VALIDATION AND LESSONS FROM THE FIELD -- APPLICATIONS OF INFORMATION ACCELERATION

by

Glen L. Urban, John R. Hauser, William J. Qualls, Bruce D. Weinberg, Jonathan D. Bohlmann, and Roberta A. Chicos

Massachusetts Institute of Technology, Cambridge, MA
Alfred P. Sloan School of Management, W.P. #3819-95

INFORMATION ACCELERATION --
VALIDATION AND LESSONS FROM THE FIELD

Glen L. Urban, John R. Hauser, William J. Qualls, Bruce D. Weinberg,
Jonathan D. Bohlmann, and Roberta A. Chicos

August 1995

Glen L. Urban is Dean of the Sloan School of Management, Massachusetts Institute of Technology, 50 Memorial Drive, E52-474, Cambridge, MA 02142, (617) 253-6615, (617) 258-6617 fax. John R. Hauser is Kirin Professor of Marketing, Massachusetts Institute of Technology, Sloan School of Management, 38 Memorial Drive, E56-314, Cambridge, MA 02142. William J. Qualls is Associate Professor of Marketing, Massachusetts Institute of Technology, Sloan School of Management, 38 Memorial Drive, E56-323, Cambridge, MA 02142. Bruce D. Weinberg is Assistant Professor at Boston University, School of Management, Room 111, 621 Commonwealth Avenue, Boston, MA 02215. Jonathan D. Bohlmann is a Ph.D. student at the Sloan School of Management, Massachusetts Institute of Technology, 38 Memorial Drive, Cambridge, MA 02142. Roberta A. Chicos is an independent consultant at One School Street, Arlington, MA 02174.

ABSTRACT

The more "really new" a product is, the more difficult it is to predict consumer response. However, accurate forecasting of really new products is an essential element in a firm's innovation strategy. Therefore many leading firms are experimenting with the use of multimedia computers to present information to consumers so that these customers can react to a really new product as if it were now in the market of the future. We explore two aspects of best practice in this area of forecasting by examining the validity of a new multimedia forecasting methodology called "information acceleration" (IA) and lessons learned from eight real-world new-product applications of the methodology.

We report two internal validations and one external validation of new product sales forecasts based upon IA. The internal validation for a new automobile suggests that a computer-simulation of an automobile showroom provides forecasts that are not significantly different from those obtained based on a physical showroom. The internal validation of a medical instrument suggests that a computer-simulation of a medical technician provides forecasts that are not significantly different from those based on allowing physicians to interact with real technicians. The external validation of a new camera product suggests that IA provides forecasts that are sufficiently accurate for managerial go/no go decisions. For that application we compare actual sales to the initial forecasts, made in 1992 for sales in 1993 and 1994, to adjusted forecasts based on the actual marketing plan, and to forecasts adjusted for an unforeseen negative *Consumer Reports* article. We close the paper with a summary of the lessons that we have learned during the past five years based on eight applications of IA to really new products.

Introduction

As firms move from an era of downsizing and re-engineering they are devoting more attention to revitalization through new products. Some new products are line extensions, some are upgrades and improvements, some are new technological solutions in established categories, and some are really new -- new products that serve major new market segments by providing breakthrough solutions to customer needs in those segments. Examples include electric vehicles, electronic home imaging systems, office-in-a-briefcase telecommunication systems, new medical instruments that change the way doctors test patients, and new architectures (with implied advantages) of integrated circuits. The more "really new" a product is, the more difficult it is to forecast customer response.

Information acceleration (IA) is one method that has been proposed to forecast customer response to really new products. IA uses a multimedia computer to accelerate information to a customer so that the customer can react to a really new product as if he or she were now in the market of the future. For example, IA can create a virtual showroom for an electric vehicle where the potential customer can "walk" around the car, "climb in," and discuss the car with a salesperson. The customer can access television advertising and consumer magazine articles, read prices in the newspaper, and even get advice from fellow customers -- all while sitting at the computer. For a more complete description of information acceleration and its application to General Motor's new electric vehicle see Urban, Weinberg and Hauser (1995). For an analysis of how customers respond to virtual information see Hauser, Urban and Weinberg (1993). These published articles suggest that IA provides managers with qualitative information that they find credible and useful. But are the forecasts valid? By what process does IA affect managerial decisions? This paper provides some insight into both questions.

Over the last five years IA has been used to forecast the sales of eight different products in the automotive, telecommunications, computer, medical instrument, and photography markets. This paper reports on our validation experience with three of those applications and on the managerial lessons that we have learned. These lessons begin to define best practice for premarket forecasting of really new products. We begin with a brief review of IA and then turn to the validations and the managerial lessons. We close with a summary of future research directions.

Information Acceleration

We illustrate information acceleration with an example. A Complete Blood Cell (CBC) count analyzer is a common medical testing device that measures the various constituents of blood (white blood cells, hemoglobin, platelets, etc.). The standard technology, which has existed for over 20 years, costs almost $100,000, requires blood to be sent to central laboratories with the corresponding one to-two-day turn-around time, and requires that the blood leave the sample tube (thus creating the potential for biohazard). A new technology provides the potential for a fundamental change in the market because of its lower price ($20,000), ease of use, and increased safeguarding against biohazards make it feasible for the tests to be done in a doctor's office.

In an IA application physicians and laboratory technicians are allowed to gather information on the new CBC analyzer with a multimedia computer. This analyzer allows immediate in-office comprehensive blood testing using a two-part test with nine parameters. The technology is based on spinning the sample in a centrifuge and reading the results optically after reagents have been added. (Existing central-laboratory technology uses electrostatic readings.) For example, Figure 1 illustrates (a) a simulated physician colleague, (b) a product brochure, (c) a magazine advertisement, (d) a simulated medical technician, (e) a salesperson, and (f) a medical journal article. (A demonstration videotape of the information sources is available from the authors.) Each source of information is designed to simulate as closely as is feasible that information which would be accessible in the market. The respondent chooses which information to access, how long to "visit" an information source, and whether to return to an information source. The respondent is not forced to visit all of the sources. Reactions to this information are conditioned on the environment that the respondent is expected to face. In our example, we described the future reimbursement rules for in-office tests (e.g., more restrictions but reimbursement for office tests lowers costs and provides for better care.) This information is given at the beginning of the IA.

After visiting each source the respondent is asked to estimate his or her probability of purchase. We use a 100-point purchase intention scale, but other scales could be used as well. By using IA within an experimental design we compare different scenarios, say with and without regulation, and we benchmark forecasts against the sales of a known product. The benchmarking is done by comparing IA results from a control product (a new product now in the market) to its actual sales results and estimating the forecasting bias (e.g., 10% overstatement). This bias factor is then applied to the IA results for the new product (see Urban, Weinberg, and Hauser, 1995, for details). Naturally, the benchmark is limited by how closely it matches the really new product. This is often a challenge simply because the product is really new, but our experience suggests that adequate benchmarks can be identified.

Because the information available in the IA may not be available to everyone or may take time to reach everyone, IA can be used in parallel with other marketing models. For example, in some applications we conditioned forecasts on "awareness through advertising" and then estimated "awareness through advertising" based on the firm's marketing plan. In other applications we have used probability-flow models that simultaneously simulate the growth in the availability of a variety of information sources. (See example in Urban, Weinberg and Hauser 1995.) In addition, we have coupled IA with conjoint analysis (Green and Srinivasan 1990). In an electric vehicle application we used IA to make a base forecast and then used conjoint analysis to modify that base forecast based on projected changes in product features. In other words, IA is a complement to existing marketing models, not a substitute.

Some advantages of IA include future conditioning, full information availability and conditioning, user control, and active search. By future conditioning we mean that the respondent can be placed in a frame of mind (and information level) corresponding to future markets. By full information availability and conditioning, we mean that all information can be accessed and the forecasts can depend upon the amount of information available. By user control we refer to the fact that information is not forced on the respondent. The respondent can select as little or as much information as he or she wants. If we select the "cost" of that information carefully, then we hope that the respondent will select only that information that he or she would have selected (will select) in a real market environment where information is costly. Finally, IA allows the respondent to search actively rather than be a passive information receptor.

Managers in our applications have found the IA to have high face validity. But does IA have internal and external validity?

Validation

Validation is difficult. Forecasts are never a single number; they are conditioned on assumptions about the market environment in which the product will be introduced and about the marketing plan with which the product will be introduced. For example, suppose we forecast that 1,000,000 units of a new computer product will be sold in year 1 as an addition to the current product line, but a year later we observe that the firm has sold 5,000,000 units. At first glance, this does not appear to be an accurate forecast. However, suppose that contrary to plan, the firm dropped all the existing products and offered only the new product. Further, suppose that the model would have forecast 5,000,000 units had it been asked to forecast for the new-product-alone scenario; then the model would appear to be valid. Alternatively in another application, if we forecast 1,000,000 units at $10, but sell 700,000 units at $20, should we assess validity against the forecast that was made at $10 or the forecast that would have been made at $20? Of course, if we make one such correction, we must make all such corrections even if they cause the forecast to deviate from actual sales. (For example, we might have sold the units at $20, but with a greater investment in advertising.)

We address validation in steps. We begin with two studies of internal validation and then report one study of external validation. In the internal validations we compare forecasts under alternative conditions to determine if these conditions affect the forecast. The first internal validation compares forecasts made based on a computer-simulated automobile showroom and a computer-simulated salesperson to those based on a physical showroom and a real salesperson. The second internal validation compares a computer-simulated technician to the doctor's actual technician. These comparisons do not depend upon external factors nor on the realized marketing plan. Once we establish some degree of internal validation, we can evaluate the external validation. The external validation compares IA forecasts to actual sales.

Internal Validation of a Computer Showroom

The application for the first internal validation was the forecast of a new (at the time in 1990) two-seated Buick sporty car, the Buick Reatta convertible. Although there were other comparable cars on the market at the time, such a concept was new to Buick. One advantage of this first application of IA was that this category enabled us to choose a known product, the Mazda RX-7 convertible, as the experimental control. By implementing a complete IA for the RX-7 and a complete IA for the Reatta, we could compare the RX-7 IA forecasts to known RX-7 sales and thus provide a baseline for the sales of the Reatta. For example, if the IA forecast for the Reatta was 50% less than that for the RX-7, then, all else equal, we would predict that the Reatta would sell 50% fewer units than the RX-7. However, all else is not equal. Thus, we use probability-flow models to incorporate such differences as advertising, word-of-mouth, pricing, dealer availability, life cycle phenomena, and industry volume. The use of a test vs. control anchoring methodology and the use of probability-flow models were already well-established at Buick (Urban, Hauser, and Roberts 1990). The incremental change in methodology, which we wished to test, was whether IA could replace the traditional automobile "clinics" where respondents viewed and drove the preproduction cars, watched advertising, saw simulated word-of-mouth, and read simulated magazine articles. Such a test was important to the sponsor because IA could be used before investing in a pilot plant to build preproduction cars and before building a sufficient quantity of hand-built prototypes.

The information that was simulated with the IA was chosen based on qualitative consumer interviews, manufacturer and dealer experience, and prior academic research. The specific information sources were *advertising* (magazine, newspaper, and television advertising), *interviews* (unrehearsed video of actual consumers), *articles* (simulated consumer-information and trade publications), and *showroom* (a chance to walk around the car, sit in the car, and ask a salesperson questions). The internal validation opportunity was made possible because the *showroom* was simulated for some of the respondents with a multimedia computer and for other respondents with a visit to view the actual car and talk to the actual salesperson -- the same salesperson that was videotaped for the computer presentation. For both sets of respondents all other aspects of the IA experience were held constant. (A videotape of the IA is available from the authors.) For more detail on the measures see either Hauser, Urban, and Weinberg (1993) or Weinberg (1992).

Sample

The target sample was chosen from the registration records of consumers who had purchased a sporty car in the last two years. Respondents were pre-screened via telephone on whether they would consider purchasing a two-seated sports car as their next car and whether they were willing to spend $20,000 or more on the purchase. Those who qualified were invited to participate in the study, given a time and location at which to appear, and promised $25 to cover incidental expenses related to participation. The final sample of 177 respondents was assigned randomly to treatments shown in Table 1.

Comparison of Dependent Measures

Two dependent measures are relevant to comparing the impact of the above experimental design: (1) the full-information judged probabilities of purchase and (2) the change in judged probability before and after the *showroom* visit. The full-information probability was measured at the end of the customer's complete IA search thus capturing any interactions the *showroom* visit might have with other information sources and with the allocation of time to other information sources. The change in probability measures are the changes in the judged probability from one source to the next and thus capture the incremental impact of the *showroom* information source. The judged probability scale was a 100-point thermometer scale that is a modification of an 11-point purchase intention scale (Jamieson and Bass 1989, Juster 1966, Kalwani and Silk 1983, Morrison 1979). It was administered after each information source and at the end of the IA.

Table 2 and Figure 2 compare the impacts of the *showroom* type and the *automobile* type on both dependent measures directly measured in the IA. There was no significant difference between the *showroom* type -- thus suggesting a degree of internal validity. There was a significant difference between *automobile* type, thus suggesting that the IA is a sufficiently sensitive measurement instrument to pick up differences between the Reatta and the RX-7.

We also compared the time that respondents spent in the showroom, not counting the time going to and coming from the physical showroom. On average, respondents spent 3 minutes and 25 seconds in the computer showroom and 3 minutes and 23 seconds in the physical showroom. This difference was not significant ($t=0.11$). These relatively brief times reflect the efficiency of examining one car and the degree of other information sources presented about the car. They are appropriate for an internal validity test. Naturally, for a real-world dealer visit the consumer might travel to and from the showroom, road-test the car, and attempt to negotiate the price with the salesperson. This would take significantly longer.

Internal Validation of Simulated Human-Interaction

The application for the second internal validation is the CBC analyzer described in Figure 1. This application enabled us to test the IA methodology with a business-to-business product for which the purchase decision involved more than a single decision maker. Although many people influence the decision to purchase a medical instrument, the primary decision participants for the CBC analyzer are the physicians who use the CBC test results and the technicians or nurses who operate the CBC analyzers. The doctor is the primary decision-maker but is often strongly influenced by technician/nurse recommendations.

The information that was simulated with the IA was chosen based on focus groups, questionnaires, and interviews with CBC analyzer manufacturers. The specific information sources (review Figure 1) were *information from colleagues* (video clips of a person portraying a physician or technician who would share his or her experiences -- a participating physician could access a physician; a participating technician could access a video technician), *product brochures, magazine advertisements, medical journal articles, sales presentation*s (video of a salesperson including a computerized product demonstration), *cost analyses by an accountant*, and *staff discussions* (opportunity for the physician to interact with a technician sharing information and opinions about the CBC analyzer). The internal validity compared a simulated technician to an actual discussion with a real technician. For more detail see Urban, Qualls, and Bohlmann (1994).

Prior to receiving information about the CBC analyzer each respondent was given information about expected health care reforms, federal regulations, and Medicare reimbursements. In order to provide seed information so that the respondents could decide whether to search for more information, we asked them which information source they were most likely to use to gather information on new medical equipment. They began the IA by receiving information from that source (e.g., salesperson visit, word-of-mouth, brochure). After the initial forced exposure, they were allowed to access all subsequent sources in any order, if at all, that they preferred.

A sample datum was a physician and a technician from the same medical practice, who were involved typically in equipment purchases. We describe first the actual-technician cell of the comparison. In the actual-technician cell both the physician and the technician went individually through an IA on the CBC analyzer. All information sources were available except the *staff discussions*. After completing the initial IA, they came together, face-to-face, to discuss the CBC analyzer and make a group purchase decision. They then returned to their individual IA to record their final judgments. (We compare the physicians' judgments to the judgments obtained in the simulated-technician cell; we use the technician's judgment for diagnosis of the stimuli. See Figure 3.)

In the simulated-technician cell the technician completed an IA and recorded his or her final judgments including a judged probability that the medical practice would purchase the CBC analyzer, a recommendation (purchase, undecided, not purchase), and an overall favorability rating (7 point scale). Based on prior qualitative analysis of technicians' reactions to the stimuli we created three simulated technicians (positive, neutral, and negative) that represent the range of possibilities. For each technician we selected the simulated technician that matched the actual technician's recommendations. The physician then completed an IA which included all information sources including the simulated technician that best represented the views of the actual technician from the physician's office.

At the end of the IA, to test the simulated-technician induction, we asked the actual technicians to view the three simulated technicians. The actual technicians judged (1) the overall favorability of the simulated technicians and (2) the overall purchase recommendations that the simulated technicians represent. On a 7-point favorability scale the actual technicians' mean rating of the simulated technicians was 6.3 for the positive simulation, 4.0 for the neutral simulation, and 2.1 for the negative simulation. These were significantly different at the 0.01 level ($F(2,60)=61.9$). Furthermore, the overall recommendations (purchase, undecided, not purchase) were significantly correlated at the 0.01 level with the categorization of the simulated technicians (*Pearson* $\chi^2=70.4$).

Sample

The target sample was chosen from independent physician practices who used CBC testing for their patients, either with their own analyzer or by using an outside laboratory. Of the 157 qualified practices, 40 agreed to send a physician and a technician. Each participating practice received an honorarium of $150. By using portable computers we were able to perform the interviews at the respondents' offices if they so requested. Of the 37 practices participating for which we have complete data, 20 were assigned randomly to the actual-technician cell and 17 were assigned to the simulated-technician cell of the IA.

Comparison of Dependent Measures

The dependent measure is the physician's judged probability on a 100-point thermometer scale of purchasing the CBC analyzer. This physician's judged probability was strongly related to the physician's overall recommendation on the 3-point scale of "purchase", "undecided", and "not purchase". The purchase point of the 3-part scale corresponded to 81.7% in the purchase probability scale; the undecided point corresponded to 41.5% purchase probability; and the not purchase part to 9.6% purchase probability. The differences across the three points are significant at the 0.01 level ($F(2,34)=57.7$). Thus, we use the 100-point judged probability in the subsequent comparisons.

Doctors and technicians had similar reactions; in both the simulated and actual conditions the probability declined with full information. (This emphasizes the danger of overly optimistic forecasts when only initial concept exposures are used. Subsequent information can reveal features that lower the benefit of the product to the respondent). There was no significant difference between the physician's judged probabilities for the actual-technician and simulated-technician cells on either the concept forced exposure measure (after exposure to the first information source), the final measure (after exposure to all chosen sources including the actual or simulated technician), or the difference between the two measures. (*t=.2* for the initial measure, *t=.2* for the final measure, and *t=.4* for the difference.) See Table 3. The differential impact that is relevant to the comparison of the actual and simulated technician is measured by either the final measure or the difference.

Prior to the experiment the respondents had no knowledge of the new CBC analyzer, thus the pre-experiment probability of purchase is not defined. What we are calling "concept forced exposure" is the measure taken after exposure to the first information source. From previous research we know that the second and subsequent sources can have a measurable impact on judged probabilities, but the largest impact usually comes from the first information source. See Hauser, Urban and Weinberg (1993) for details. Nonetheless, we can compare the impact of second and subsequent information sources on the final judged probabilities. While the effect of these information sources is larger than the difference between the actual vs. simulated technician, that effect is only significant at the 0.26 level (*t=1.15*). However, if we assume that the original state of awareness and purchase probability are zero, the significance would be at the .01 level (*t=6.1*). The true (latent) initial probability and the true statistical significance is probably between that corresponding to the difference from the forced exposure and that corresponding to the difference from the assumed original state of unawareness.

To examine the impact of the actual- vs. simulated-technician further, we tested a series of covariates including judged performance, judged cost, technician favorability, confidence, current-system evaluation, whether or not the medical practice already has a CBC analyzer, CBC test volume, the number of medical instruments in the practice's office, the percent of patients on Medicare, and the years of experience. These covariates (and the actual-vs.-simulated dummy variable) were included in a regression with judged probability as the dependent measure. The adjusted R^2 was 0.55 and the actual-vs.-simulated dummy variable was not significant (*t=0.63*), although several other variables were significant (e.g., price, satisfaction with current system, volume of CBC medical tests). We obtained a similar result when Tobit analysis was used to account for the non-negativity of the dependent measure. For details see Urban, Qualls, and Bohlmann (1994).

In summary, there was no measured difference between the actual technician and the simulated technician. The IA appeared to be more sensitive in picking up differences between the latent initial, concept forced exposure, and final judged probabilities, but these may not be highly significant due to the small sample size of 37 medical practices.

Other Comparisons

At the end of the IA, the physicians were asked to rate on a 7-point scale the realism and influence of the various information sources that they searched. Of the simulated information sources, the simulated technician was rated as the most realistic. On the other hand, the physicians who interacted with an actual technician rated that information source significantly higher in influence than the physicians who interacted with a simulated technician ($t=2.1$). See Table 4. Thus, although the simulated technician was perceived as highly realistic, the physicians did not feel that it was as influential as an actual technician. However, the lower perceived influence did not have a measurable (significant) effect on the judged probability of purchasing the CBC analyzer.

Our interpretation of the data is that the simulated technician does yet not fully replace interactions with an actual technician, but that the simulated technician provides a viable means to model some of the multi-person interactions in business-to-business situations. With more experience, improved simulations, and larger sample sizes, we predict that IA has the potential to provide an internally valid representation of human interaction.

External Validation of Actual vs. Forecast Sales for a New Camera

Pre-analysis -- Contemporaneous (Internal) Validation

In an earlier paper we reported briefly on a partial validation of a new camera (Urban, Weinberg and Hauser 1995). In that validation, IA was used to forecast the sales of a new camera. The information sources included television advertising, a simulated mass merchandise store environment, simulated word-of-mouth communication, and a simulated consumer magazine article. The validation was only partial because the new camera was already on the market. Sales, awareness, distribution, word-of-mouth, and other marketing variables were known at the time of the forecast. (Sales of the new camera were known by management, but this knowledge was not used to make the forecast.) The sample consisted of 100 respondents for the new camera and 100 respondents for a control camera for which market information was known. The respondents were chosen from those who had not yet received information about the relevant camera. Forecasts were based on a probability flow model that accounted for the fact that some respondents gain awareness before going to the store and some gain awareness at the store. The flow model also accounted for the fact that cameras are purchased both for personal use and as gifts -- especially in December.

In this partial validation, the (known) marketing variables were used in the flow model and purchase probabilities and store-visit probabilities were the basis of the sales forecast. (Before the forecast, biases in the purchase probabilities were removed multiplying the index of actual-to-predicted results in the control camera condition times the purchase probabilities in the test camera condition). In this limited forecasting situation (which we label as an internal validation rather than an external validation), the IA did quite well. Forecasts were within 10% of actual sales during the first and second years. Based on these initial results the camera company decided to use IA to forecast the sales of a really new camera before market launch.

New Home Camera Premarket Forecast Validation

Based on the confidence generated in the pre-analysis, in 1992 the firm used an IA to measure premarket customer response to a new camera with a novel film format and film handling capability. The firm believed that this camera was a substantial improvement at its price point. The design of the IA was similar to that in the partial validation. That is, the same types of information sources and a similar probability flow model were used.

The control product for the IA was the current camera product line that the firm was offering. (The test cell contained the new camera plus the current camera product line.) Because the pricing strategy was not known at the time of the IA, the respondents engaged in a discrete-choice pricing experiment in which they made tradeoffs of various prices and features (e.g., picture format, camera style). The forecasts were made in 1992; the camera was introduced to the market in August of 1993.

Differences Between Marketing Plan in the Initial Forecast and Marketing Plan as Realized

The initial forecasts were based on the marketing strategy that was in the marketing plan at the time of the IA measurement. Later, when the product was introduced to the market the firm had changed its marketing plan. In particular,

- Advertising spending was 30% above plan.
- Product distribution was above plan in mass merchandising channels. However, the camera was out of stock in the fourth month in about 8% of the distribution outlets.
- Advertising copy and tactics were better at generating awareness than had been planned. For example, advertising test scores were above average. In addition, the largest mass merchandiser featured the camera in its advertising during the first year.

- The price was 30% above plan in mass merchandising channels and 50% above plan in other channels.
- *Consumer Reports* published an article in May 1994 (June issue) that was much less favorable than had been anticipated (it criticized the new picture format).

Some of these changes would have increased the forecast had they been known in 1992 while others would have decreased the forecast.

Because the IA forecasts are based on a probability flow model we are able to simulate what would have been forecast had these changes in the marketing plan been known at the time of the forecast. (See examples in Urban, Hauser, and Roberts 1990.) For example, we use a previously calibrated response function based on an advertising agency's historical data to predict how the increased advertising spending would increase advertising awareness. We used the discrete-choice pricing model to predict the effect of the price change. This new probability of becoming aware is then used in the probability flow model to make the new forecast.

Impact of the *Consumer Reports* Article

The impact of the May 1994 *Consumer Reports* article is more difficult to forecast. For a new camera a negative article can have a substantial impact on the probability of visiting a store to look for the camera, the probability of purchasing one in the store, and on the word-of-mouth that might be generated based on the article.

To begin to measure the impact of the article, the firm used a paper-and-pencil survey of 42 respondents. Roughly half (20) of the respondents were shown the *Consumer Reports* article and roughly half (22) were shown the simulated article that had been used in the IA. For this convenience sample the probability of visiting a store after viewing the *Consumer Reports* article was just 26% of the probability of visiting a store after viewing the simulated article. Similarly, the probability of purchase (once in a store) after viewing *Consumer Reports* was 21% of that based on viewing the IA article. In the IA, 57% of the respondents "visited" the simulated-article information source. We do not know how many actual camera consumers read *Consumer Reports*, but we know it is a smaller percentage than 57% based on overall readership surveys for consumer durables. Making various sensitivity assumptions about the true percentage of *Consumer Reports* readership in this camera class, about the word-of-mouth impact on non-readers, and about the independence of the probabilities measured for the store visit and the purchase and reflecting the small sample variants, we estimate through sensitivity analyses the *Consumer Reports* effect to be a multiplier between 25% and 77%. That is, we would reduce the IA forecast by a multiplicative factor of 25% - 77% for those months following the *Consumer Reports* article.

This is a rather large range which reflects a need for further study. Below we suggest modifications to the IA measures which enable the firm to adjust the forecasts after it observes the *Consumer Reports* article. Then post event estimates provide data for future IA-based forecasts.

Comparison of Actual Sales and Forecasts

Figure 4 reports the actual sales of the new camera from September 1993 through December 1994. Because the camera was introduced during August 1993, we merged the data and the forecasts for August 1993 with those for September 1993. To disguise the data we have indexed the largest number in figure 4 to a value of 100. All other data are reported relative to that number. The errors are reported as a percent of the actual sales in a month. They are averaged over the relevant months.

Initial Forecast. The initial forecast does reasonably well with a mean absolute error (MAE) of 27% up to and including May 1994, but is off by a MAE of 106% after May 1994. (We obtain similar results with root mean squared error [RMSE]. For example, the RMSE is 30% up to May 1994 and 108% after May 1994.) However, some of these errors are due to the difference between the planned marketing strategy and the actual marketing strategy. As a comparison, Urban and Katz (1983) report an MAE of 22% before adjustment and 12% after adjustment for validation of pretest market forecasts for new package goods.

Adjusted Forecast. When we adjust for the differences in the marketing plan, the forecast is off by an MAE of 5% up to including May 1994, but off by an MAE of 72% after May 1994.

Effect of Consumer Reports. We can compute a *post hoc* adjustment factor by treating the May 1994 *Consumer Reports* article as an "event." That is, we use one degree of freedom to determine a single multiplier for post-May forecasts such that the average of all post-May forecasts equals the average of all actual sales. This multiplier does not guarantee a good fit, because there is still significant monthly variation in actual sales following May 1994. Naturally, this adjustment factor will capture both the *Consumer Reports* effect and any other unobserved effects that happened in May 1994. It is possible that the adjustment factor may overstate the effect of *Consumer Reports*. When we statistically estimate and apply this event-analysis-based "*Consumer Reports* adjustment factor" of 42% to the adjusted forecasts after May 1994, we obtain a much improved fit -- an MAE of 11% after May 1994. (This adjustment factor is within the broad range forecast by the firm's paper-and-pencil comparison of the *Consumer Reports* article and the IA-based article described in the section above).

Readers should interpret the external validations for themselves. The initial forecasts vary substantially from actual sales, but so does the marketing plan. The adjusted forecasts do much better before the *ConsumerReports* event, but not thereafter. After statistically adjusting for the negative *ConsumerReports* article the forecasts do well, but we used post-event data to make those adjustments.

Our interpretation is that given the challenge of forecasting for a really new camera, the forecasts were sufficiently accurate to make a go/no go managerial decision at the time of the forecast and sufficiently accurate to optimize the initial marketing plan. The 27% MAE before adjustment and the 5% MAE after adjustment provide a sufficiently narrow range (given the uncertainty inherent in a really new product) that management can decide whether or not to launch the product. However, we must improve the forecasts to incorporate events such as the

Consumer Reports article. Had additional measures of alternative magazine copy been taken in the IA, we believe that the sensitivity to *Consumer Reports* could have been modeled without the *post hoc* adjustment. In our experience the IA forecasts are a major improvement over managerial judgment in terms of both accuracy and insight provided. Furthermore, IA measurement is synergistic to other marketing measurements such as conjoint analysis. (We return to this issue in a later section.)

Summary of Validation Experience

IA is a relatively new measurement and forecasting technology. We expect that its ability to forecast will improve as more research teams experiment with the methodology. Based on the Reatta study and the CBC study we are confident that simulated information sources that approximate physical environments and human interaction can be created. Based on the camera study it is our judgment that the IA methodology has sufficient external validity for many of the managerial decisions that are based on early premarket forecasts. However, IA is not a panacea. Clear challenges remain and its forecasts must still be used intelligently and with caution by managers.

Lessons from the Field

In this section we share our experience over the last five years during which we used IA to forecast the sales of eight products. Throughout this period we learned new lessons on how IA provides more than just a forecast (lesson 1), how to match the tool to the problem (lessons 2 to 3), and how to make maximum use of the methodology (lessons 4 to 7). Some of these lessons apply to forecasting in general (perhaps marketing models in general), while others are specific to the IA methodology. We hope that by sharing our experience we can begin a productive debate which will lead to improvements in forecasting tools and improvements in practice. We invite the reader to compare our experience with his or her own experience.

<u>Lesson 1. IA focuses a cross-functional team on all aspects of the product launch and decisions.</u>

In order to undertake IA measurement the product team must create the information stimuli. We have found that the process of creating the stimuli enhances communication among members of the cross-functional product team and hence integrates inputs from R&D, engineering, manufacturing, marketing, and finance. Some examples include:

- In the electric vehicle studies, a large product team used the model to carry out strategic analyses of the product launch that required inputs from engineering, marketing, and production. By creating ads early the team focused on a target group of customers "who feel personally responsible for the environment".
- In the camera studies, the product team created a "war room" in which IA stimuli and forecasts were posted on a wall beside the overall project schedule and success factors.
- In the telecommunications studies, engineering, network specialists, and sales staff worked as a team to create an integrated product design. When they selected the stimuli for the IA it became clear to the team that they had not thought through the service benefits that the products would deliver to customers. The IA caused them to focus on the use of the product by customers rather than on the technology of the physical product device.
- Also in the telecommunications studies, whenever the IA analyses forecast sales below initial targets, the product team used the diagnostic information in the IA analyses to improve the design of the product to match market needs.

IA is not the only marketing tool that enhances cross-function teams, but it is an important tool. It is synergistic with other methods such as qualitative research and Quality Function

Deployment (QFD, Griffin and Hauser 1993). For example, in the electric vehicle studies, the product team began the IA after completing a QFD analysis. However, because the IA required a statement on environmental benefits, the team returned to the QFD analyses to examine these issues and improve the design.

As well as increasing communication, the future conditioning aspect of IA enhances the product team's sensitivity to consumer dynamics. Some examples include:

- word-of-mouth feedback based on initial sales of the Buick Reatta,
- improved features and reduced costs in subsequent generations of electric vehicles,
- new infrastructure such as recharging stations for electric vehicles or fiber optics for the communications product,
- competitive entry of new brands (Japanese electric vehicles) and alternative technologies (hybrid electric vehicles), and
- lead users (some medical practices will adapt CBC earlier).

By using IA to simulate the full product line, the firm is better able to evaluate the <u>net</u> profits due to the launch. Here are a few examples:

- The electric vehicle which was simulated in the IA was a two-seated vehicle, but, partially based on the IA forecasts, the firm decided to introduce at a later time a van and a sedan. By including an additional IA search on a van or sedan and by including a conjoint analysis, we were able to make forecasts for the entire product line.
- In the telecommunications study we used a test vs. control design to forecast not only the sales of the new product but its effect on the sales of existing cellular services. (The control cell had only the existing product available; the test cell had both the existing product and the really new product available.)
- In a new microprocessor study, we forecast the sales of the new microprocessor and the sales of existing microprocessors.

Many of these team benefits generalize beyond IA applications (and have been reported in the past, e.g., Lilien, Kotler, Moorthy 1992). We illustrate them here with examples from IA applications so that, combined with other methods in this volume, the field can draw empirical generalizations about the application of marketing models to new product development.

Our internal and external validity tests suggest that IA is sufficient for go/no go decisions on the capital investment, but we have found that managers are unwilling to commit such large sums of money on IA alone. They often have other sources of information such as their own judgments, indications from qualitative market research, econometric forecasts, or product placements with selected customers. These parallel inputs, often unknown to the IA team at the time of measurement, provide confirming or disconfirming managerial information with which to judge the face validity and potential accuracy of the IA. Such multi-source information is important in the risky world of really new products in order to highlight the possibility of (and hopefully minimize) both false rejection and false acceptance.

Lesson 2. IA is most appropriate for risky products which require large capital commitments.

All of the products that we tested required substantial capital commitments -- the smallest was $100 million while the largest was over $1 billion. This is not by accident. With current technology, IA is still expensive ranging from $100,000 to $750,000 per application. Much of this cost is the result of preparing realistic stimuli, although some of the cost still comes from the challenge of multimedia programming. The only firms that have been willing to invest in IA are those firms for whom an accurate forecast is a critical input to a large, risky capital investment. In these cases the value of the information provided by IA justifies its cost. With current technology, IA is unlikely to have a sufficient payback for products with low capital risk (e.g., a new cake mix or shampoo) or products that are not really new. (In some of our applications firms tested products that were not really new in order to test IA. Once they were confident with the methodology, they applied it to really new products.)

Fortunately, the cost of IA is likely to decrease in the future as the cost of multimedia computing decreases. These advancements include new programming tools which will reduce labor costs and more-powerful, less-expensive hardware. The advent of two-way cable systems could eliminate the costly central facility or the need to bring the IA technology to respondents. Rapid prototyping could decrease the costs of the physical prototype products and improvements in computer-aided design and animatics could replace physical prototypes altogether.

Lesson 3. The IA team should provide more than a forecast to the firm.

Although the management team that is close to the product launch decision relies on multiple sources of information, they often use IA to communicate to top management. For example, we have found that IA is quite convincing to top auto managers because the IA represents the future buying environment completely and realistically in the same manner they think of auto buying. In the auto case the IA was so convincing that managers considered using it as a way to actually sell cars, not just evaluate new autos. This face validity is an advantage when the goal is to gain support for a good program. IA can be used to illustrate the critical success factors and to warn of potential threats. In one case the project leader, who personally wanted a go decision, used IA to conduct analysis and recommend to management that the product not be introduced. The IA convinced him to change his decision. This project manager was commended for his analysis and promoted based on this carefully formulated recommendation.

However, IA can be misused. For example, in one of our applications the project leader believed that his career depended upon a go decision so he manipulated the profitability forecast by reducing the marketing variables (and their cost) without changing the forecast. The product was launched and sales were below the forecast that he had reported to top management. It quickly became clear that the product would not make its forecast without the marketing support that was in the original plan.

We believe that successful applications are more likely if (1) the IA team is deeply involved throughout the decision process so that the methodology can be explained at all levels and (2) if top management takes the time and effort to understand the strengths and the weaknesses of the IA forecast and to understand the various contingencies in the forecast.

The first three lessons have dealt with the managerial use of IA. The final four lessons deal with methodological recommendations that reflect current best practice.

Lesson 4. Develop alternative future scenarios.

The initial stimuli in IA attempt to place the respondent in a future simulated purchasing environment. For example, in the electric vehicle study we used future conditioning to present the respondent with a world in which there was more concern about pollution, an improved electric-vehicle infrastructure was in place, and government regulations were favorable. In each of our applications management made its best guess as to the likely future scenario -- we simulated only one scenario. In each case management found it hard to justify the expense (in terms of new stimuli and increased sample size) to simulate alternative future scenarios. As a result, we have had to add questions to the end of the IA in order to measure the sensitivity of the judged probabilities to changes in the future conditions.

While this has proven adequate, we feel that the forecasts can be improved with a variety of future conditions. For example, had the camera IA simulated a more negative *Consumer Reports* article, the camera firm might have been ready to respond quickly when the article appeared in May 1994. Multiple future scenarios would enable management to use robust designs to develop a product and marketing plan that does well in most scenarios or the most-likely scenario (Taguchi and Clausing 1990).

Lesson 5. Select the control product carefully.

The control product serves two purposes in IA. (1) The sales forecast is indexed to the control. (2) Market response parameters for advertising, word-of-mouth, distribution, etc. are based on the control. We have found that the usefulness of IA is enhanced when a good control product is available. For example, the on-line service, Prodigy, was an excellent control for a new home shopping service. On the other hand, cellular telephone service was not as good a control for a new digital mobile service. Cellular service was acceptable as an index to the sales forecast, but there were so many differences in the early life cycle (e.g., high prices and limited service areas) that it was difficult to estimate the diffusion and word-or-mouth parameters for the new digital service.

Lesson 6. Combine IA with other marketing models such as conjoint analysis or logit analysis.

Because IA simulates the future environment and provides full, realistic product information to consumers through an active search of information sources, the base-line forecast is likely to be more accurate than would be possible by conjoint analysis alone. On the other hand, the cost of IA makes it prohibitive to have different test cells for all of the variations in features that the product team is likely to consider. In seven of the eight IAs the sponsor requested, and we provided, either a conjoint analysis or a discrete choice measurement at the end of the IA. In these cases an IA stimulus was matched to one of the conjoint analysis (or discrete choice) stimuli. This combined use enhanced the face validity of the resulting forecasts. With the indexed forecasts we were able to use the conjoint analysis (or logit analysis) capabilities to simulate product changes and competitive response.

Lesson 7. Two hours is a feasible interview time.

Some of the eight IAs used a 1-1/2 hour interview time and some used a 2-hour interview time. We have found that the multimedia computing environment is sufficiently engaging to respondents that they remain interested for a full 2 hours. The variety of stimuli in a vivid format with easy-to-answer questions maintains the respondents' attention. However, we have had to use substantial incentives to recruit respondents -- $50-$75 for consumer product interviews and $150-$200 for business-to-business interviews.

Computer literacy has not been a problem. Even for the lower-priced mass market products we have found that, with 10 minutes of training time, over 95% of the respondents can participate in an IA survey. The ability to use a mouse and respond to questions on the screen is a simple task for most people. In addition, we collect qualitative information by having respondents speak directly into a microphone attached to the computer. They find this mode of input more natural and easier than typing.

Summary

This paper reports on two internal validations and one external validation of forecasts made with IA. The internal validations, which compare (1) a computer-simulated automobile showroom to a physical (simulated) automobile showroom and (2) a computer-simulated medical technician to interactions with a real technician, suggest that a multimedia computer can portray information sources with a high degree of realism. Forecasts based on the computer-simulated stimuli are not significantly different than the more-traditional laboratory stimuli. The external validation suggests that IA has the potential to forecast actual sales, particularly if the firm sticks with its marketing plan and if unanticipated events such as a negative *Consumer Reports* article do not occur. The comparison of the adjusted forecasts to actual sales suggests that IA has the potential to forecast the effects of changes in the marketing plan (e.g., changes in prices and advertising).

Despite these initial indications of success, IA is still a developing technology. Many challenges remain. For example, the test vs. control design assumes that, given full information, stable preferences can be measured. Although the initial external validation covered two years into the future, forecasts are often made for five years into the future. Thus, the assumption of stability requires further testing.

Another important challenge is modeling the learning that takes place as consumers gain experience with the product and invent new uses. The use of VCRs for time shifting television programs, the change in cooking patterns based on the availability of microwave ovens, and the varied uses of personal computers are examples of in-use learning. In order to predict the sales of these new products we must also predict the impact of these new uses. Perhaps future research will develop a "learning accelerator" to model this process.

References

Green, Paul E. and V. Srinivasan (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice," Journal of Marketing, 54, (October), 3-19.

Griffin, Abbie, and John R. Hauser (1993), "The Voice of the Customer," Marketing Science, 12 (Winter), 1-27.

Hauser, John R., Glen L. Urban, and Bruce D. Weinberg (1993), "How Consumers Allocate Their Time When Searching for Information," Journal of Marketing Research, 30 (November), 452-66.

Jamieson, Linda F. and Frank M. Bass (1989), "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods," Journal of Marketing Research, 26, (August), 336-45.

Juster, Frank T. (1966), "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design," Journal of the American Statistical Association, 61, 658-696.

Kalwani, Manohar U. and Alvin J. Silk (1983), "On the Reliability and Predictive Validity of Purchase Intention Measures," Marketing Science, 61 (September), 243-286.

Lilien, Gary L., Philip Kotler, K. Srihar Moorthy (1992), Marketing Models, (Englewood Cliffs, NJ: Prentice-Hall, Inc.).

Morrison, Donald G. (1979), "Purchase Intentions and Purchase Behavior, Journal of Marketing, 43 (Spring), 65-74.

Taguchi, G. and Don Clausing (1990), "Robust Quality," Harvard Business Review, (68), (January-February), 65-75.

Urban, Glen L., John R. Hauser, and John H. Roberts (1990), "Prelaunch Forecasting of New Automobiles," Management Science, 36 (April), 401-21.

----, and Gerry M. Katz (1983), "Pre-test-Market Models: Validation and Managerial Implications," Journal of Marketing Research, 20 (August), 221-34.

----, William J. Qualls, and Jonathan D. Bohlmann (1994), "Information Acceleration of High Tech Industrial Products: A Market Feasibility Test," Working Paper, MIT Sloan School, Cambridge, MA 02142.

----, Bruce D. Weinberg, and John R. Hauser (1995), "Premarket Forecasting of Really New Products," Journal of Marketing, 59, forthcoming.

Weinberg, Bruce D. (1992), "An Information-Acceleration-Based Methodology for Developing Preproduction Forecasts for Durable Goods: Design, Development, and Initial Validation," Unpublished Ph.D. Thesis, MIT Sloan School of Management, Cambridge, MA 02142.

Table 1. Sample sizes in auto study.

		Showroom	
		Computer	Physical
Automobile	Buick Reatta	71	43
	Mazda RX-7	40	23

Table 2
Analyses of Variance

SOURCE OF VARIATION	FULL-INFORM. PROBABILITY			CHANGE IN PROBABILITY		
	D.F.	MEAN SQ.	F-STAT.	D.F.	MEAN SQ.	F-STAT.
Main Effects						
Video vs. Physical	1	144.3	0.22	1	10.4	0.05
Reatta vs. RX-7	1	2835.1	4.33	1	653.8	2.96
Interaction	1	392.7	0.60	1	0.1	0.00
Residual	171	655.4	-	173	220.8	-

Table 3. CBC Experimental Results.

		Technician Interaction	
		Actual	Simulated
Judged Purchase Probability (Physician)	Forced Exposure to the Concept	35.4%	37.2%
	Final	31.9%	29.5%

Table 4
Rated Realism and Influence of the Information Sources by Physicians
(Depending on assigned condition either the simulated or actual technician was rated.)

Information Source	Realism	Influence
Physician Colleague	5.8	5.9
Product Brochure	5.1	4.3
Magazine Advertisement	5.5	4.3
Sales Presentation	5.5	5.1
Journal Article	6.0	5.7
Accountant Memorandum	5.8	6.0
Simulated Technician	6.1	4.9
Actual Technician	na	5.8

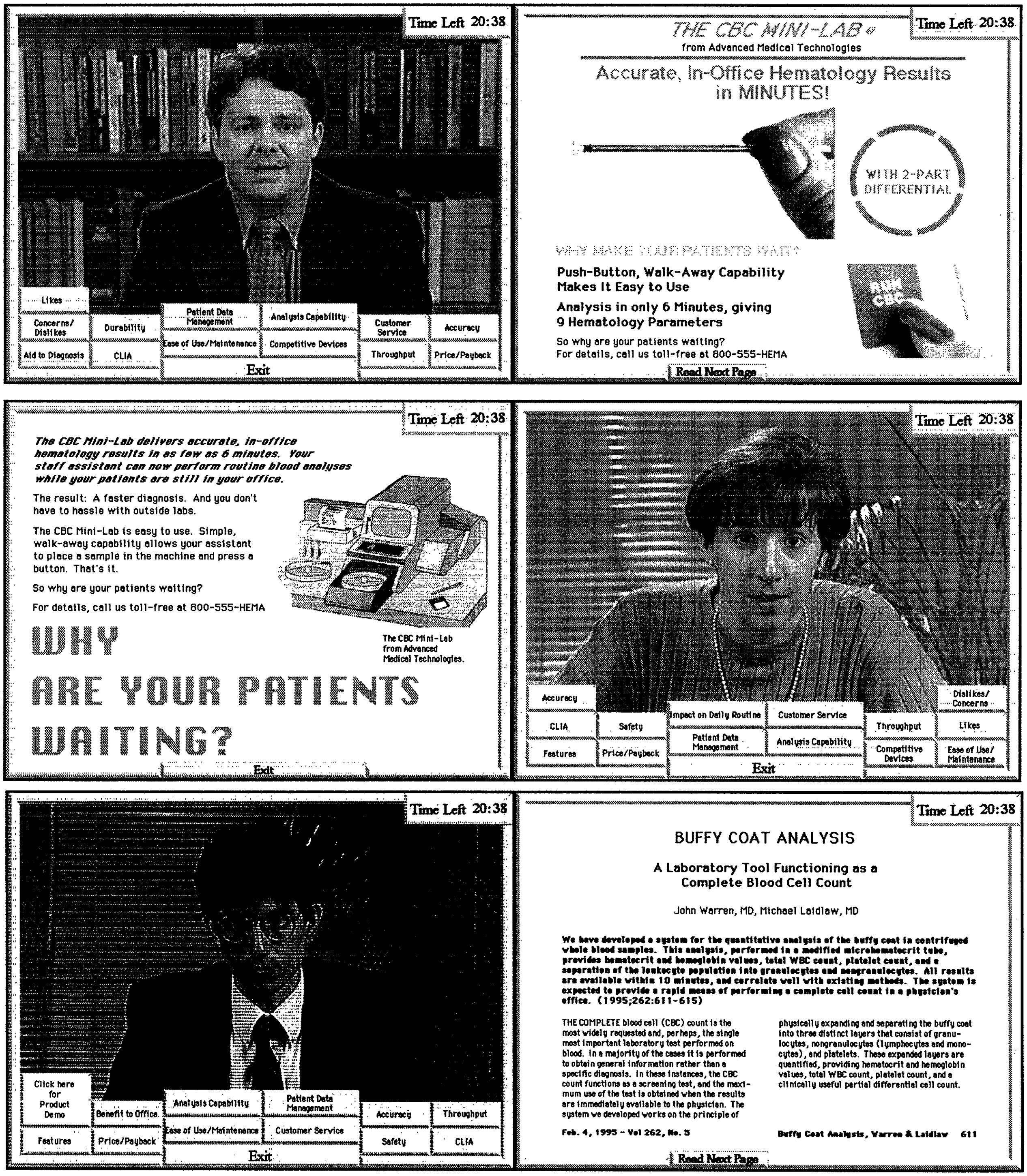

Figure 1: Example Screens from Medical Equipment Information Accelerator. (a) Physician Colleague, (b) Product Brochure, (c) Magazine Advertisement, (d) Simulated Medical Technician, (e) Salesperson, (f) Medical Journal Article.

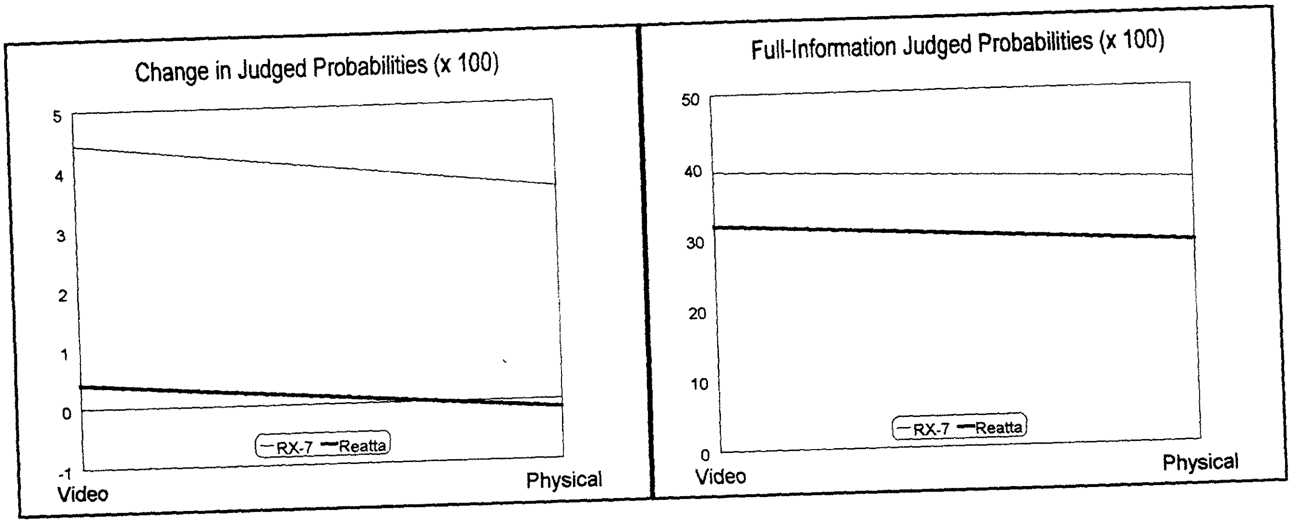

Figure 2
Comparing the Effects of a Video vs. Physical Showroom

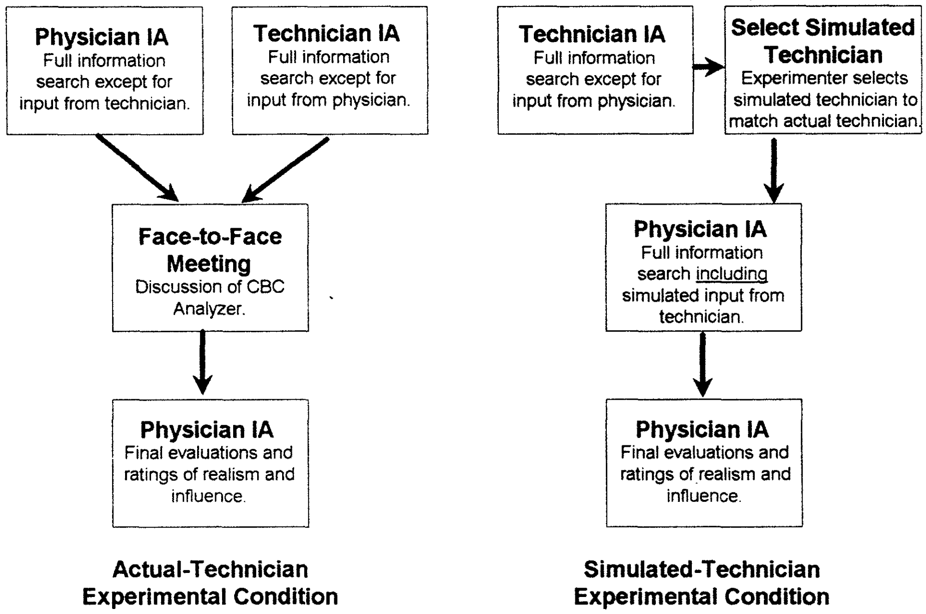

Figure 3
Experimental Design for CBC Analyzer Internal Validation

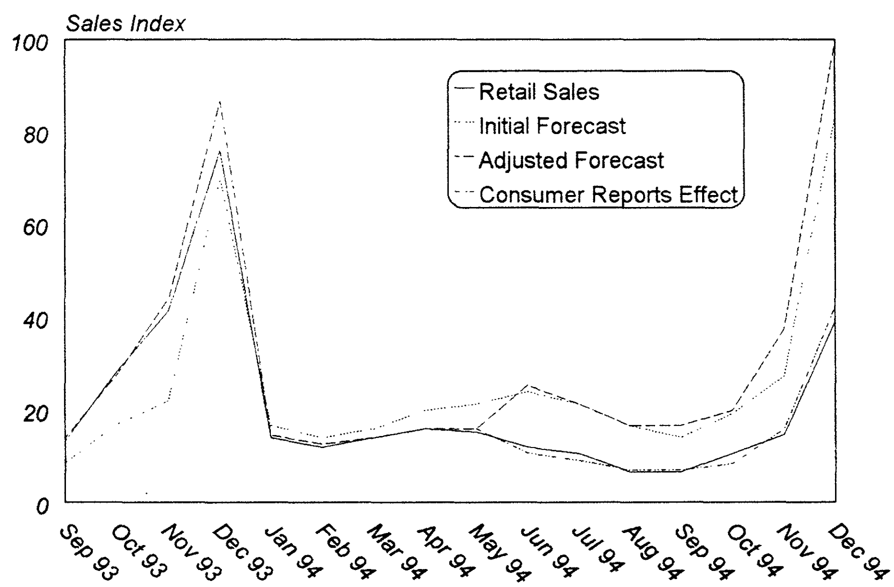

Figure 4
External Validation of New Camera Forecasts