The International Center for Research on the Management of Technology

Validation and Lessons From the Field -- Applications of Information Acceleration

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

January 1996
WP # 141-96

Sloan WP # 3882

Sloan School of Management
Massachusetts Institute of Technology
38 Memorial Drive, E56-390
Cambridge, MA 02139

ABSTRACT

There is strong management interest in the use of multimedia stimuli to gather data with which to forecast consumer response to really new products. These vivid methods have high face validity and are attractive to top management, but these methods have only begun to be tested for validity. In this paper we evaluate one virtual representation called information acceleration (IA) that has been applied eight times to consumer and business-to-business products. We report on three tests of validity -- two internal and one external. The first internal test compares the ability of IA to represent a physical automobile showroom and salesperson. The second internal test compares the ability of IA to represent the interpersonal interaction of a medical technician with a physician when evaluating a new medical instrument. The external test compares forecasts, made in 1992 for a camera launched in 1993, with actual sales for 1993 and 1994. We also compare actual sales to forecasts modified for the actual marketing plan and 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 real-world applications of IA.

Acknowledgment

The work reported on in this paper was funded, in part, by the International Center for Research on the Management of Technology (ICRMOT) at the Sloan School of Management, Massachusetts Institute of Technology. The authors gratefully acknowledge this support.

Introduction

Faced with a need to revitalize through new products, many firms are exploring ways to make their new product development process more effective and more efficient. In this paper we focus on one aspect of their efforts to understand and forecast consumer response better. Specifically, we explore the ability of multimedia computer representations to represent products, people, and situations. Early computer simulations of choice began in the 1980s (Brucks 1985, 1988, and Johnson, Payne, Schkade, and Bettman 1986) and have expanded into virtual shopping in the 1990s.

There is strong management interest in the use of multimedia representations for understanding and forecasting consumer response. Multimedia representations of potential new products can be developed earlier, quicker, and with less cost than actual prototypes. If forecasts based on multimedia representations are sufficiently accurate and provide sufficient diagnostic information for product improvement, then products can be developed in less time with less risk and with a greater potential for profit. Many managers believe that the multimedia representations are more vivid and more realistic than traditional concept descriptions and, hence, provide a useful middle ground between traditional concept descriptions and actual physical prototypes.

In light of this managerial interest, it is not surprising that academics and practitioners have developed and used creative representations of new products. For example, Burke, et. al. (1992) created a computer-simulated supermarket and showed that purchases of existing products made by consumers over a seven month period were reasonably valid when compared to actual market shares. Researchers and consultants have used computer attribute response simulation to design products (e.g. Green and Srinivasan 1990, Moskowitz 1995) and consumer laboratory prototypes to test services with consumers (e.g. Santosus 1994). For this paper we study on one virtual representation, called information acceleration (IA), which has been used to test new consumer durable and business-to-business products (and the associated marketing campaign). For example, Urban, Weinberg, and Hauser (1996) describe how IA was used to create a virtual showroom for an electric vehicle in which the potential consumer could "walk" around the car, "climb in," and discuss the car with a salesperson. The customer could access television advertising and consumer magazine articles, read prices in a virtual newspaper, and even get advice from fellow consumers -- all simulated on a multimedia computer. When supplemented by a test drive of a prototype vehicle, this procedure allowed forecasting of the life cycle sales for the electric vehicle.

While IA has had a commercial impact, there have been no published validations. This paper seeks to explore the validity of some aspects of IA and review lessons learned from field applications.

Validation

IA and other multimedia representations have their greatest managerial potential for really new products. However, before relying on IA for really new products where a billion dollars might be at stake (as was the case with electric vehicles), firms wanted to test IA with new, but not really new, products. After all, data to validate the electric vehicle forecasts are unlikely to be available prior to the year 2000. Thus, we address validation in steps. See figure 1.

Step 1. We begin by comparing the ability of IA to represent an automobile showroom. By comparing forecasts based on a multimedia showroom to those based on an actual car in a physical showroom with an actual salesperson present, we test the ability of IA to represent physical prototypes.

Step 2. We next compare the ability of IA to represent a medical technician as part of a physician-technician buying dyad. This enables us to test whether IA can represent the information provided by other people in the buying process.

Step 3. Once we establish that the components of the multimedia representation are valid representations of physical products and human interaction, we test the ability of a complete IA to forecast actual sales of a new product.

Because steps 1 and 2 do not depend upon external factors nor on the realized marketing plan, we call steps 1 and 2 internal validations. We call step 3 external validation.

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 sufficiently new to Buick that Buick wanted to test the ability of IA to represent the Reatta. Prior to the development of IA Buick had used traditional automobile "clinics" where respondents viewed and drove preproduction cars, watched videotapes of advertising and simulated word-of-mouth, and read simulated magazine articles. Not only were clinics expensive to run, but Buick wanted to develop a methodology that would enable them to make forecasts before investing in a pilot plant to build preproduction cars and before building a sufficient quantity of hand-built prototypes.

Following standard practice in obtaining clinic-based measures, we obtained measures both for the new product, the Reatta, and for a "control" product, the Mazda RX-7. In actual forecasts we compare clinic measures for the new product to clinic measures for the control product and use probability-flow models to incorporate phenomena such as advertising, word-of-mouth, dealer availability, life cycle, and industry volume. See Urban, Hauser, and Roberts 1990 for examples and details. For the purposes of validation we use the control car to establish that the IA can measure significant differences between the Reatta and the RX-7. Establishing significant differences among cars is important because we interpret as internal validation the fact that there is no significant difference between a multimedia showroom and a physical showroom.

The information 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*. For the *showroom,* one set of consumers had a chance to walk around the car, sit in the car, and ask a salesperson questions. For another set of consumers this opportunity was simulated on a multimedia computer. For both sets of respondents all other aspects of the IA experience were held constant. Respondents are allowed to explore information sources in any order, to revisit information sources, and to spend as much time on any information source as they want (subject to an overall time limit). 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 sporty car as their next car and whether they were willing to spend $20,000 or more on the purchase. They were not screened on any computer experience. The goal was to get typical consumers for the Reatta's market segment. 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 multimedia-based showroom: (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. 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 second internal validation focuses on a business-to-business product for which the purchase decision involved more than a single decision maker. The product was a new technology for a Complete Blood Cell (CBC) count analyzer. The primary decision participants for this medical instrument 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.

A CBC 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 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 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. (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. The new technology changes usage patterns because it makes it feasible to do comprehensive testing in the physician's office.)

So that they might evaluate the CBC in the (future) context in which it would be sold, each respondent was given information about expected health care reforms, federal regulations, and

Medicare reimbursements. The information about the CBC analyzer that was simulated in the IA was chosen based on focus groups, questionnaires, and interviews with CBC manufacturers. Some examples of this information are given in Figure 3. These include: information from *product brochure* or *magazine advertisements*; *medical journal articles*; *sales presentations* (video of a salesperson demonstrating the product); *information from a colleague* (video of a person portraying a physician or technician who shares his or her experiences - a participating physician could access a physician; a participating technician could access a technician); *cost analyses by an accountant* (not shown). *Staff discussions* were also available as an information source. To test internal validity we allowed some physicians to come together face-to-face with their staff while for other physicians the staff discussions were simulated on the computer. For more details on the information sources see Urban, Qualls, and Bohlmann (1994).

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 4.)

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.

Sample 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.

Prior to the experiment the CBC analyzer was sufficiently new that respondents had no knowledge of its features. Thus, we needed to provide some initial information to respondents so that they could decide whether they wanted to search for more information. To do this we asked them which information source they were most likely to use to gather information on new medical equipment. They then began the IA by receiving information from that source (e.g., sales presentation, medical journal articles, product brochures). We call this induction "concept forced exposure." After this initial exposure, they were allowed to search all subsequent information sources in any order if they so chose.

In both the simulated and actual conditions the probability as judged by doctors and technicians declined with full information. (This emphasizes the danger of overly optimistic forecasts when only initial concept exposures are used. Subsequent information sources can reveal features that lower the perceived benefit of the product). 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.

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 concept 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 probability, concept forced exposure probability, and final judged probability, but these differences 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 a better 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 1996). In that partial validation, IA was used to forecast the sales of a new, but not really new, camera. The validation was only partial because the new camera was already on the market. Awareness, distribution, word-of-mouth, and other marketing variables were known to both management and the forecast team. Sales were known to management but not the forecast team. As reported, forecasts in this internal validation were within 10% during the first and second years. This gave management confidence that forecasts could be made using multimedia-based measures. With this confidence, management decided to use IA to forecast the sales of a camera with a novel film format and film handling capability. Management believed that this really new camera was a substantial improvement at its price point.

Premarket Forecast of a Really New Camera

Standard IA methodology was used. The information sources included television advertising, a simulated mass merchandise store environment, simulated word-of-mouth communication, and a simulated consumer magazine article. The sample consisted of 671 respondents chosen as representative of the new camera's market segment. Of these, 202 were given information on the current camera product line and 469 were given information on 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). Forecasts, made in 1992 for a camera that was introduced in August 1993, were based a test-vs.-control design supplemented with a probability-flow model that accounted for the facts (1) that some respondents gain awareness before going to the store and some gain awareness at the store and (2) cameras are purchased for both personal use and for gifts -- especially in December.

The camera was introduced and we now have data from the first sixteen months of sales. However, before we compare the forecast sales to actual sales we first describe the events that took place during those months. Specifically, the firm changed its marketing plan and *Consumer Reports* published an unexpected article about the camera. (This article was not part of a normal review of cameras; it was a special article on this one new camera.)

Differences Between the Marketing Plan in the Initial Forecast and the 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. This new probability of becoming aware is then used in the probability flow model to make the new forecast. Similarly we use the discrete-choice pricing model to predict the effect of the price change.

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 the *Consumer Reports* article, but we believe 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 the best we can do given data that could have been collected as soon as the article became available.

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. These post-event estimates provide data for future IA-based forecasts.

Comparison of Actual Sales and Forecasts

Figure 5 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 5 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 post-May 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 may also act as a "recalibration" factor to account for cumulative unobserved errors prior to June 1994. Thus, 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 *Consumer Reports* event, but not thereafter. After statistically adjusting for the negative *Consumer Reports* event 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 or, at least, with less *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

Multimedia-based stimuli provide the basis for new measurement and forecasting methods. Based on the Reatta study and the CBC study we are confident that simulated information sources can be created to approximate physical environments and human interaction. We expect that the ability of such stimuli to represent products and human interaction will improve as more research teams experiment with these stimuli. Based on the camera study we believe that the IA methodology (multimedia stimuli, test-vs.-control measures, and probability-flow models) has sufficient external validity for many of the managerial decisions that are based on early premarket forecasts. However, multimedia stimuli, and their use in IA, are not a panacea. Clear challenges remain and 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 have used IA with multimedia stimuli to forecast the sales of eight products. In the previous section we interpreted three quantitative studies that attempted to infer the applicability of multimedia stimuli for forecasting the sales of really new products. This section is very different. In this section we summarize what we have learned from field experience. We attempt to do this through examples, but we caution the reader that these are our impressions. We invite the reader to compare our experience with his or her own experience and we invite productive debate on the future of multimedia methods.

Lesson 1. Developing Multimedia Stimuli Helps to Focus Cross-Functional Teams

Our first lesson applies to forecasting in general, perhaps marketing models in general. These benefits have been reported in the past, e.g., Lilien, Kotler, and Moorthy (1992). We illustrate them here with examples 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.

We have found that the challenge of creating realistic information-source stimuli forces the new product team to integrate inputs from a variety of functional areas. Some examples include:

- In the electric vehicle study, the creation of the stimuli forced the team to develop advertising concepts earlier than they otherwise would have. This caused them to define carefully their target group of consumers.

- In the telecommunications studies, engineering, network specialists, and sales staff worked as a team to create an integrated product design so that they could develop the stimuli. In this process it became clear that they had not thought through the service benefits that the products would deliver to their customers.

In order to simulate a future environment the team must agree on the implications of that future environment. Some examples include:

- The need to simulate the impact of new infrastructure (recharging stations for electric vehicles or fiber optics for communications products).
- The need to simulate future generations of products required technology assessment (batteries for electric vehicles), competitive forecasting (the entry of Japanese vehicles), and generic alternatives (hybrid electric vehicles).

In order to simulate one product, the team had to plan for the entire product line including new cameras (camera study), vans, sedans, and two-seaters (electric vehicles), and cannibalization of cellular telephone sales (communications product).

<u>Lesson 2. With today's multimedia technology, IA is most appropriate for risky products which require large capital commitments.</u>

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).

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 and/or the internet 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. IA and multimedia tools help the product development team to communicate with top management

Multimedia representations have strong face validity and are vivid. Product development teams have found that top management is comfortable with IA's ability to represent the future buying environment realistically. This face validity has proven 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 used IA to convince management that a product not be introduced. He was commended for his analysis and promoted.

The first three lessons deal with the managerial use of IA. The final four lessons reflect our recommendations for improving 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-of-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 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 of multimedia stimuli and one external validation of forecasts made with multimedia stimuli and 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, 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.

References

Brucks, Merrie (1985), "The Effects of Product Class Knowledge on Information Search Behavior," *Journal of Consumer Research,* 12 (June), 1-16.

Brucks, Merrie (1988), "Search Monitor: An Approach for Computer-controlled Experiments Involving Consumer Information Search," *Journal of Consumer Research*, 15 (June), 117-121.

Burke, Raymond R. (1993), "Virtual Shopping: A New Tool for Consumer Research," Harvard Business School working paper, Boston, MA. September.

-----, Bari A. Harlam, Barbara E. Kahn and Leonard M. Lodish (1992), "Comparing Dynamic Consumer Choice in Real and Computer-simulated Environments," *Journal of Consumer Research,* 19 (June), 71-82.

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.

Imanura, Kayo, Junji Nomura, Hikaru Ohata, Robert J. Schultz (1992), "Virtual Space Decision Support System (VSDSS) and The Production/Sales Unification System Applied to It," IEEE International Workshop on Robot and Human Communication, 0-7803-0753-4/92.

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.

Moskowitz, Howard (1995), *Food Concepts and Products: Just in Time Development*, (Trumbull, CT: Food and Nutrition Press, Inc.).

Santosus, Megan (1994), "The AT&T Test Kitchen," *CIO*, (May), 36-48.

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

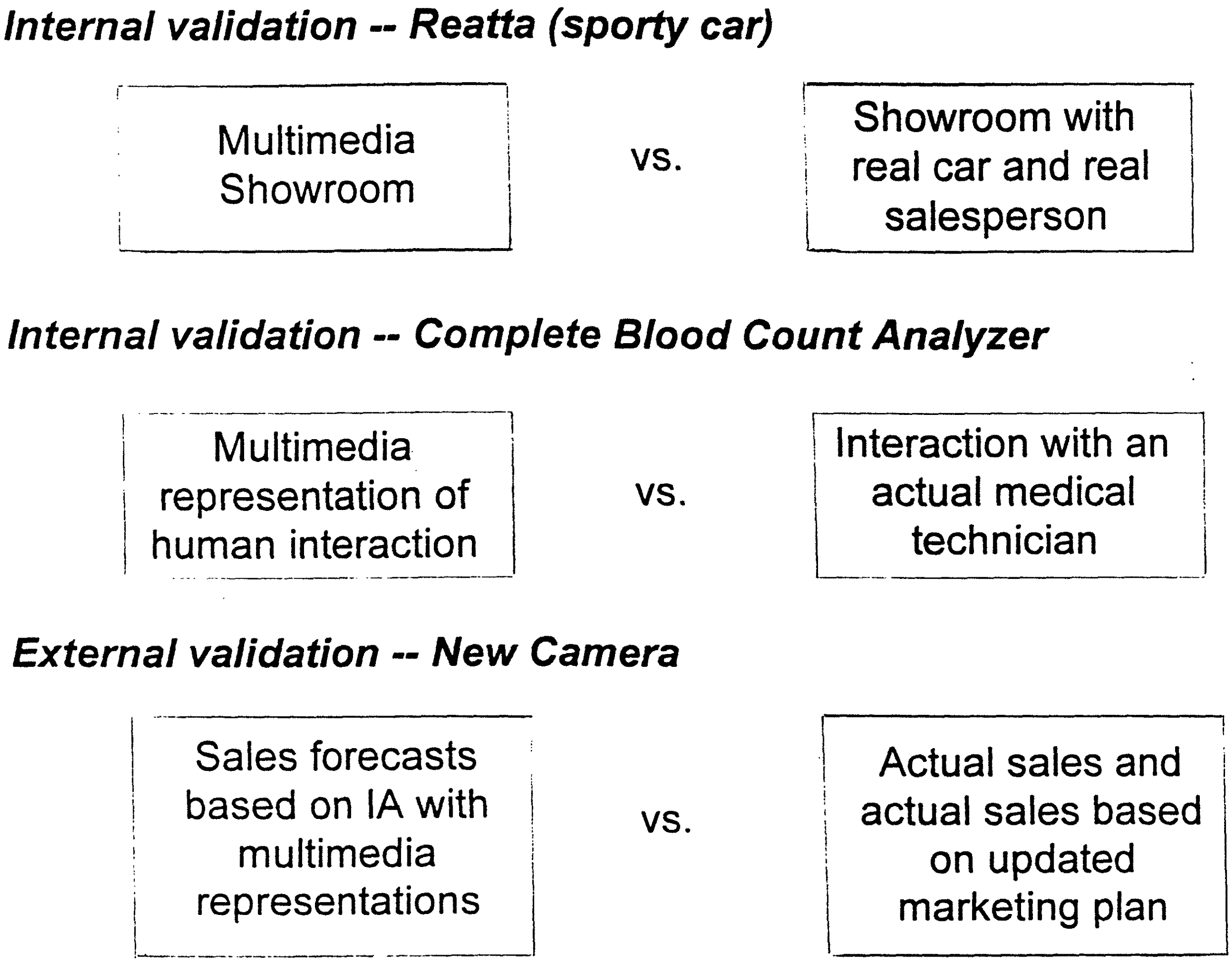

Figure 1. Three related validations.

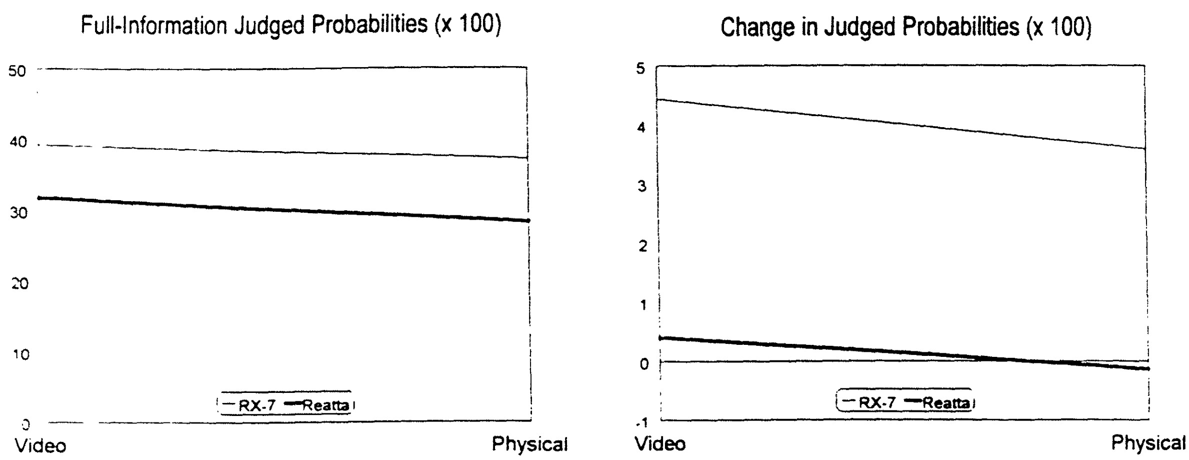

Figure 2
Comparing the Effects of a Video vs. Physical Showroom

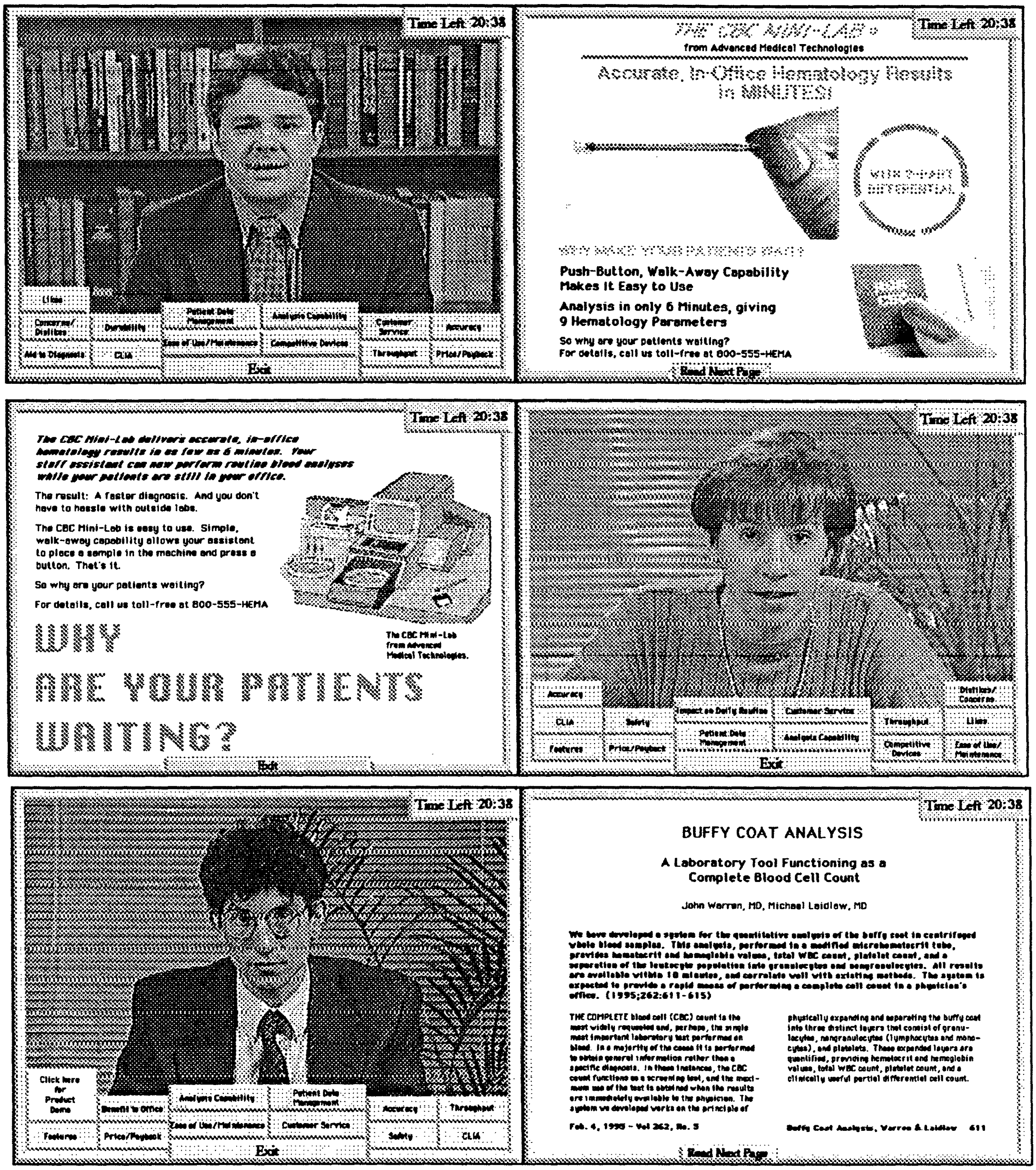

Figure 3: Example Screens from Medical Equipment Information Accelerator. (a) Physician Colleague, (b) Product Brochure, (c) Magazine Advertisement, (d) Simulated Medical Technician, (e) Salesperson, (f) Medical Journal Article.

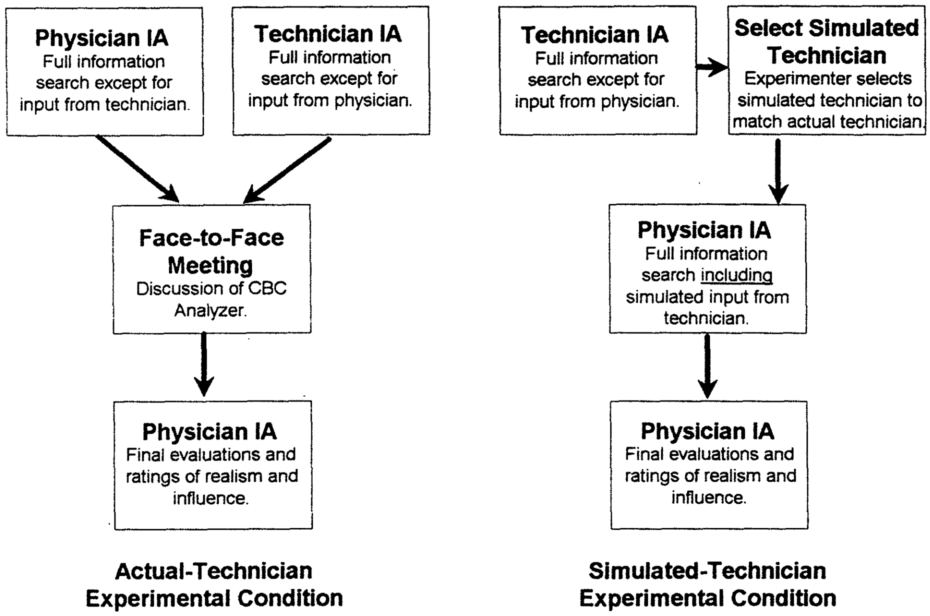

Figure 4
Experimental Design for CBC Analyzer Internal Validation

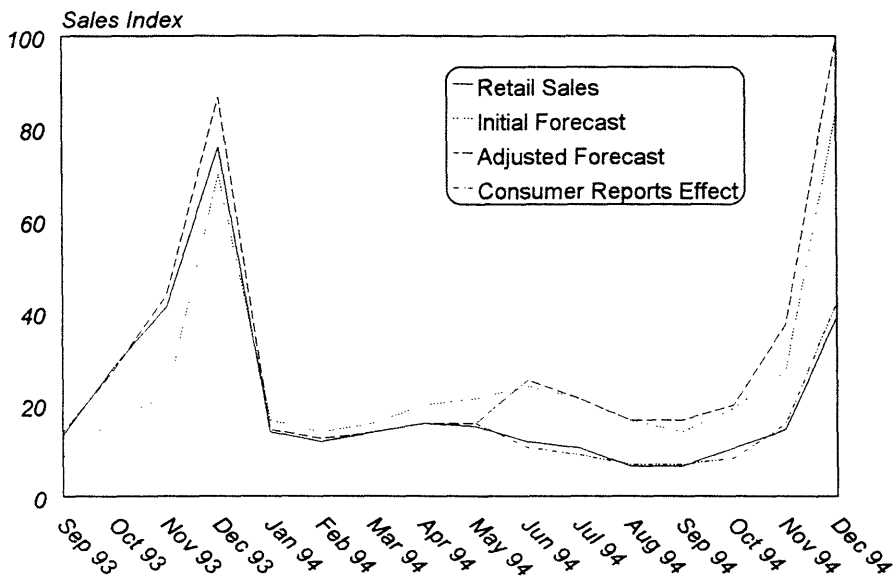

Figure 5
External Validation of New Camera Forecasts