 minutes (standard deviation 7.7). While the subjects worked, confederates in the positive (negative) groups made positive (negative) comments even though subjects were instructed to keep as silent as possible. For example, positive comments included several different loudly-whispered expressions of admiration and excitement, while negative comments included groans, sighs, and loudly-whispered expressions of how painful the exercise seemed. These scripted outbursts were timed at about 5–6 minute intervals to ensure a proper balance of salience and realism.

The experimenter reacted to the first outburst, asking if there was a question or problem, then reminded subjects to be as quiet as possible. Just before the second outburst (involving obviously sympathetic interaction between the two confederates), the experimenter walked out to the hall for a moment and returned as if he was not aware of the outburst. Just before the third (large) outburst, the experimenter walked quickly to the door as if there was an urgent message and seemed to miss it once again. Depending on timing, the fourth or fifth (minor) outburst was scheduled to occur at about the time when the experimenter was preoccupied distributing packet 3 to the first real subject who completed the task. The experimenter again admonished subjects lightly for talking.

Minor outbursts continued on fairly regular intervals as subjects began to hand in their work. These outbursts were kept brief and were only meant to "renew" the salience of the joy or frustration for the remaining subjects. We considered it necessary to treat the outbursts as unsanctioned behavior because we needed to control the setting as much as possible. If real subjects were allowed to speak out loud, then each group would be different; our level of analysis would then be at the *group* rather than at our target, the *individual* level. Therefore, the risk of making subjects uncomfortable with outbursts was considered to be outweighed by the added control from ensuring that all groups within a treatment were exposed to identical messages.

We considered it necessary to treat the outbursts as unsanctioned behavior
because we needed to control the setting as much as possible

After collecting packet 3, each subject was thanked, paid, and followed out into the hallway. As a manipulation check, the experimenter asked each subject if the outbursts distracted him or her. He then reminded the subject not to discuss any aspect of the experiment (including the outbursts) with anyone else until the experiment was completed (in three days).

The manipulation check was not very effective because subjects seemed reluctant to disclose that they heard the outbursts. About a third of the subjects said they heard nothing, but when the experimenter pointed in the general direction of machines 2 and 6, all but five subjects suddenly acknowledged that they were aware of the outbursts, acting as if they had at first misunderstood the question. This conflict between their initial and revised answers was puzzling

[4] Workstation 9 was never intended to be used, but the lack of a ninth station might have raised subjects' suspicions.

to the experimenters, who speculated that the subjects might have indeed been made a bit uncomfortable by the outbursts. Most of the subjects quickly assured the experimenter that the outbursts did not affect their performance.

The procedure met the goals described earlier. Face-to-face WOM statements were provided by using confederates. Cognitive engagement was maximized through consistent and repetitive outbursts. Source experience and peer similarity were simultaneously achieved by using multiple student confederates.

Due to the small number of subjects, univariate t-tests comparing only the positive and negative groups were used to test each hypothesis. This is appropriate because no non-hypothesized relationships were explored.

Results

Results will be discussed in three sections, covering user attitudes, behavior and behavioral intentions, and performance.

Attitudes. Cell means for the custom-made 7-item attitude scale[5] are shown in Table 1. As Hypothesis 1 predicts, the negative group is significantly lower than the positive group.

Behavior and Behavioral Intentions. Behavior was examined using conventional behavioral intentions surrogates, and also using a new approach: counting the number of voluntary "what-if analysis" iterations performed by subjects. Subjects were asked if they would use the software again, and if they would purchase the software.[6] Cell means (see Table 2) both differed as predicted by Hypotheses 2 and 3a. Once again, the control group appears to be more similar to the positive group than to the negative group.

In accordance with Hypothesis 3b, actual behavior was examined. Subjects in the two groups appeared to perform about the same number of additional, voluntary "what-if" analysis iterations (see Table 3). Although the control group mean appears to be larger than either of the experimental group means, even the most liberal testing failed to indicate a significant difference, as did the appropriate, conservative post-hoc comparisons [12].

We also investigated the degree to which the optional iterations correlated with intentions to purchase or use the software. Both correlations were non-significant and only 1% of the variation in the number of optional iterations would be explainable by knowing each of the intentions if the correlations had been significant.

Performance. The final, exploratory hypothesis was examined by measuring two performance variables: a 10-point quiz score and a 10-point task score. As Table 4 illustrates, mixed results were obtained for Hypothesis 4. Means for the quiz scores for negative and positive group subjects differed significantly in the hypothesized direction, but not the means for the task score.

Discussion

There were significant findings in each of the three categories we examined. These findings appear to support the assertion that word-of-

Table 1. Attitude Scores
Mean Scores on Semantic Differential Item Indicating Overall Attitudes About the Software (sum of 7 items; 7=positive; 1=negative)

	Negative Group n=16	Control Group n=18	Positive Group n=16	Negative vs. Positive (1-tailed)
Mean Score (std. dev.)	31.6 (7.8)	36.3 (6.5)	37.1 (8.6)	t=1.87; p=0.035*

*significant at the p=0.05 level

Table 2. Behavioral Intentions Scores
Mean Scores on Semantic Differential Items (7=positive; 1=negative)

	Negative Group n=16	Control Group n=18	Positive Group n=16	Negative vs. Positive (1-tailed)
Likelihood of using again	4.0 (1.6)	5.1 (1.9)	5.4 (1.8)	t=2.40; p=0.012*
Likelihood of purchasing	3.4 (2.0)	4.4 (2.0)	4.8 (2.0)	t=1.94; p=0.031*

*significant at the p=0.05 level

Table 3. Behavior Scores
Mean Number of Times Subject Performed Optional Task (minimum of 0; maximum of 8)

	Negative Group n=16	Control Group n=18	Positive Group n=16	Negative vs. Positive (1-tailed)
Number of times	0.9 (1.4)	1.9 (2.2)	0.9 (1.2)	t=−0.01; ns

not significant at the p=0.05 level

Table 4. Performance Scores
Mean Scores on Quiz and Task (higher score indicates better performance)

	Negative Group n=16	Control Group n=18	Positive Group n=15	Negative vs. Positive (1-tailed)
Score on 10-item quiz	8.1 (1.3)	8.8 (1.0)	8.8 (0.9)	t=1.85; p=0.038*
Score on 10-point task	9.7 (0.6)	9.2 (1.4)	9.2 (1.3)	t=−1.42; ns

*significant at the p=0.05 level

[5] Cronbach's alpha = 0.89

[6] The two behavioral intention scores were highly correlated (r=0.78; p<0.001), exhibiting satisfactory levels of concurrent validity.

mouth communication can be a significant and important determinant of training outcomes, namely, attitudes, behavior, and performance. Along with a short discussion of each of the major findings, we discuss some of the limitations of the study.

Major Findings. Our subjects who were exposed to unfavorable WOM statements appeared to adopt unfavorable attitudes toward the software, in comparison to the subjects exposed to positive statements. Interestingly, the control group mean appears to be virtually equivalent to that of the positive group, which would suggest that negative WOM comments are more potent than positive comments. This finding is consistent with that of much of the marketing literature, which has found that negative WOM is more salient and hence has more of an impact than positive WOM: "negative information tends to be more diagnostic or informative than positive or neutral information" ([13], p. 460).

There are two alternative explanations for the apparent lack of effect in the positive group. First, our positive treatment might simply not have been as convincing as our negative treatment. Another possible explanation might be that the use of word-of-mouth communication during training may have had offsetting effects. Positive outbursts by the confederates may have indeed created in the subjects a more positive frame of mind, but the outbursts may also have created somewhat of a negative feeling toward the confederates or the setting because of the distractions they created. These conflicting feelings might have neutralized the positive treatment. Nevertheless, the WOM literature would support our first explanation.

Both measures of behavioral intentions appeared to differ as a result of the negative WOM communication. However, actual behavior (voluntary iterations in "what-if analysis") did not differ between groups. Several alternative explanations can account for this unexpected finding.

Perhaps the additional work was not meaningful for them, or perhaps subjects tried to finish quickly. Another possible explanation is the skewness of the results; nearly half of the subjects performed none of the optional iterations, and almost 10% performed only one of the two required iterations. Finally, different subjects might have possessed several different types of motivation for continuing the iterations; experimental artifacts such as subjects who attempt to please the experimenters (or *not* please the experimenters) might have influenced subject behavior to a greater degree than did the treatment.

Performance results were also mixed. The quiz score appeared to be significantly different between the positive and negative groups, but not the actual task score. The lack of variability of the task score might explain the failure in detecting a significant difference. Over half of the subjects (31) received perfect scores on the task. Only six subjects made more than one error on the task. There was much more variation in the quiz scores (although quiz scores were also quite high). Our pilot testing failed to reveal this lack of discrimination of the task score, and future studies should be conducted with more difficult tasks and testing.

While most of the hypotheses were fully or partially supported by the data, there are several other limitations of this study. First, the subjects were MBA students in a laboratory, not working professionals using the software to accomplish real tasks. However, nearly all of the MBA students from which the sample was drawn have previous work experience, which raises the extent to which we can generalize the results. Also, the time period of the study was extremely short; the entire cycle from the WOM communication to task performance was completed in under an hour. A future study in the field that examined long-term effects could be an especially fruitful venture. However, such assessment will be difficult without strong controls that are available in an experimental setting. Finally, the random assignment of treatments to groups rather than subjects to treatments made us vulnerable to systematic dependent variable differences. However, our testing revealed that our final results were unaffected by bias in subject assignment.

Implications

In this study we demonstrated how word-of-mouth communication can affect the outcomes of training, even keeping constant the characteristics of the software and the training approach. The findings show that the same training method will be less successful when trainees are exposed to negative WOM rather than positive WOM. Trainers concerned with end-user attitudes and performance need to be wary of negative WOM before, during, and after training, because such communication can undermine training efforts. Both researchers and practitioners can benefit from the implications of this experiment. In general, this study has provided evidence that researchers should consider WOM communication as one possibly important determinant of trainee attitudes, behavior, and even performance. Also, there are many unanswered questions pertaining to the preference heterogeneity of software.

Although software would initially appear to be quite heterogeneous in how it is perceived by users, and thus less susceptible to word-of-mouth effects, perhaps another factor could explain the strength of our results. Many of the items offered as examples in the definitions by marketing researchers (e.g., hair

Appendix. Questionnaire Items

Software Evaluation

Overall reactions to the software:

terrible			neutral			wonderful
1	2	3	4	5	6	7
frustrating			neutral			satisfying
1	2	3	4	5	6	7
dull			neutral			stimulating
1	2	3	4	5	6	7
difficult			neutral			easy
1	2	3	4	5	6	7
inadequate power			neutral			adequate power
1	2	3	4	5	6	7
rigid			neutral			flexible
1	2	3	4	5	6	7
not useful			neutral			useful
1	2	3	4	5	6	7

Future Use

How likely are you to use this software again if given the opportunity ?

never			neutral			definitely
1	2	3	4	5	6	7

Assuming you have the resources and need, would you purchase this software ?

never			neutral			definitely
1	2	3	4	5	6	7

Quiz items

Ten multiple-choice questions asked subjects how to:

1. Combine spreadsheets, text, and graphics
2. Size frames
3. "Jump" inside a frame
4. Create a graph
5. Set margin width
6. Underline cells
7. Perform "what-if" analysis
8. Select more than one cell
9. Cancel a selection
10. Describe its overall features available

styles; restaurant dinners) are highly heterogeneous in the extent to which different people react to the same treatment, and therefore word-of-mouth communications are not as valuable in predicting a given person's reaction.

Interestingly, highly complex computer software is likely to be even higher in preference heterogeneity; software opinions tend to be passionate and highly varied. What, then, could possibly explain the value of WOM communication? The answer is likely to be related to the high amount of effort involved in personally experiencing software. This effort raises the value of others' negative opinions; a bad report could save a person significant installation and learning time.

Researchers should investigate the reasons positive WOM communication was ineffective in enhancing training outcomes in our study. It would be useful to determine if negative information indeed tends to be more informative, if our positive treatment was not as convincing, or if distraction is a key difficulty in disseminating positive WOM. Perhaps future research could also determine the potential effects of varying characteristics of the source or characteristics of individual subjects.[7]

Practitioners should probably pay more attention to detecting any rumors and gossip that occur before new software is introduced. The potential effects of negative WOM communications on training outcomes should be of great concern, whether the software is produced or purchased, required or optional. In the case of required software, user attitudes, behavior, and performance can suffer. In the case of optional software, users who have been exposed to negative communications can choose to adopt unsupported packages or fail to adopt any software, both undesirable outcomes.

The new evidence of a performance effect can be important to consider when designing training programs; some time should probably be spent conveying accurate information to users before beginning the sessions, perhaps including videos of peer testimonials.

Although theoretical and empirical findings tell us that positive word-of-mouth information is not likely to help significantly, it might be very valuable to at least maintain active, open, and honest communication with users. Such communication can help detect negative sentiment early enough to correct any misunderstandings and perhaps take the opportunity to make favorable changes where the negative communications are correct.

User attitudes, behavior, and performance are important but elusive information systems training outcomes. By extending our understanding of these variables and their interrelationships, we might avoid needless difficulties when introducing new systems.

Acknowledgments

The authors would like to thank Kieran Mathieson for comments on earlier versions of this article. An earlier version of this article was presented at the Fifteenth International Conference on Information Systems, December 1994, Vancouver. Please contact the first author for a copy of the experimental materials.

[7] For example, see work by Petty and Cacioppo [19] who investigate source credibility and ability of a receiver to process a message.

References

1. Ajzen, I., and Fishbein, M. *Understanding Attitudes and Predicting Social Behavior.* Prentice-Hall, Englewod Cliffs, NJ, 1980.
2. Bearden, W.O., Netemeyer, R.G., and Teel, J.E. Measurement of consumer susceptibility to interpersonal influence. *J. Consumer Research 15,* (1989), 473–481.
3. Bostrom, R.P., Olfman, O., Sein, M.K. The importance of learning style in end-user training. *MIS Q. 14,* 1 (1990), 101–119.
4. Brancheau, J. C., and Wetherbe, J. C. The adoption of spreadsheet software: Testing innovation diffusion theory in the context of end-user computing. *Info. Syst. Research 1,* 2 (1990), 115–143.
5. Card, S., Moran, T. and Newell, A. *The Psychology of Human-Computer Interaction.* Erlbaum, Hillsdale, NJ, 1983.
6. Carroll, J.M., Smith-Kerker, P.L., Ford, J.R., and Mazur-Rimetz, S.A. The minimal manual. *Human-Computer Int.* 3, 2 (1987–88), 123–154.
7. Davis, S.A., and Bostrom, R.P. Training end users: An experimental investigation of the roles of the computer interface and training methods. *MIS Q, 17,* 1 (1993), 61–85.
8. DeLone, W.H., and McLean, E.R. Information systems success: The quest for the dependent variable. *Info. Syst. Research 3,* 1 (1992), 60–95.
9. Feick L.F., and Higie, R.A. The effects of preference heterogeneity and source characteristics on Ad processing and judgments about endorsers. *J. Advertising, XXI,* 2 (June, 1992), 9–23.
10. Galletta, D.F., Abraham, D., El Louadi, M., Lekse, W., Pollalis, Y., and Sampler, J. An empirical study of spreadsheet error-finding performance. *Accounting, Management, and Information Technology, 3,* 2 (1993), 79–95.
11. Ginzberg, M.J. Early diagnosis of MIS implementation failure. *Manage. Sci. 27,* 4 (1981), 459–478.
12. Hays, W.L. *Statistics, 2nd. ed.* Holt, Rinehart, and Winston, NY, 1988.
13. Herr, P.M., Kardes, F.R., and Kim, J. Effects of word-of-mouth and product-attribute information on persuasion: An accessibility-diagnosticity perspective. *J. Consumer Research 17,* (1991), 454–462.
14. Mathieson, K. Predicting user intentions: Comparing the technology acceptance model with the theory of planned behavior. *Info. Syst. Research 2,* 3 (Sept. 1991), 173–191.
15. Moran, T.P. Guest editor introduction, special issue on the psychology of human-computer interaction. *Comput. Surv. 13,* 1 (1981), 1–6.
16. Olfman, L. A Comparison of Applications-Based Training Methods for DSS Generator Software. Doctoral dissertation, Indiana University., 1987.
17. Olfman, L. and Bostrom, R.P. End-user software training: An experimental comparison of methods to enhance motivation. *J. Info. Syst. 1,* 4 (1991), 249–266.
18. Oliver, R.L. and DeSarbo, W.S. Response determinants to satisfaction judgments. *J. Consumer Research 14* (Mar. 1988), 495–507.
19. Petty, R.E., and Cacioppo, J.T. The elaboration likelihood model of persuasion, in L. Berkowitz, Ed., *Advances in Experimental Social Psychology, 19,* Academic Press, NY, 1986.
20. Richins, M.L. Negative word-of-mouth by dissatisfied consumers: A pilot study. *J. of Marketing, 47* (Winter 1983), 68–78.
21. Rogers, E.M. *Diffusion of Innovations 3rd ed.* Free Press (Macmillan publishing), NY, 1983.
22. Sein, M.K. Conceptual Models in Training Novice Users of Computer Systems: Effectiveness of Abstract vs. Analogical Models and Influence of Individual Differences. Doctoral dissertation, Indiana University, 1988.
23. Tse, D.K., and Wilton, P.C. Models of consumer satisfaction formation: An extension. *J. Marketing Research, 25* (May 1988), 204–212.
24. Vessey, I., and Galletta, D.F. Cognitive fit: An empirical study of information acquisition. *Info. Syst. Research 2,* 1 (1991), 63–84.
25. Webster, C. Influences upon consumer expectations of services. *J. Services Marketing 5,* 1, (Winter 1991) 5–17.

About the Authors:

DENNIS F. GALLETTA is an associate professor of Business Administration at the Katz Graduate School of Business, University of Pittsburgh. Current research interests include end-user computing and user training, particularly in the areas of end-user attitudes, behavior, and performance. email: galletta@vms.cis.pitt.edu; URL: http://www.pitt.edu/~galletta

MANJU K. AHUJA is a doctoral candidate in Management Information Systems at the University of Pittsburgh. Current research interests include computer-supported cooperative work, virtual organizations, organizational learning, and management of systems development. email: manju@vms.cis.pitt.edu

AMIR HARTMAN is a doctoral candidate at the Katz School of the University of Pittsburgh. Current research interests include social and organizational implications of information technology, sociology of scientific inquiry, and telecommunications policy. email: ahartman+@pitt.edu

A. GRAHAM PEACE is a doctoral candidate in Management Information Systems at the University of Pittsburgh. Current research interests include software piracy, privacy issues, accessibility, and systems analysis and design. email: agp@vms.cis.pitt.edu

Authors' Present Address: Katz Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15260.

THOMPSON TEO is a lecturer in the Faculty of Business Administration at the National University of Singapore. Current research interests include strategic impacts of information technology, adoption and diffusion of information technology, telecommuting, and end-user training. **Author's Present Address:** National University of Singapore, Department of Decision Sciences, 10 Kent Ridge Crescent, Singapore 0511; email: fbteosh@nus.sg