Quantitative Early-Phase User Research Methods: Hard Data for Initial Product Design

Christopher N. Chapman, Microsoft Corporation
Edwin Love, University of Washington Business School
James L. Alford, Volt Information Sciences

Abstract

We describe questions that commonly arise in early-phase user research for new technology products concerning customer needs, priorities, and market definition. We suggest that methods adopted from marketing research, statistics, and game theory may be helpful for user researchers to answer those questions. We show how these methods have been applied to real problems and decisions for a new product line at Microsoft. These methods are especially appropriate for HCI professionals because they require solid experience with experimental research and statistical methodology and complement other user research tools. The methods may be most effective when combined with detailed research on user tasks, goals, and interaction models. When research is synthesized in this way, it can make a strong contribution to product definition and business strategy.

1. Introduction: Problems of early-phase user research

Technology companies are gradually moving away from a purely technology-centered approach to development and are embracing the ideas of user-centered design and human-computer interaction (HCI) approaches such as usability engineering [18][1]. It is now commonplace within technology firms for engineers and managers at every level to inquire about customer needs, or to postulate ideas in terms of customer scenarios and value propositions.

When organizations are customer-focused and have credible internal sources of user information, they ask increasingly more of their HCI and user specialists. This has led some technology organizations to expand the role of usability engineering and to cast it instead as user research [11]. User research is an expansion and generalization of usability engineering, where the goal is not merely to study and optimize interaction design, but to know as much as possible about users and their needs in general. This includes understanding customers' unmet needs, their preferences among potential solutions, and the specifics of how they would interact with those solutions.

This focus on end-to-end customer needs has brought together two disciplines that have historically worked rather independently: usability engineering (or user research) and market research [13]. Traditionally, market research has determined the general direction of development by specifying marketing goals and products, while usability engineering has worked to optimize the behavioral interactions between a particular product and its users. This is easiest to see in terms of a product lifecycle: the province of market research has been the earliest stage of development (product selection) and the latest stage (sales), while the domain of usability engineering has been what comes between (development and engineering). However, when user research is considered holistically, there is intersection throughout the lifecycle. Understanding user interaction should inform the earliest stages of product and market selection.

In its traditional form, neither user research nor market research is able completely to fill the needed role. User research, especially in its role doing usability engineering, may be close to the engineering teams, but may not possess the methodological foundation and tools to conduct rigorous early-phase research. For example, to determine early user needs, user research typically applies qualitative methods such as focus groups and field observation [11]. Such research is valuable, yet incomplete. Market research may possess quantitative tools to address the needs, but does not have the complete view of user behavior that user research has. Thus, market research may not even know what the problems are or how to address them. It is only when one unites the two disciplines that one is able to delineate a complete, end to end discipline of rigorous user research.

Our goal is to outline the specific research questions that often arise in early product research and to describe quantitative methods that may be able to address those questions. In the course of doing this, we demonstrate case material in which these methods have been applied to a product line at Microsoft. Throughout the paper, we discuss early-stage research conducted for Microsoft in the process of establishing a new product line of consumer webcams. We take it for granted that such research should use methods that are describable, quantifiable, replicable, and defensible. In short, this research should be scientific.

We describe four general methods that are helpful at different stages of the early development cycle: ranking methods that assist with early concept selection; conjoint analysis, which helps with feature selection and product refinement after a concept is selected; customer definition methods to identify target users and markets for the defined product; and a game theoretic approach to examining business strategy.

Our discussion of each specific method is cursory; our goal is to give an overview of how various research fits together over the course of early product development, rather than to give tutorials on individual methods. We wish to inspire HCI practitioners to learn more about potentially useful methods that may be new to them.

2. Methods for early-phase research

2. 1 Concept selection

A product lifecycle begins when a firm develops potential product concepts through such means as technological innovation, brainstorming, market observation, or qualitative user research. At this point, the product team needs customer data in order to choose among various concepts. Which do users want most? In our experience, the most commonly applied research methods are unitary metrics of desirability, along with stack ranking of concepts. Users are asked to score multiple concepts on a Likert-type scale for desirability or some other attribute; or they are asked to stack-rank their concepts.

The problem is that absolute scales and stack-ranks do not encompass the entire question about preference among the given concepts. Unitary ratings do not indicate a tradeoff among concepts; one may be slightly more "preferable" or "desirable" or "important" or "likely to be purchased", but what does a Likert-type score really mean for comparing customers' choices? If several choices are all "important", should one choose among them on the basis of an average Likert score? Likewise, a stack-ranking can yield an ordered list, but how significant is the difference between #1 and #2, or #1 and #3? If it turns out that #1 is too difficult to develop, will #2 be nearly as good, or is it markedly inferior? One needs a method that (a) forces user tradeoffs among choices; (b) yields a comparative ranking of concepts; and (c) allows comparison of absolute preference among the choices.

One method that fills this need is the Maximum Difference ("MaxDiff") rating method [12][4][3]. In MaxDiff, respondents repeatedly select their most and least preferred choices (or important, useful, or any other adjective) from short lists of concepts (drawn from a longer list). This is a generalization of paired comparison methods, and yields a stable ranking of K concepts, presented L at a time, after approximately 3K/L user choices [16]. We have commonly used MaxDiff to score 15-35 items, although the method is applicable to lists of arbitrary length. Scores are estimated at the individual respondent level for each item using a Hierarchical Bayes model, and are then averaged across individuals for overall item scores [17]. The results are appropriate both for comparisons among items and among groups of individuals [3].

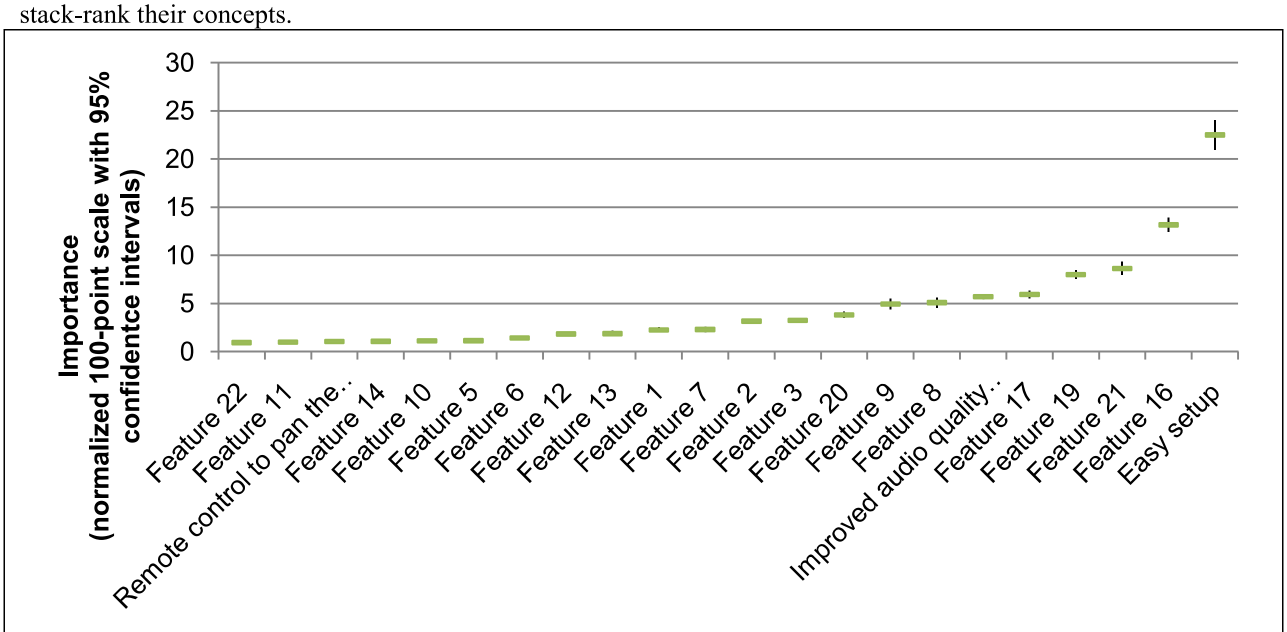

Figure 1: Webcam concept/feature importance ratings

In Figure 1 we show the result of a MaxDiff assessment of webcam features conducted with 1008 online respondents in the US (most features are masked for product confidentiality). The interest level in a given feature may be read on the Y axis. For instance, “Easy Setup” received an average importance score of 23 of a total 100 points, with a 95% confidence interval ranging 21-25. Overall, a small number of features were highly desired by respondents, while there was a long list of features that garnered almost no interest. From this data, we know not only the stack ranking of concepts, but also the absolute magnitude of each, allowing both absolute and relative comparisons.

When interpreting MaxDiff, it is helpful to consider both the confidence intervals around individual data points and the absolute magnitude of data points. The confidence intervals (as plotted in Figure 1, although too small to distinguish for some concepts) can help in determining whether observed average preference levels show a true distinction, and help to determine groupings among data points. The absolute magnitude is important because MaxDiff is standardized to an absolute 100-point scale. This means that random data alone should yield scores that average 100/K, where K is the number of concepts. If a concept’s observed average score is not significantly greater than 100/K (e.g., a score of 4.5 in Figure 1), then it would be difficult to argue that the concept is important, unless one has additional data or rationale.

The MaxDiff method can be applied generally to any list where items are represented by a single brief description and presented in a tradeoff task with respect to a defined metric. We have successfully used the method to gather data on feature preference, use case importance, and general product descriptions. At Microsoft, this method – applied several times, to various markets and sets of issues – gave us confidence that we had correctly prioritized user needs for our new line of webcam products.

2.2 Product refinement

After use cases and features have been prioritized through concept selection, one may turn to the specific definition of one or more products. This poses new research questions, such as the interaction of product attributes and how to maximize the customer’s perception of value within a given development budget or product cost. An appropriate method to use for these questions is traditional choice-based conjoint (CBC), in which products are defined in terms of feature attributes and levels [8][9]. CBC typically requires respondents to make trade-offs among fully-specified products in relation to price, and yields estimates of the importance of individual features and the value that customers perceive for each feature. The tasks present randomized lineups of hypothetical products. On each task, a respondent selects the product that he or she prefers. Multiple randomized tasks are performed for each respondent. An example of a single CBC task is shown in Table 1.

In our experience, CBC is very helpful to establish the answer to three questions: what are the features that a product must have, as opposed to those that are of less importance? How much value do customers assign to a given feature or feature level? How would one expect a given product to do against its competitors with a different set of features?

Conjoint analysis yields several kinds of information, including effect size estimates for both main and interaction effects (which are appropriate for assessment by statistical tests such as Chi-square), and choice proportions for each level of main effects and interactions. Choice proportions show the overall likelihood that a given main or interaction level was chosen when presented, across all of the randomized choice tasks, and represent a simple overall metric of importance for an individual feature level. We find that choice proportions are both informative and easy to communicate to product teams and management.

Table 1: Sample Choice-Based Conjoint Task

If these were your only options, which product would you purchase? (select one)

[Brand A]	*[Brand B]*	Microsoft	NONE:
Desktop camera	Notebook camera	Desktop camera	If these were my only choices, I wouldn’t buy anything at this time.
$39	$79	$49	
Metal case	Plastic case	Plastic case	
No microphone	Detachable microphone	Built-in microphone	

Figure 2 shows a sample choice proportion result from a CBC study, in this case, a small study of a new product concept, N=56 ([16]; for brevity, we omit discussion of statistical significance in all of our examples). The given Feature "X" could be made available in 3 different ways (e.g., similar to the three microphone options in Table 1). We plot the likelihood of selection across all choice tasks, broken out by price point, for each of the three potential feature options.

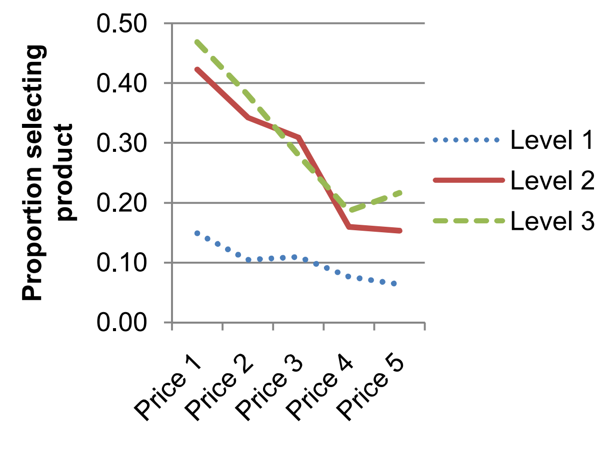

Figure 2: Conjoint results for feature "X"

The resulting choice proportion chart may be interpreted in two ways. Reading the vertical separation on the Y axis, one may interpret the relative difference in *likelihood to select* a product with different feature levels. On the X axis, one may interpret the relative differences in *perceived customer value*. In either case, it is important to interpret the results as being about the relative *difference* between feature levels; they do not represent market share, absolute purchase likelihood, or absolute willingness to pay as measured in real currency. In Figure 2, we can interpret the likelihood to purchase at Price 2 as being approximately 3x greater for feature Level 2 than for Level 1 (reading vertically at Price 2 on the X axis). Similarly, we see that at the same likelihood to select (approximately 15%), the sampled customers perceive Level 2 to be worth more than Level 1, with an incremental perceived value that is equivalent to Price 4 minus Price 1 (reading horizontally at 15% likelihood on the Y axis).

Overall, the results in Figure 2 demonstrate clearly that Level 1 of Feature "X" is inadequate; across all the CBC trials, almost no one chose any product with Level 1 of the feature, at any price. Levels 2 and 3, on the other hand, showed only a small difference in customer preference across price points. In the present case, this is informative for product design. We were able to assess that Level 1 is essentially unacceptable to customers, while we could choose between Levels 2 and 3 on the basis of other product goals, including engineering cost and alignment with other dimensions of product strategy.

2.3 Customer definition

At the simplest level, a product requires only one thing: a customer willing to pay. How does one determine which products will appeal to which users? A simple but suboptimal answer may be obtained by releasing a product and observing who buys it. Better market results may be obtained if there is some degree of initial market determination, followed by fitting between a product and a target market. This fitting may take the form of adjusting the product specifications to the market, or adjusting a marketing message.

The problems of determining a market are extremely complex, and involve a large degree of recursion (products help establish markets and vice versa). We cannot cover the topic exhaustively, but would like to call attention to three techniques that we have found helpful.

First, one may simply hypothesize a difference among groups on the basis of some theory or observation. For instance, for a new product at Microsoft, it was initially predicted that the product would appeal more to young adults (defined as 18-30 years old) rather than adults (35-55 years old). We examined this hypothesized difference with a variety of approaches, including interviews, focus groups, and field trials, along with choice-based conjoint (CBC) and MaxDiff exercises. A typical result is shown in Figure 3, plotting CBC results for the two groups for overall product interest vs. price (US results, total N=56). (For clarity of interpretation, we often find it useful to plot fitted, exponential regression lines.) In Figure 3, Young Adults show *less* interest in the product than Adults across a wide range of prices.

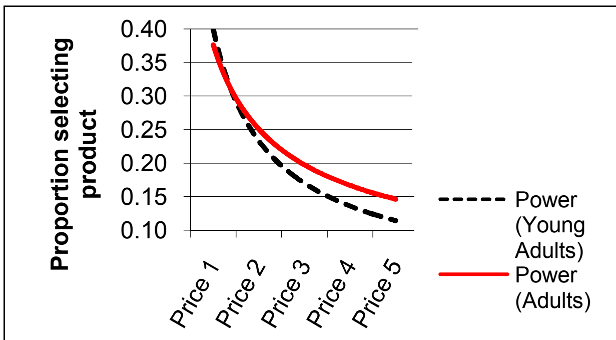

Figure 3: Conjoint likelihood by target group

After obtaining similar results across multiple countries and multiple research formats, we concluded that the initial hypothesis was incorrect; young adults would not necessarily make a better market than adults. We combined a deep understanding of users' needs and tasks (obtained through traditional qualitative user research) with the quantitative results of multiple CBC and MaxDiff exercises. Because the subsequent analysis was both deeply descriptive as well as quantitatively rigorous, product management was convinced to adopt a different strategy.

Simply determining that a given target customer group shows less interest is, in itself, not a particularly important finding; it may be that the target group is also concurrently underserved. In that case, a less interested customer group might, in fact, have higher market potential than a more interested but better-served group. It is important to include behavioral analysis and other user-centered design approaches to assess the opportunity in depth [13]. In the present case, market interest findings aligned with analysis of use cases and the degree to which needs were being served among each market group. This yielded a powerful conclusion that united qualitative and quantitative data with the results of user task analysis.

A second technique to understand the customer definition is to find dimensions that underlie observed interest. This is similar to the above hypothesis-testing approach, but adds an exploratory phase, in which dimensions of interest are identified in a dataset (e.g., through factor analytic approaches). The goal is to uncover the relationships that are of most direct product importance; and also, because this is applied research, to find relationships that are easily understandable and actionable.

For the Microsoft webcam product line, we conducted a series of large, online surveys in multiple countries, collecting a variety of data on demographics, technology usage, personality characteristics, and webcam product interest. There was team debate about whether webcam interest would be more closely related to personality style, such as extraversion, or to general interest in technology products; there were plausible arguments for each. Our survey included a psychometric scale measuring Extraversion [10] (among other personality factors) and a scale of Technology Enthusiasm derived from a confirmatory factor analysis of items identified in previous studies [2]. Measures of association showed no significant relationship between webcam interest and Extraversion (US survey, $r=0.04$, $p=0.26$, $N=737$), but a moderately strong relationship between webcam interest and Tech Enthusiasm (US survey, $r=0.38$, $p \ll 0.001$, $N=753$).

For ease of presentation and team discussion, we then divided respondents into groups (three evenly distributed groups for Extraversion; three non-uniformly distributed groups for Tech Enthusiasm, based on clustering results described below). Figures 4 and 5 show the results, plotting the proportion of people in each group who own, want, or don't want webcams. The differences clearly establish that Technology Enthusiasm is a more explanatory underlying dimension than Extraversion for these data.

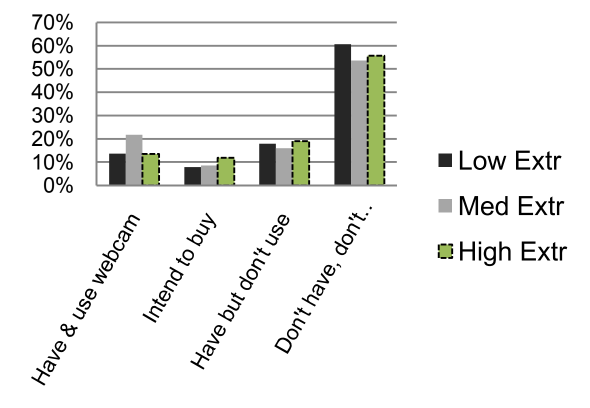

Figure 4: Ownership by extraversion group

We have found it helpful not only to examine the strength of such associations, but also to find easily interpretable forms to present the results to product teams, as in Figure 5. The importance of such results can be great. In the current example, for instance, designing a product for technology enthusiasts would emphasize attributes such as having the latest technology, being innovative, and having the most technically advanced feature levels. Designing for extroverts, on the other hand, would emphasize different attributes, such as specific use cases or marketing messages.

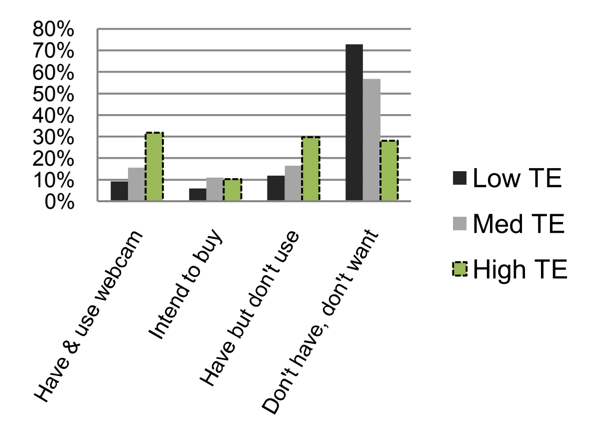

Figure 5: Ownership by technology enthusiasm (TE) group

A third technique to examine customer definition is to use clustering methods to find previously unknown groupings of customers. One significant problem with traditional clustering techniques is that there is no clear way to determine the number of clusters [7]. A clustering algorithm typically yields N solutions, where N is the number of respondents, with individual proposed clustering models that comprise every size from 1 to N clusters. In practice, when such exploratory clustering is conducted on a dataset with many measures of potential interest, it can be very difficult to select among the multiple results.

One alternative is to use a clustering method that relies upon distributional assumptions and maximum-likelihood techniques to find a model with an optimal number of clusters [6][7]. We have used the model-based MCLUST algorithm to detect clusters of interest, i.e., customer groups who may share similar characteristics [5]. In one example, with data from an online survey of 1008 respondents, we used the MCLUST technique to find an optimal cluster solution on the basis of the self-reported importance of using various technology items (cell phone, laptop computer, digital video recorder, etc.). The resulting optimal cluster model found 4 clusters, which were primarily differentiated in terms of the total number of technology devices owned. We used the sizing of those 4 clusters to establish the sizes of our Technology Enthusiasm groups previously noted (combining the two lowest-scoring clusters into a single "Low" group). The cluster sizing determined by MCLUST is shown in Figure 6 (software [5][15]).

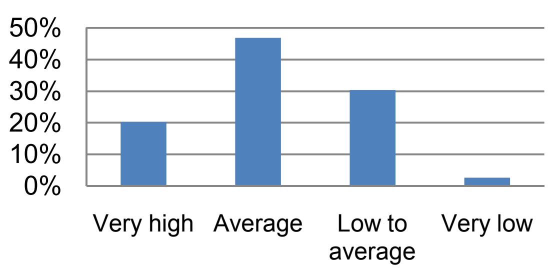

Figure 6: Cluster sizing for technology enthusiasm

For user researchers, there are many advantages to conducting this kind of exploration of customer definitions. It allows one to test hypotheses about user groups, to ensure that predictions derived from theory, speculation, or qualitative research are accurate. It affords the opportunity to discover new relationships that may significantly inform product design. It serves to identify the attributes to use when screening participants for studies such as focus groups and usability tests. Finally, it yields hard data that may be useful in convincing engineering teams or management of the appropriateness of a design strategy.

3. Informing business strategy

Given a defined product space and user target, one can explore the business implications of various product decisions. One way to do this is to conduct market simulation exercises, in which respondents' choice tasks (from conjoint exercises) are used to estimate the relative preferences that customers would have for various product configurations. This works by constructing product configurations and then estimating for each individual in the dataset which product would be preferred on the basis of that individual's attribute preferences as observed in the conjoint trials.

The results of such a simulation may be expressed in multiple ways, including product preference share and estimated revenues. These are relative values, since the market size cannot be known through the simulation. However, they can be used to estimate customer preferences, assuming that all other untested factors are equivalent. The results can also be used with other sizing information such as census or sales data.

We have found it helpful to apply a game theoretic framework to the results of such simulations [14]. Table 2 shows the results for an actual revenue simulation where Microsoft and a primary competitor would make a new webcam "Y" with one of two possible feature sets (yielding product Y1 or Y2). We modeled this as a 2-player, non-zero sum game, with players making simultaneous choices.

Results are shown in Table 2, where entries are estimated revenue results by product line in standardized units for Microsoft and Competitor A, in that order, according to the product made by each. For example, if Microsoft makes product Y2 while the competitor makes product Y1, then Microsoft will expect a standardized revenue of 4.1 units, and the competitor will expect 3.8 units, all other factors being equal.

Table 2: Results of Webcam Product "Y" Simulation in Game Strategic Form

		Competitor A	
		Product Y1	Product Y2
Microsoft	Product Y1	(2.3, 5.4)	(1.9, 6.7)
	Product Y2	(4.1, 3.8)	(2.9, 5.2)

As shown in Table 2, both Microsoft and Competitor A obtain higher results by making product Y2, regardless of what the other player does. As an example, if Competitor A makes Y1 (first column), then Microsoft will expect 2.3 revenue units for making Y1, but 4.1 units for making Y2.

Better results from making Y2 occur for every pair of potential strategies – reading across each row and down each column – for both players. Because of this, it is rational for each player to assume that the other one will in fact make product Y2. Also, if either player makes Y2, it forces the other player to respond, since otherwise that player's revenue would drop substantially. If it turns out that either player in fact makes Y1, then the other player will benefit hugely by making Y2 instead. In game theoretic terms, the present simulation shows strict dominance for the Y2 strategy over Y1, for both players. Although most product scenarios require a more complex model than a simple 2-player, 1-choice game, this general approach is extensible to a variety of strategic situations [14].

In the present model, the expected total market (i.e., the sum of revenues for both players) grows if at least one of the players makes product Y2. This demonstrates the value of using a non-zero sum approach, since it allows an estimate of how the overall market might grow due to differential user interest in a feature. Such growth may benefit both competitors.

We have found that this kind of analysis – when combined with extensive user research that backs up its findings in depth – is highly interesting to product management and is able to inform strategic choices. These results show the value of having comprehensive user research, so that such strategic choices can be identified.

The performance of such methods can be assessed empirically through reliability analysis or experimental research. We typically assess reliability both within groups and across locations. One might additionally examine validity and accuracy through quasi-experimental studies. For instance, one might offer products at multiple locations with different prices or feature messaging and examine whether the observed outcomes match prediction.

Another important measure of success is impact on the development process. In our case, engineering teams and upper management have welcomed the new methods, and the process described here is now part of our standard early research process. We have also seen interest from other product line research groups within the company; our group has mentored several other teams to adopt these methods.

These kinds of methods are easiest to apply when a product category is well-defined, as with traditional consumer goods. Users must understand the product in order to assess its features or use cases. With careful attention to how research is conducted and interpreted, we believe these methods are applicable to many kinds of innovative products [13]. For instance, with a novel product, user familiarity might be achieved through longitudinal interaction with prototypes.

4. Why user researchers are right for these methods

Why are these methods right for user researchers? Some of them, such as conjoint analysis, are fundamental to an existing and quite different discipline: market research. One might wonder whether we are simply defining user research to take over the space of market research. However, we believe there are good reasons why user researchers should be considered as users of these methods.

First, technology products can be substantially more complex than many traditional products such as consumer packaged goods. Understanding customers' needs requires detailed attention to user tasks, goals, use cases, and interaction paradigms. Without such detailed attention to the specifics of customer needs, it would be easy to design a product that meets a high-level specification on paper, but fails to deliver a usable product. Because they study and understand the specifics of product interaction, HCI specialists such as user researchers bring detailed user knowledge to the design and interpretation of higher-level quantitative studies. It is not the result of a single market study that should determine product direction, but an analysis of both market-level and individual needs. User researchers are well positioned to address both.

Second, many HCI practitioners have background in experimental methodologies that are helpful to design and interpret quantitative research studies. To conduct and interpret quantitative studies well, one needs substantial familiarity and experience with experimental design, hypothesis testing, interpretation of statistical tests, and familiarity with the problems of analysis of statistical power, multiple comparisons, and aggregation of results (such as meta-analysis). Anyone can operate a survey tool or statistics package, but interpreting the results requires specialized knowledge. A strong background in experimental behavioral research is ideal preparation for this kind of analysis. (On the converse side, user researchers without such expertise may be well-advised to avoid these methods.)

Third, user researchers often have close ties to product engineering teams. This is helpful both to determine which aspects of a product should be studied, and to ensure that the results are interpreted and made actionable to the engineering team as the product lifecycle continues. Close communication with the engineering team may lead to a research study that includes product specifics that are important but would have been unnoticed without such close cooperation. As the product lifecycle continues, user research shifts from early needs into direct product engineering (e.g., usability testing). A user researcher who was involved deeply in early-stage product research will be able to ensure continuity in understanding and interpreting user needs.

To be sure, such expertise and lifecycle engagement does not necessarily require embodiment in a single individual or a single discipline. It would be possible for multiple people to share roles, or for people in multiple disciplines (e.g., usability engineering and market research) to collaborate sufficiently to ensure adequate coverage. However, in practice, it may be easier for one person or discipline to have responsibility for both user needs and customer understanding: it reduces the overhead of collaboration, ensures continuity of knowledge across the lifecycle, and yields the most consistent approach to conducting research and synthesizing its results.

5. Conclusion

The cases and applications discussed here highlight methodological tools that are available but may be unknown or unused within the applied HCI community. These methods do not replace traditional HCI methods in product design but complement them. They may contribute increased validity for some kinds of user research studies and lead to a more scientific approach to higher-level questions about user needs. These tools are able to assist with customer definition, feature definition, product configuration, market identification and sizing, and product line strategy. HCI practitioners with strong experimental methodology backgrounds may be the right people to use and apply these methods, especially in collaboration with an organization’s marketing team. If researchers adopt these methods when appropriate, we believe they may be better able to understand and deliver what customers need. That should lead, in turn, to improved business results as well as more influence within organizations for HCI practitioners.

6. References

[1] Beyer, H., and K. Holtzblatt. *Contextual Design: Defining Customer-Centered Systems*. San Francisco: Morgan Kaufman, 1998.

[2] Chapman, C.N., Love, E., Alford, J.L, and Farrow, R Video Communications Market Segmentation, Microsoft Corporation, March 2007 (unpublished internal report).

[3] Cohen, Steve and Paul Markowitz. “Renewing Market Segmentation: Some New Tools to Correct Old Problems,” ESOMAR 2002 Congress Proceedings, 595-612, ESOMAR: Amsterdam, The Netherlands, 2002.

[4] Finn, A. and Louviere, J. J.. “Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety,” *Journal of Public Policy and Marketing*, 11:1, 12-25, 1992.

[5] Fraley, C., and Rafferty, A. E. “MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering.” Technical Report 504, Department of Statistics, University of Washington, September 2006 (revised January 2007).

[6] Fraley, C., and Rafferty, A. E. “Model-based clustering, discriminant analysis, and density estimation.” *Journal of the American Statistical Association* 97:611-631 (2002). (Also available as Technical Report 380, Department of Statistics, University of Washington, October 2000.)

[7] Fraley, C., and Rafferty, A. E. “How many clusters? Which clustering method? - Answers via Model-Based Cluster Analysis.” *Computer Journal* 41:578-588, 1998. (Also available as Technical Report 329, Department of Statistics, University of Washington, February 1998.)

[8] Green, P. E., J. D. Carroll, and S. M. Goldberg (1981), "A General-Approach to Product Design Optimization Via Conjoint-Analysis," *Journal of Marketing*, 45:3, 17-37, 1981.

[9] Green, P. E. and Srinivasan, V. "Conjoint Analysis in Consumer Research - Issues and Outlook," *Journal of Consumer Research*, 5:2, 103-23, 1978.

[10] Harms, P. D., Wood, D., and Roberts, B. W. “Twenty Item Personality Inventory.” Champaign, IL: University of Illinois, Department of Psychology, 2007 (manuscript in preparation).

[11] Kuniavsky, M. *Observing the User Experience: A Practitioner's Guide to User Research*. San Francisco: Morgan Kaufmann, 2003.

[12] Louviere, J. J. “Best-Worst Scaling: A Model for the Largest Difference Judgments,” Working Paper, University of Alberta, 1991 (unpublished).

[13] Love, E., and Chapman, C.N. “Issues and cases in user research for technology firms”. Proceedings of the Sawtooth Software Conference 2007, Santa Rosa, CA, October 2007.

[14] Myerson, R. B. *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard Univ. Press, 1997.

[15] R Development Core Team. *R: A language and environment for statistical computing (version 2.4.1)*. R Foundation for Statistical Computing, Vienna, Austria, 2006.

[16] Sawtooth Software. *SSI/Web, version 6.0.6*. Sequim, WA, 2006.

[17] Sawtooth Software. “The MaxDiff/Web System Technical Paper.” Sequim, WA: Sawtooth Software, 2005.

[18] Shneiderman, B. and Plaisant, C. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 4th ed. Reading, MA: Addison-Wesley, 2004.