More than a Feeling: Understanding the Desirability Factor in User Experience

Carol M. Barnum

The Usability Center

Southern Polytechnic

State University

1100 S. Marietta Pkwy

Marietta, GA 30060 USA

cbarnum@spsu.edu

Laura A. Palmer

The Usability Center

Southern Polytechnic

State University

1100 S. Marietta Pkwy

Marietta, GA 30060 USA

lpalmer2@spsu.edu

Abstract

Interest in understanding the "desirability" factor in user experience continues to grow while the use of post-test questionnaires to measure desirability continues to be problematic. Microsoft created a toolkit to address desirability in studies, and their use of the product reaction cards from that kit was presented at conferences in 2002 and 2004. Since then, however, little has been published about how others have used the cards to measure desirability. We began using the product reaction cards in 2006, and we report on the results in case studies from the past several years. We find that the cards prompt users to tell a rich and revealing story of their experience. Triangulating these findings with post-test questionnaire data and direct observation strengthens the understanding of the desirability factor.

Keywords

Desirability, satisfaction, emotion, user experience, usability, usability testing, user research, usability research, methods

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

H5.2. User Interfaces – Evaluation/methodology

General Terms
Human Factors, Experimentation, Measurement

Introduction

The ISO definition of usability (9241-11) has three major elements for gauging the usability of products for specified users with specified goals in specified contexts of use: effectiveness, efficiency, and satisfaction. Usability evaluations generally do a good job of uncovering effectiveness and efficiency issues, but may not do so well in addressing user satisfaction. While effectiveness and efficiency are certainly important to overall system usability, satisfaction may be the element that makes or breaks successful adoption.

Post-task and post-test questionnaires often ask users to rate their level of satisfaction with the product. However, it is widely recognized that these ratings tend to give higher values (the acquiescence bias) to satisfaction and usability than observations of user experience indicate. In a review of 180 published usability studies, Hornbaek [3] sums up the problem of using questionnaires: "The challenges are to distinguish and empirically compare subjective and objective measures of usability: to focus on developing and employing measures of learnability and retention . . . [and] to extend measures of satisfaction beyond post-use questionnaires" (p. 79).

Recognizing the limitations of post-test questionnaires and the time-consuming nature of semi-structured interviews following testing, Microsoft usability engineers created an instrument to get at the essential, elusive quality of desirability [2]. Their work was adapted for a larger scale study by the MSN 9 team [8].

Since these two reports were published, little more has been published on how other companies may be using the product reaction cards, which comprise 118 words or phrases, presented to participants, who then choose the cards that reflect their experience.

Tullis and Stetson [7] included a version of the card set in their comparative evaluation of five different feedback mechanisms, four of which were questionnaires and one of which was the cards. The SUS (System Usability Scale) proved most reliable in indicating the preference of users when comparing two websites. However, the authors' comments about the cards indicate the need for further study:

> When evaluating only one design, possibly the most important information is related to the diagnostic value of the data you get from the questionnaire. In other words, how well does it help guide improvements in design? That has not been analyzed in this study. Interestingly, on the surface at least, it appears that the Microsoft Words might provide the diagnostic information, due to the potentially large number of descriptors involved.

We wanted to pursue this potential for diagnostic information, so we began using the product reaction cards in our studies in 2006 and have used them often since then. We have found that these cards unlock information regarding the user's sense of satisfaction in a more user-centered way than any other tool or technique we have tried. In our experience, the reason

for the success of product reaction cards is simple: this tool provides a way for users to tell the story of their experience, choosing the words that have meaning to them as triggers to express their feelings—negative or positive—about their experience.

We share the results from several of our studies, following a review of the ways in which Microsoft and others have used the cards.

Microsoft's creation and use of the cards

The complete set of 118 product reaction cards represents a broad spectrum of options and dimensions of desirability. The words and phrases created by Microsoft were obtained from market research, prior user research, and team brainstorming. Based on their observations of the higher-than-average positive response from participants in completing post-test questionnaires, the team established the ratio of positive to negative words at 60% positive and 40% negative or neutral words.
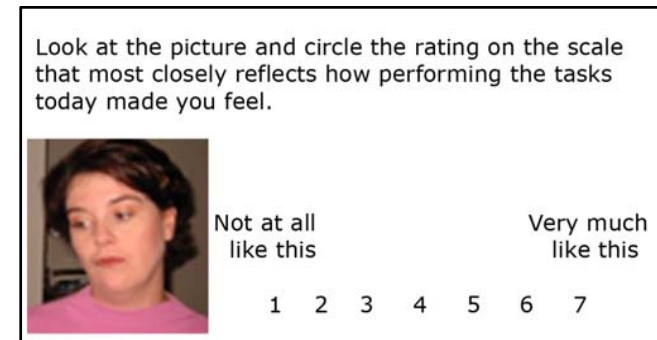
The first studies [2] using the cards did so as part of the desirability toolkit, which had two components:

- A faces questionnaire, in which participants were asked to look at a picture of a face and circle a rating on a scale that most closely identified with how performing the task made them feel.
- The product reaction cards, in which participants were asked to choose as many cards as they liked to express their feelings about the experience; then narrow their choices to their top five and explain these. All cards were recorded, with details about the top five cards included in the analysis.

The faces questionnaire (Figure 1) worked well for

some, but confused others; so it was abandoned.

The product reaction cards were refined, keeping the balance at 60% positive and 40% negative or neutral and keeping the number at 118.

Look at the picture and circle the rating on the scale that most closely reflects how performing the tasks today made you feel.

Not at all like this　　　　　Very much like this

1　2　3　4　5　6　7

figure 1: Example item (1 of 6 faces) from the faces questionnaire

The second reported study [8] using the cards did so in a very different format. This study required large numbers of participants to evaluate four different design options for a selection of the best option on the basis of user preference for visual design. Focus groups were primarily used with the cards being a part of and an impetus for the discussion.

The resulting design direction was based on the data collected from this study. Project success was measured by customer feedback from the beta launch and subsequent launch to the general public. A large number of spontaneous, positive comments from the 25,000 beta users were supported by survey response data, which showed that users consistently rated the

appearance of MSN 9 above even those features rated most *useful* in the product. Among all the methods used, the team concluded that future testing would focus more on the product reaction cards and less on traditional Likert-type statements and the other methods used in this study.

How others have used the cards

Very few reports of the use of the cards have been shared with the user experience community. One of these comes from studies of desirability at Yahoo! [5]. Using its own version of the cards, which aligned with brand values, the team instructed participants to select cards to express their responses to various designs. Two studies were conducted: a qualitative study and a quantitative study.

In the qualitative study, individual participants in a lab setting were shown different visual designs for Yahoo! Personals, and asked to select the card that best matched their response to each design. Participants were then interviewed to understand their basis for selecting the cards they did.

In the quantitative study, various images were embedded into a survey and participants were asked to respond with their choice of a word/phrase from the list that matched their feelings about each design. This process generated a lot of data for analysis and compilation into quantitative results.

Others [1, 6] have reported anecdotally that they have listed the words on a sheet of paper, handing the sheet to the participant to review and select the words, then scrambled the order of the words on the sheet to give to the next participant.

The format of our studies

Our interest in using the product reaction cards derived from the fact that we wanted to know more about the users' experience than we felt we were getting from post-task and post-test questionnaires. As most of our studies are qualitative, often as part of iterative development, we wanted to know more about an aspect of our studies that defied the regular measurements of mouse clicks, optimal paths, or help desk calls. We wanted to know what users *felt* and, specifically, we wanted to add an element of methodological plurality to our studies. Including a qualitative method, such as the product reaction cards, would, we believed, provide a way to triangulate our findings from other feedback mechanisms, with the potential to produce more meaningful and substantive results for our clients.

When we first started using the cards in our studies, we used the sample set of 55 cards presented in the original proceedings article. We later enlarged the size to the full 118 cards available from Microsoft's website [4].

Our method is as follows:

- After completing the study (and in some cases, after certain scenarios within the study), we ask each participant to go to a table with the cards spread out in a random pattern.

- We ask them to look over the cards, and pick up 3 or 4 or 5 cards (making the suggestion in a way that maintains a flexible requirement on the number of cards) that match their experience of working with the product.

- We then ask the participant to bring the cards back to the desk, place them under our document camera so that we can record them, and tell us what each card means to the participant.
- We record the comments made by the participant, both on video and in our log (for analysis in the findings meeting).
- We return the cards to different places on the table, so that they are arranged differently for the next selection round, whether that selection is made by the same or a different participant.

As early as our first study, we saw that overall feelings, thematic word clusters, and repeated word choices gave us a new feedback mechanism to understand the user experience. The repetition of selected cards in each study has been uncanny, at first, and then consistent through repeated use. In addition to noting the number of times the same cards are selected by more than one participant, we have noted that the cards, when grouped, convey positive themes or concepts such as "efficient" and "time-saving," as well as negative findings, such as "annoying" and "stressful." The thematic groupings provide yet another assessment of how participants responded to the site.

In the case studies presented here, we also demonstrate how results can be represented visually. Qualitative results—such as language-based outcomes—can also be presented quantitatively to indicate how the results cluster as multiple users select the same or similar cards. Our illustrations of ways to present these findings visually are designed to indicate flexible and innovative options for visual representation in usability practices. We do not mean to suggest that

more than one method should be used in any particular study; rather, we want to show a range of options that can be used to present the findings from users.

Our rationale for presenting a number of case studies, rather than just one or two, is to demonstrate the consistency in results obtained from using the cards. Because so few studies have reported on the findings from using the cards, we believe that numerous examples will help others see that the cards can add to our understanding of users' experience.

Case study 1: Computer Network Monitoring Application
Our first study with the cards was conducted in 2006 for a software company that produces enterprise computer network monitoring and management applications. In this study, six participants—all IT professionals—were asked to perform a series of typical tasks with the new remote-based application. The study had several goals: the client wanted to gain feedback about features such as the navigation and user interface of the new system and wanted to know if the participants could easily learn the new system. Our scenarios were constructed accordingly and we used the cards to gain a sense of the users' feelings about this new application.

Among the cards chosen, the breakdown of positive versus negative language was 30 cards and 13 cards respectively, which was very close to Benedek and Miner's finding of 60% positive/40% negative choices as the basis for their card creation. The negative card selections were our immediate focus, as we wanted to know the aspects of the product that evoked a negative experience.

Four of the six participants chose negative words, with *time-consuming* chosen by three participants and *frustrating and inconsistent* chosen by two participants.

Although we expected more positive language choices than negative choices, we didn't expect that the four participants who chose negative cards would also have positive impressions about the application. While the product's strongest negative dimension was *time-consuming*, the product was also *usable* according to four participants. Cards including *flexible* and *accessible* were selected by three of the six participants. We clearly had a trend developing with participants and their impressions. However, it would be several more years and multiple studies later before a more compelling pattern of consistency in user choices emerged.

Case study 2: Destination Teaching website
Destination Teaching is a website organized as a resource to prospective teachers in the state of Georgia (USA). It compiles information about education programs around the state, scholarships and loans for prospective teachers, types of teachers needed by the state according to subject and grade level, and information about teacher certification.

From the six participants in this small study, a total of 18 words were selected: 17 positive and one negative. Among the positive words, all six participants chose *useful* and 3 of the 6 chose *organized*. However, when the positive words are clustered, the strength of the positive findings becomes even more obvious, as shown in Figure 2.

accessible comprehensive connected easy-to-use fast **organized** relevant simplistic time-saving usable **useful**

figure 2. A tag cloud illustrates that the two most frequently chosen words were *useful* and *organized*, with all other words represented with equal weight.

With these tallied results, we saw that the product reaction card findings provided an entirely different lens of assessment from the perspective of the participants. Although the participants had rated the site well in their questionnaire responses, we learned from their choice of cards how positively they felt about the teacher education site. Our study demonstrated that not only was the site's current content rated as valuable, but that the design facilitated participants' information-seeking practices.

Case Study 3: Comparative evaluation of hotel reservation process
In this usability study with a major hotel group's website, 12 participants compared a new site prototype with the current site. Participants were asked to make a reservation and include add-on options such as tickets to a local sports event, using both versions of the site. The feedback from the product reaction cards revealed that the negative word choices for the prototype totaled 57% and the positive words totaled 43%.

Participants' experience with the current site was, however, in stark contrast. For the current site, the positive word choices were 86% of the total while negative words were only 14%. Clearly, there was a meaningful difference in how the participants were responding to the two versions of the site, as revealed by their card choices.

Figure 3 presents the positive words selected for both the prototype and current site, as well as word clusters. The prototype site was rated much less positively (43%) than the current site. Fewer words were selected by participants to describe their experiences with the current site; in total only six unique positive words were selected. The repetition of words among the participants was also low; the expression *easy-to-use* was selected twice. No other words were selected by more than one participant. The strength of word clusters was low with only two clusters of three words each.

For the current site, we saw a distinct and measurable difference along several unexpected dimensions. The number of positive words chosen was also greater than for the prototype site. Participants indicated that their range of positive feelings about the current site was also greater than for the prototype site. In total, nine unique positive words were selected.

As represented in figure 3 several words were selected repeatedly by participants to describe their experience with the current site. *Efficient* and *time-saving* were selected three times each. They clustered with the individual selections of *straight-forward, accessible*, and *inviting.* When visually represented in this cluster, we gain a big-picture view of factors contributing most

strongly to participants' positive assessment. The words *useful, predictable, usable*, and *valuable*—chosen only once each—also demonstrate, via their small clusters, important secondary influences to the positive ratings.

Prototype Site
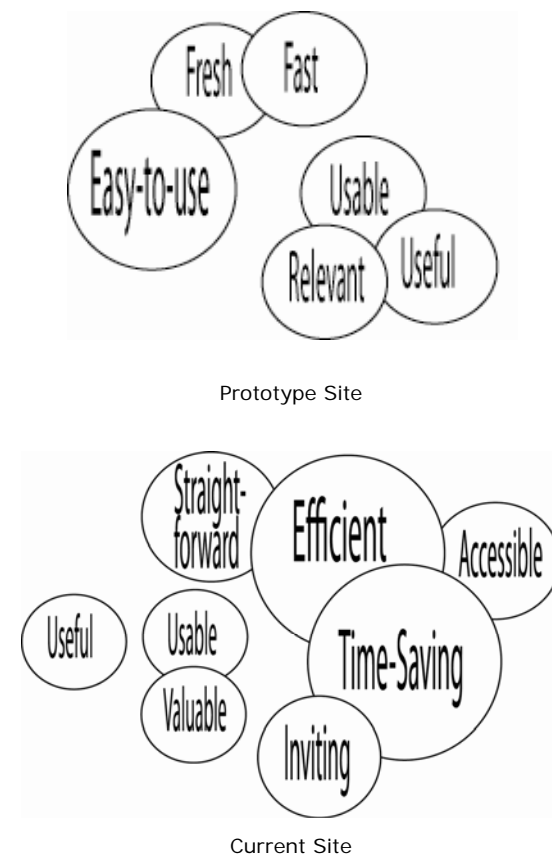
Current Site

figure 3. Positive language for prototype site vs. current site

The overall negative word rating for the prototype site was 57%. However, the significance of this number was

clearer when we looked at the selected words, their frequencies, and their cluster effects. The word *annoying* was selected by three participants and formed the core of the cluster. *Stressful* and *not valuable* were both chosen twice. Figure 4 represents this cluster and provides a strong message about what participants thought. The words chosen for the current site indicated that only 14% of the total reflected an undesirable experience. The word *time-consuming* occurred twice; *annoying, cluttered*, and *inconsistent* were selected once each. The word cluster for the current site provides a much different picture as compared to the prototype site.

In the next phase of this series of comparative usability analyses, 12 participants compared the experience of making a hotel reservation with the newest prototype site and the current site and then with two other competitor sites. Unlike our other work with the cards, there was no clear favorite site. The product reaction card selection told us that participants were, at best, mixed-to-negative about their experience with all four sites. The prototype site and competitor #2 were overwhelmingly disliked with 66% and 75% negative words chosen as descriptors.

As shown in Figure 5, individual card selections for all of the sites indicate that each site was rated as *frustrating* by at least one participant. The hotel group's current site was the most frustrating with participants making three selections of that word. The prototype site was also *frustrating* (2), *confusing* (3), *busy* (2), and *not valuable* (2). Clearly, participants made their preferences known through their selections from the product reaction cards.

Prototype Site

Current Site

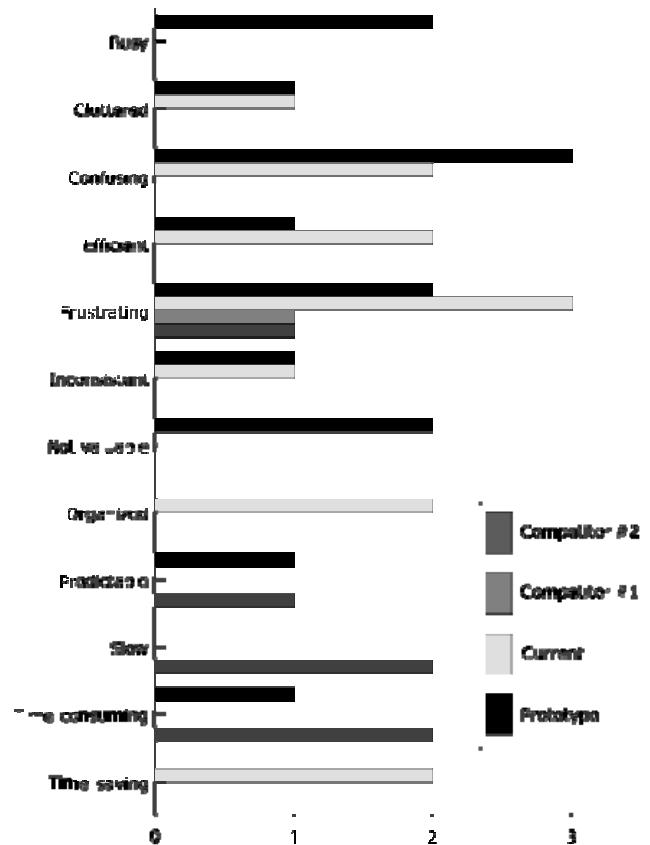**figure 4.** Negative language for prototype site vs. current site

figure 5: Card frequency selections for prototype, current, and two competitor sites for hotel study.

Case Study 4: Hotel group study on fee-based loyalty program enrollment
In our continuing work with the major hotel group, we conducted a study to elicit participant feedback regarding the enrollment process for a fee-based loyalty rewards program. The hotel group wanted to

know how the current site and a prototype site compared to the enrollment process for a fee-based loyalty rewards program offered by a major airline (Delta's Crown Room program). There were two phases to this study and six participants were recruited for each phase.

In this first phase, all the participants were current customers and members of the hotel's free loyalty rewards program.

Most compelling for us in this study was the language associated with speed and ease-of-use for the current fee-based loyalty rewards program site. Sixteen total cards were selected by participants; *complex* reflected the only negative choice. The word *fast* was selected three times; *efficient* was selected once. It was easy to see that participants liked the speed at which the system functioned. They were also positive about how easy the current site was to use as shown in Figure 6. The terms *straight-forward* (4) and *familiar* (3) repeated strongly among participants.


figure 6. Cards chosen for ease of use for the current site. Positive cards far outweigh negative cards in this category.

Participants rated the prototype site experience as primarily positive, with ten of the 16 words selected being positive descriptors. The expression *easy-to-use*

was selected twice; no other words were repeated. Single card choices were *hard-to-use, confusing, cluttered, slow, boring, frustrating*. However, as with our other studies using the cards, we saw the emergence of strong, categorical language groupings that framed the user experience around categories, such as *speed* and *ease-of-use.*

Ease-of-use, for example, was a highly conflicted category (Figure 7) with the number of negative and positive words being equal: *easy-to-use* (2), *simplistic, hard-to-use, frustrating, confusing*. However, participants expressed their confidence in the system via *trustworthy* and *secure*. As well, the value of the site's content was seen as *comprehensive* and *useful*.
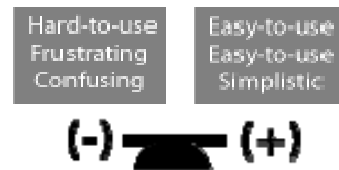
figure 7. Mixed language expressing an equal divide among participants for ease of use in prototype site.

When it came to describing their experience with Delta's Crown Room registration process, the cards showed that it did not fare well (Figure 8). Nine of the 15 words chosen were negative. Three participants chose the word *confusing* to describe their experience with the registration process. On the positive end of the spectrum, the word *usable* was chosen twice to describe the site. However, like the hotel's fee-based loyalty program prototype site, the remainder of the selected cards did not consist of repeated words,

although categorical themes emerged again for ease-of-use.
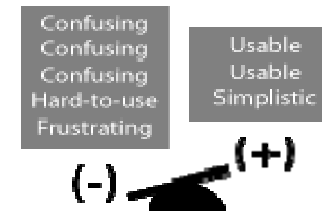
figure 8. Cards chosen for ease of use for airline loyalty program site. Negative cards outweighed positive cards in this category.

Participants' card choices across this comparative usability study made a clear statement: the prototype was not, as the developers had hoped, a new and improved version of the current site. Instead, the results indicated that more work on the prototype would be required prior to the next round of testing.

A week later, we conducted the second phase of this study, which compared the current version of the loyalty program's enrollment process with two new prototypes of the site. Of the 17 words selected to describe the current site, 10 positive words were chosen (58%). Prototype 1 also elicited 17 word choices from participants, 15 of which were positive (88%). Prototype 2 was described by participants with only 13 word choices, nine of which were positive (69%).

This time, it was clear from the card choices that participants felt only marginally positive about their experience with the current site. With the exception of

one individual, every participant had at least one negative word in his or her card selections. The words *easy-to-use* and *usable* were selected twice to describe the current site; however, *confusing* and *time-consuming* were also selected twice. For prototype 1, only two participants included negatives in their choices: both chose the word *confusing*. Like prototype 1, prototype 2 had only two participants select negative words. The word *confusing* was selected twice with *inconsistent* and *cluttered* selected once each. The value in reviewing these responses is both in the choices themselves and the comparison of choices made by participants working with the current and the prototype sites.

Our next phase of analysis provided insight into the defining characteristic for user preference. Although all three sites were described by the same four words— *easy-to-use, organized, familiar, and usable*—it became clear when we further analyzed the findings that prototype 1 was described as *easy-to-use* by five out of six participants. Thus, while all three versions of the site were *easy-to-use*, prototype 1 outranked the others, as shown in Figure 9.
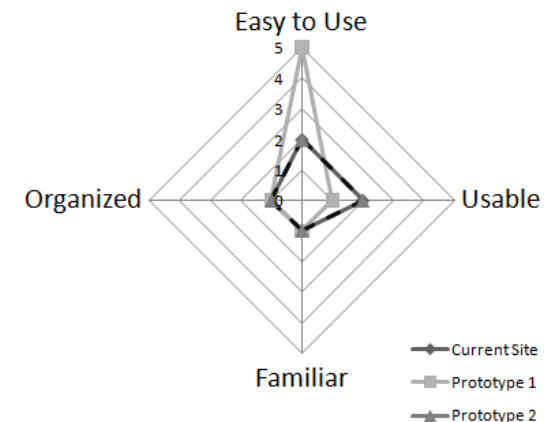
figure 9. Comparison of card choices related to ease-of-use category for current site, P1 and P2.

When we looked at prototype 2 in more detail to see which card choices differentiated it from prototype 1, the *speed* category was the one significant area of difference. Prototype 2 was described as *time-saving* (2) and *fast* (1); only one reference to speed—*time-saving*—was found in the choices for prototype 1.

The compilation of participant preferences indicated that the current site was the least preferred of the three versions. We expected that the remaining five participants would select prototype 1, as the cards indicated it was more positively rated and the ease-of-use language was strong. However, the results were split.

Two participants selected prototype 1 as their preferred site while three selected prototype 2. This slight preference towards the second version was interesting in light of the fact that prototype 2 was deemed less

easy-to-use than prototype 1 but was seen as time-saving. This difference in the category of speed might have been the variable that tipped the decision slightly in favor of prototype 2. Without the cards, however, we might not have understood this basis for participants' preferences.

Case Study 5: Call center application
Our most recent use of the cards examined the experience of call center employees using a new application for managing customer calls. As the company was eliminating an older, green-screen system and replacing it with a web-based, graphical interface, it was critical to gain insight about the employees' reactions to the radically different look and feel as well as their interest in learning to use the new system. Additionally, the call center employees would be working with one part of a complex system that represented the upcoming consolidation of multiple data sources into a centralized information source.

Eleven participants used the new interface in a series of typical telephone-based interactions with customers. Of the total cards chosen by participants, 51 of the 56 choices were positive (93%); only 5 words (7%) reflected a negative assessment of working with the new interface. As the current system involved switching back and forth between databases to answer customer queries, the participants chose cards that expressed their pleasure with the speed and ease-of-use this new system provided. That the system felt familiar and that it was easy to understand also resonated positively with the participants, especially as they were using the system without any training. Five of the 11 participants selected *easy-to-use* from the 118 cards (as shown in Figure 10), among other words picked more than once.
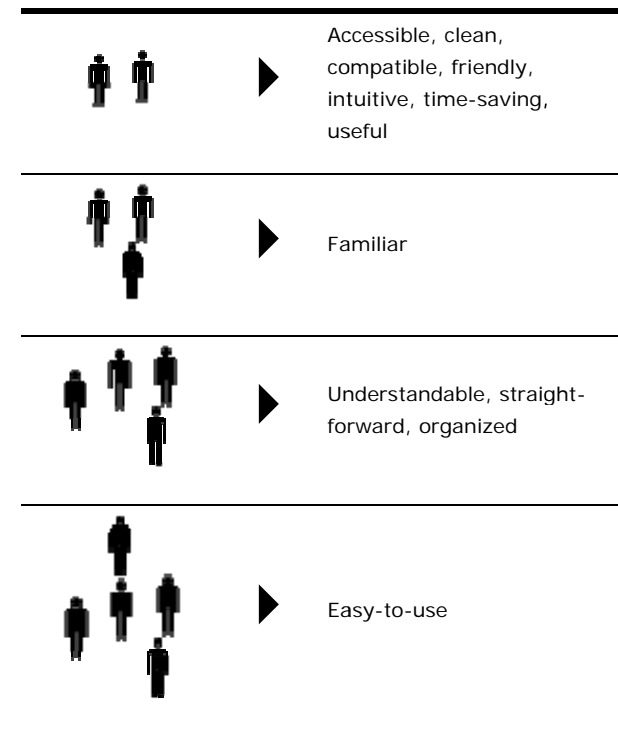
figure 10. Number of participants selecting the same cards to describe their experience with the new call center interface.

- Accessible, clean, compatible, friendly, intuitive, time-saving, useful
- Familiar
- Understandable, straightforward, organized
- Easy-to-use

Limitations
Our studies have been primarily small in scale, with as few as six participants in a single study. The results of larger-scale studies are needed to determine if the findings reported here are seen in larger studies.

The issue of higher-than-expected positive self-reporting (the acquiescence bias) does not necessarily go away with the use of product reaction cards versus standard post-test questionnaire instruments.

The time to administer the card choice process and the time needed for users to share the relevance of their card selection may restrict the use of the cards in studies that are tightly controlled by time constraints. However, we have found that the card selection process and report takes far less time than semi-structured interviews.

The cards should not be used as the sole means of getting participants' feedback regarding their experience. They work best when used along with other satisfaction survey instruments or when used as a baseline for comparison in iterative studies.

Results

From these case studies and our continuing use of the product reaction cards, we have found they help us understand the nebulous, critical feeling of desirability in user experience. Whether the study is of a single product or a comparative evaluation, card selection gives our participants the means to share their story of their experience with us.

The language choices and themes created for us by our participants help us triangulate our findings with other methods of usability data collection, such as logger's notes, standard post-test questionnaires, and video reviews. The points at which findings intersect can provide meaningful direction for the iterative development process.

References

[1] Barker, I. Presenting findings about emotional responses. *Simpler is better* (2007, June 12) http://simplerisbetter.wordpress.com/2007/06/12/presenting-findings-about-emotional-responses/#comments

[2] Benedek, J., & Miner, T. Measuring desirability: New methods for measuring desirability in the usability lab setting . (2002) http://www.microsoft.com/usability/UEPostings/DesirabilityToolkit.doc

[3] Hornbaek, K. Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies 64* (2006), 79-102.

[4] Microsoft Product Reaction Cards. http://www.microsoft.com/usability/UEPostings/ProductReactionCards.doc

[5] Rohrer, C. Desirability studies. (2009, January 14) http://www.xdstrategy/com/blog

[6] Travis, D. Measuring satisfaction: Beyond the usability questionnaire. (2008, March 3) http://www.userfocus.co.uk/articles/satisfaction.html

[7] Tullis, T.S. & Stetson, J. N. A comparison of questionnaires for assessing website usability. Usability Professionals Association Conference, 2004. http://home.comcast.net/~tomtullis/publications/UPA2004TullisStetson.pdf

[8] Williams, D, Kelly, G, Anderson, L,, Zavislak, N, Wixon, D, & de los Reyes, A. (2004). MSN9: New user-centered desirability methods produce compelling visual design. Proc.*CHI 2004,* ACM Press (2002), 959-974.