Beyond Technology Transfer: Quality of Life Impacts from R&D Outcomes

Vathsala I. Stone
Michelle Lockett
Douglas J. Usiak
Sajay Arthanat
University at Buffalo, NY

Abstract

This paper presents methodology and findings from three product efficacy studies that verify the quality of life benefits resulting from prior research, development, and transfer activities. The paper then discusses key lessons learned with implications for product evaluation practice. The studies assessed the quality of three assistive technology (AT) products transferred to market by the University at Buffalo's Rehabilitation Engineering Research Center on Technology Transfer (T[2]RERC) and their value to consumers with disabilities. The purpose was to focus on outcome evaluation and seek evidence of effectiveness for the transfer process. Findings showed differences among the three products regarding their impact on end users in terms of satisfaction with product quality and product acceptance. The product most successful on all quality and value indicators was an automatic jar opener designed for persons with limited hand function. The other two–a computer software product designed to facilitate mouse pointer use by persons with limited hand function or with low vision, and a voice interactive thermostat, designed for persons with total or partial visual impairment–were less successful. They showed mixed results. Not many consumers were satisfied with the technical quality or usability of the latter two products. Of the two, the thermostat was slightly better accepted and valued by users. Differences in impact were found to follow from differences in the way evaluation information was utilized by the three product development processes. A case is made for systematic and timely use of evaluation throughout the development process in shaping a product of quality and value, in the context of the intended end users of AT.

Key words: Outcomes Research, Assistive Technology, Product Evaluation, Technology Transfer, Efficacy Assessment, Quality of Life

The T[2]RERC at the University at Buffalo, in partnership with the Western New York Independent Living (WNY-IL), has been seeking to improve the quality of life for persons with disabilities bringing new and improved technologies and products to market. Applying and perfecting a systematic process (Lane, 1999) over its three cycles of funding from the National Institute on Disability and Rehabilitation Research (NIDRR), the T[2]RERC has transferred to date over 50 technologies and products into the AT market. The development of an operational model and its demonstrated success in accomplishing technology transfer (TT) were acknowledged by experts at the State of the Science Conference held by the T[2]RERC in 2003 at the conclusion of its second cycle of funding (Lane, 2003). As Lane reports, their responses about how to advance the field of TT to the point of establishing it as an academic discipline emphasized the need to first establish TT as a formal process through continued research. The need for studies about existing models using longitudinal data was pointed out, as was the importance of continued study of the T²RERC model in other contexts and in comparative settings.

As it pursued the study of the model into its third cycle of operation, the T²RERC recognized that an extended evaluation of the model addressing its long term outcomes, as established in the NIDRR long-range plan, was in order. Accomplishing transfers evaluates only part of T²RERC's mission. It means outcomes have been achieved as intended in the form of transferred products, using a systematic transfer process. At that point, some critical questions remain unanswered: In what ways does the new product provide beneficial impacts on the quality of life of people with disabilities? To what extent do the transferred products represent improvements over existing devices that were already available in the marketplace? These questions point to a need for research "…devoted to systematic efficacy trials aimed at demonstrating how well the technologies being developed actually work" as opposed to 'letting the market-place decide' (Fuhrer, 2002, p. 13). The implied concern is about a product's impact on users, which includes product efficacy and through it, the validity or worth (Joint Committee on Standards for Educational Evaluation, 1994) of the T²RERC's transfer effort.

Following up on outcomes from its TT process, the T²RERC undertook an in-depth study of the efficacy of three selected products transferred by the center. This paper reports on this series of three studies comparing and contrasting key findings about beneficial impacts on users.

As conceptualized by Lane (Lane, 1999; Lane, Leahy, Bauer, & Stone, 2008), the transfer effort relates to the movement of technology 'from mind to market.' The T²RERC intervenes in the path of TT, at appropriate points between the 'idea/concept' stage and the final product stage, facilitating its entry into the marketplace as a commercialized product. Whether this intervention is accomplished through a demand-pull (Bauer, 2003), supply-push (Leahy, 2003), or corporate collaboration strategy (Leahy, 2005), the goal is to develop a new AT product designed to better meet the functional needs of users with disabilities, that is, better in relation to existing options currently available in the marketplace.

The product's ability to improve the user's functional capability is evidence of its value to the user and hence supports the value of the process that transferred the product. All three outcome evaluation studies presented here assessed the AT products' efficacy with focus on the intended beneficiaries – people with disabilities. In doing so, they do in fact seek the 'proof of the pudding' for the effectiveness of the T²RERC's transfer process in the context of the sponsor's mission to improve the quality of life for persons with disabilities.

In this article we present the rationale that guided the efficacy study for three products. We then present the three cases, describing the method and results for each. In doing so, we focus on a limited number of key variables common to the three cases. At the end, we summarize, compare, and contrast the three cases, draw conclusions from the overall experience, and end with a discussion of lessons learned and future directions. A summary of the T²RERC intervention into the development process of the three products is included in a later section, while full reports of individual case studies are addressed in our Resource Guide (Stone, Lockett & Usiak, 2009).

Rationale and Guiding Concepts

Efficacy as Quality and Value

Efficacy is a term used in product development practice. As a synonym of effectiveness it has been commonly used as something 'having an effect' and therefore useful or valuable, as well as something 'working well' and therefore meritorious and possessing quality. The U.S. English dictionary and thesaurus equivalents as per MSN Encarta are: "effectiveness or the ability to produce the desired result" (Microsoft, 2009a); and "effectiveness, efficiency, usefulness, worth and value" (Microsoft, 2009b). These equivalents reinforce both of the common usages, suggesting that efficacy is a global term that includes both the quality and value aspects of something being evaluated. In evaluation literature the concepts of *quality* and *value* are roughly equivalent to a product's *merit* and *worth*, the terms used to define evaluation (Joint Committee on Standards for Educational Evaluation, 1994; Scriven, 1991).

In the specific context of products devised for enhancing the functional capabilities of people with disabilities, we understand efficacy as impacting and improving their functional capabilities and independent living. This perspective ties the value of a product to consumer perception of quality and value, as well as to consumer satisfaction about how well individual needs are met. Ideally the products will meet the consumer's quality and value requirements, leading to long term use of the product to enhance daily living and independent functioning.

Each study assessed the quality and value of the transferred device in focus and sought to determine whether the device was an improvement over existing alternatives. In controlled onsite trials, the new product was compared to other products present in the marketplace at the time of transfer and was expected to offer equivalent functional benefits. In home trials, consumers also compared the new product to alternative strategies they had previously used for accomplishing the same function without the new product. Each study also assessed the value of the new product to the consumer through acceptance and use or disuse of the product over a 4- to 6-month timeframe, as well as through the consumer's response to an opportunity to acquire the product.

Evaluation's Role in Technology Transfer

Systematic evaluation is a major component of the T^{2}RERC's transfer effort. Careful evaluation helps steward transfer efforts through each step of the process, from the initial idea to new product in the marketplace. In particular, primary and secondary market research (Malhotra, 1999) captures and provides consumer and market needs. Consumer evaluations in two successive focus group interviews capture and provide features for evolving prototype versions. New product development and commercialization is the goal of the evaluation, and the principles of product evaluation are the most relevant.

Product evaluation is the most mature sector within the field of evaluation (Scriven, 1991). The concepts of *formative* and *summative* evaluations widely used in systems development and program evaluation contexts owe their origins to the principles of product development (Scriven, 1973). Tied directly to product quality and often also to value, efficacy is the focus of all product development. Quality assurance is an essential part of the product development cycle, and value is often a simultaneous concern in optimizing the product's quality. Evaluation enlightens the entire process of product development, stepping in before product conceptualization and offering guidance through prototype design and construction to final product manufacturing. Formative evaluation includes iterating cycles of testing of the prototype/product against its quality standard, followed by continuous product improvement.

Summative evaluation of the final product version assesses it against desired quality and checks its readiness for final distribution. Formative evaluation shapes the product's quality; summative evaluation verifies if the quality is at par with what was expected. The above mentioned concepts are embedded in Figure 1.

Figure 1 describes the systems approach to product development. This is our specific product development version adapted from a generalized model applicable to any system development, the CIPP (context, input, process, and product) model of evaluation proposed by Stufflebeam and colleagues (Stufflebeam & Shinkfield, 2007; Worthen & Sanders, 1973). The four types of evaluations in this model capture useful data to respectively inform four successive management decisions of design, structure, implementation, and product recycling. Need provides the basis for designing a relevant system (or product), input information makes it feasible to put together, process evaluation enables optimal implementation, and product evaluation helps improve the output to optimal quality. Evaluation thus enlightens the development process. Used beyond the process, it provides post-commercialization guidance.

Issues of quality and value are routinely addressed by professional product developers, whether as part of their technical quality control routine or in their consumer satisfaction assessments and marketing surveys. Such evaluations tend to be isolated and conducted at specific stages of the product's development and commercialization, rather than systematically span the entire development cycle as shown in Figure 1. Yet, in order to maximize evaluation's potential to enhance a product's quality and value to the intended end user, it is

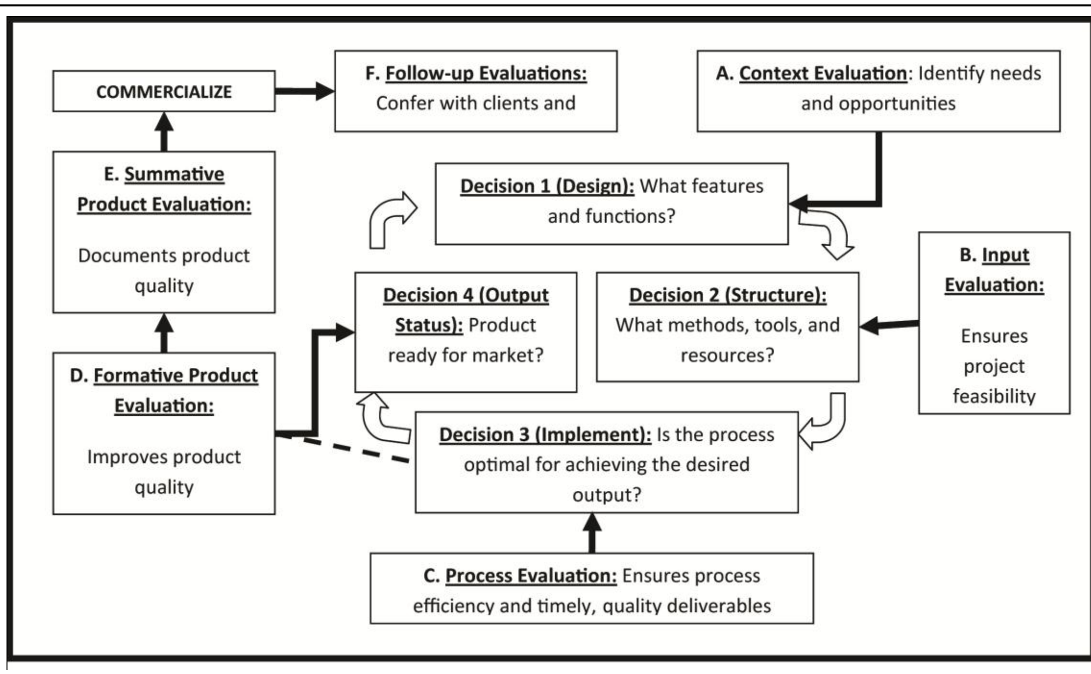

Figure 1. Evaluation enlightens development.

crucial that it occur continuously at all stages of product development and contribute to enlightened management decisions. It is easy to see from the diagram how a pro-active approach where consumer and market needs drive design priorities is superior to an after-commercialization satisfaction survey that may be too late to act upon. Likewise, if formative evaluations (Scriven, 1973) are to be valuable in perfecting product quality, they should not only be iterative but also timely, with all iterations occurring before the product is commercialized. Ideally, a summative evaluation should also occur before commercialization and document product effectiveness, including in real-life situations. This can be difficult, given practical constraints on time, finances, and personnel expertise in the industry sector. Each of these constraints has impeded the utilization of evaluation to its maximum potential with consequences such as poor product quality, product languishing in market, or users abandoning products.

The T²RERC's transfer effort bases its rationale on the Product Development Institute's Stage-Gate® framework (www.prod-dev.com; Kahn, Castellion, & Griffin, 2005), the roadmap for driving new products from idea to launch successfully and efficiently. Used by most major companies today, it brings a management rather than an academic perspective to new product development. Although less explicit and detailed as the CIPP model, the T²RERC roadmap emphasizes evaluation with sufficient concurrence and overlap with the CIPP concepts.

The T²RERC's information capture starts at the pre-design stage and spans the entire path from concept to prototype to product and, if needed, to post commercialization. Consumer and market needs guide the redesign of prototype with the relevant AT features. Consumers then evaluate the prototype in iterative focus groups and surveys and the prototype is refined before licensing to the manufacturer (Stone, 2003). Summative evaluation of the final product is usually limited to pre-production. Full-blown summative evaluations before product commercialization are hard to incorporate into corporate realities. All the same, evaluation is a key contributor to shaping the product as desired by the T²RERC technology transfer process. The quality and value of a transferred product as captured by the efficacy studies is directly tied to this role of evaluation.

Designer and Consumer Perspectives in Measuring Quality and Value

Product designers and developers are a primary group interested in product quality, as are consumers and their advocates who have an equal stake in a product's performance. A review of product evaluation literature reveals interest from both stakeholder groups expressed in different terms. Usability is often the designer expressed product quality, viewed as the optimal match of device features/functions to user characteristics (Green & Jordan, 1999). In an extended view, Popovic (1999) considers it best incorporated during the design process. It is considered best to design with users rather than designing for users, to avoid problems in the user-product interactive interfaces. The concept of usability evolved in the study of human-computer interaction, but has now broadly transcended into the world of consumer electronics (Han, Yun, Kwahk, & Hong, 2001) and home appliances (Rich, Sidner, Lesh, Garland, Booth, & Chimani, 2006). It is often considered as a reflection of the product's ergonomic quality (Babbar, Behara, & White, 2002; Dzida, 1995).

As per the International Standards Organization (1998) definition, usability of day-to-day products would determine how consumers can interact with them to successfully complete a task in relation to the effort, time and accuracy involved in using them. This approach includes considerations of both effectiveness and efficiency.

From the consumer perspective, quality standards are tied to a product's ability to meet users' expectations. Some consumer-perceived attributes of usability may include a product's ease of use, comfort in use, safety, and reliability. Whereas a designer might view 'meeting user expectations' as an end, or the result of selecting and arranging the desirable product features and functions, the consumer might consider it the starting point or a necessary condition for accepting and using the product to full satisfaction. In this sense, a product's usability and acceptance in consumers' eyes are important indicators for efficacy evaluation. Batavia and Hammer (1990) proposed a preliminary standard set of 17 consumer-expressed quality criteria that could be applied to AT devices and products as measures of their usability. These 17 consumer-expressed criteria were later refined and reduced to 10 by the T^{2}RERC's consumer ideal product study (Lane et al., 1997; Stone et al., 2009). These 10 consumer-expressed criteria are: (a) effectiveness, (b) durability, (c) reliability, (d) safety, (e) comfort, (f) learnability, (g) maintenance/ reparability, (h) portability, (i) operability, and (j) affordability.

In addition to usability, success of consumer products also depends on factors such as technical excellence, functionality, cost, and after-sales customer support (Babbar et al., 2002; Dumas & Redish, 1994; Han et al., 2001). From a consumer perspective technical excellence in design and manufacturing of a product is manifested through factors such as product's appeal or aesthetics, durability, cost and customer support.

Involving consumers in product design has been central to many concepts ranging from the well-known Universal Design (Center for Universal Design, 2007; Story, Mueller, & Montoya-Weiss, 2002; Trace Center, 2003), Design-for-all (Design for All Foundation, n.d.), Trans-generational Design (Pirkl, 1991) to the recently coined Nana-technology (Carle as cited by Jennings, 2006). Although these concepts may have their own subtleties with respect to definition, principles, cultural, and geographic significance, their commonality is in terms of their goal to facilitate the usability of products for all users regardless of their abilities or disabilities.

The above considerations guided the efficacy assessment in all three case studies. Efficacy indicators were defined using designer expertise and consumer experience, as described later in the methods section. Our approach was to combine designer insight about helping people achieve the operational objectives for which they are responsible, with the consumer's desire to use AT to fulfill life's roles (Rouse, 1991). Three examples of designer-expressed indicators of usability (International Standards Organization, 1998; Jordan, 1998) are: (a) effectiveness (i.e., the extent to which a goal or task is achieved); (b) efficiency (i.e., the amount of effort required to accomplish a goal); and (c) satisfaction (i.e., the level of comfort the user feels when using a product and how acceptable the product is to users as a vehicle for achieving their goals).

Taken together, the T^{2}RERC consumer criteria and the designer principles for usability by all represent a fairly comprehensive basis for evaluating a new product's efficacy.

Focus of the Study: Three AT Products Transferred by the T^{2}RERC

The following summarizes key characteristics of the three products assessed by the efficacy study, and salient points of the T^{2}RERC development and support activity that led to their transfer. Among the dozens of transfers by the T^{2}RERC, the impact of these three products on persons with disabilities was of unique interest. The Lids Off™ represented the best-case scenario in the sense that the transfer process involving the T^{2}RERC's intervention and the company's use of its support very closely followed the transfer model. The product was a big success in the marketplace in terms of sales volume, although it was a universally designed product, not targeting specific disabilities. The Point Smart on the other hand, did not represent the best case in terms of T^{2}RERC's full intervention but did target persons with specific disabilities. The Kelvin thermostat was of interest because of its focus on sensory impairment (persons with blindness), an especially challenging area in T^{2}RERC's experience in terms of responding to accessibility issues. All the same, each company's internal capabilities and constraints influenced their ability to apply the T^{2}RERC's input in their device development. These decisions determined the extent to which the finished product incorporated the features and functions identified by the consumer and/or designer criteria. These points are discussed in the final section of the paper.

Lids-Off™, Automated Jar Opener by Black & Decker

Lids-Off™ is an electrical household appliance manufactured by the Applica division of Black & Decker. It is designed to assist with the task of opening jars, especially useful for consumers with limited hand function. The device uses a motor driven gear system that uniquely grips and breaks the vacuum seal on a jar to unscrew its lid. It is a table top model that allows also for one handed use.

The T^{2}RERC actively facilitated the design, development, and commercialization of Lids-Off™. We provided primary and secondary market research information as well as formative evaluation (Boxes A and D in Figure 1) of the developing prototype. This included (a) desirable functions and features captured through consumer focus groups reacting to the initial prototype, and (b) step-wise input for refining consecutive prototype versions through focus groups and surveys. Market evaluation data including purchase value and intent to purchase enriched the company's commercialization perspective. However, our support by way of summative evaluation (Box E, Figure 1) before commercialization was limited to informal estimates due to the company's time lines and practical constraints. Also, post commercialization evaluations (Box F, Figure 1) were not part of the support. The several hundred thousand units of the product that were sold in the first year of its launch was possibly the pay-off of this collaborative process. More details of the collaboration on the development of Lids-Off™ from its crude prototype are described in Arthanat, Stone, and Usiak (in press).

Point Smart, Mouse Driver Software by Info Grip

Point Smart by Info Grip (2003) is mouse driver enhancement software designed to make any computer pointing device accessible for users with physical limitations, such as poor motor control or visual impairments. Point Smart works with different computer access hardware such as trackballs, a pen mouse, and other AT devices, including augmentative alternative communication (AAC). Its accessible features enable the user to navigate the pointer over the computer screen with minimal touch, or exertion, on the mouse hardware. It makes it easy to control the direction and speed of the mouse cursor and its positioning on the target.

Point Smart was conceptualized by a graduate student at the RERC on Wheeled Mobility at the University at Pittsburgh. The T^{2}RERC collaborated with him for its manufacturing and launching into the marketplace, with permission from the University at Pittsburgh Technology Transfer Office, the intellectual owners of the design and prototype. After review by Info Grip, a manufacturer and distributor of software products, the prototype interfaces were redesigned and its use was extended to include Windows® 95, 98, 2000, and XP operating systems, before it was released into the market. The T²RERC advised on the development of the product and shared in the development costs. It facilitated the design, development, and commercialization less actively than it did in the case of Lids-Off™. However, Info Grip was receptive of all the T²RERC's recommendations. The prototype was quite advanced and beyond the design (Box A in Figure 1) stage, making it too late for primary and secondary market research information, or for systematic formative evaluation. Unfortunately therefore, the product was brought to market without formal and systematic consumer involvement. Aside from T²RERC's monetary support for making the product compatible with newer operating systems, all T²RERC's evaluative support was informal. Practical constraints further caused the product to be launched to market before all technical limitations could be resolved. Also, as with Lids Off™, formal summative evaluation and follow-up evaluation were not provided by the T²RERC.

Kelvin, Voice-Interactive Thermostat by Action Talking Products

Kelvin is a voice-interactive, fully programmable thermostat designed for visually impaired consumers. It is manufactured by Action Talking Products, LLC (2008) for Innotech Systems, Inc. and distributed by Independent Living Aids. Users can operate it by pushing its buttons, all of which talk. Or, once users program the thermostat, it responds to voice commands – to lower or raise the temperature at specific times of day; or to adjust the temperature at set intervals over long periods like weekends or vacations. It can control both heating and cooling.

The T²RERC's intervention into the development and transfer of the Kelvin was similar to its involvement in the case of the Lids-Off™ jar opener. Systematic evaluative support was provided through consumer and market evaluation data for the prototype design and formative evaluation phases. Consumer-desired functions and features were identified through initial consumer focus groups and used in the development phase of Kelvin. However, follow-up focus groups were not conducted until after Kelvin was brought to market. Revision: Although post-developmental evaluation (summative evaluation of final product) was part of the T²RERC intervention, not all key features suggested by consumers were included in Innotech's alterations to Kelvin. The post-commercialized Kelvin, therefore, lacked these features. Another important difference between Lids-Off™ and Kelvin was in the production phase. Both devices were produced overseas, but Black and Decker owned the manufacturing plant and had direct control over its manufacturing protocols, where as Innotech outsourced its operations, tying Kelvin to the quality controls used by the outside manufacturer.

Evaluative Questions

The purpose of the T²RERC efficacy study was to investigate the quality (merit) and value (worth) of the project's transferred products, based on how well they met the needs of end-users with disabilities, the project's ultimate beneficiaries. Two main questions drove the study.

Table 1
Lids-Off™ Study Participant Distribution by Functional Limitation, Age, and Gender

Functional Limitation	*n*	%	Age group	*n*	%	Gender	*n*	%
Paralysis	12	24	25 – 34	3	6	Male	10	20
Weakness	47	94	35 – 44	9	18	Female	40	80
Tightness or Cramping	33	66	45 – 54	12	24			
Tremors	11	22	55 – 64	23	46			
Lack of Control	24	48	65 – 75	3	6			
Absence of Extremity	1	2						
Joint Restriction	33	66						
Swelling	16	32						
Fatigue	41	82						
Pain	37	74						
Total	50		Total	50	100	Total	50	100

1. How do products transferred through T²RERC compare in *quality*, with other products and or methods available to consumers with disabilities at the time of transfer?
2. To what extent do users with disabilities value the products transferred through the T²RERC, compared to alternatives available to them at the time of transfer?

The questions directed the methodology for studying the efficacy of the three products mentioned earlier. We next present and discuss the method and results, case by case. As mentioned earlier, we present findings selectively focused on key indicators of quality and value that were common to all three products under study. We then follow it by discussing contrasts and commonalities. For additional discussion of aspects unique to each study readers are referred to Stone et al. (2009). Findings were also appropriately synthesized and fed back to the respective manufacturers for product improvement.

Case One: The Lids-Off™ Automated Jar Opener

The Lids-Off™ study was the front runner of the efficacy study series. It helped us learn about both the efficacy of Lids-Off™ and the effectiveness of the transfer process. By piloting our proposed methodological framework, it also provided a master plan for the two subsequent studies on Point Smart and on Kelvin, so that iteration of procedures helped improve and consolidate the methodology itself. Contextual adaptations of the methodology are addressed separately under each case.

Method

Procedures

Described below are the procedures followed in the Lids-Off™ study for sampling, data collection, and data analysis.

Sampling. Lids Off targeted consumers with limited hand control. Sampling was purposive. As consumers with disability were the 'experts,' evaluating the device for merit and worth, we sought to maximize, in a limited participant sample, both the variety of functional needs (vis-à-vis jar opening) as well as related consumer experience. Physical impairments with upper extremity limitations and discomfort were included while excluding cognitive limitations that interfered with the ability to judge and report on device performance. The heterogeneity and the relatively small size of the target disability population drove the final sample size. From an initial sample pool of more than 100 qualified individuals, we were able to form a random sample of 50 satisfying all the purposive criteria.

Table 1 shows the distribution of the sample of 50 participants by functional limitation as well as by age and gender. The sample covered a variety of disabilities, such as spinal cord injury and multiple sclerosis, and the majority had arthritis.

As the table shows, weakness, fatigue and pain in hand mainly characterizes this group. There were more female participants than male participants. Participants' ages ranged from 25 to 75. The median age was 55, with 48% younger than 55 and 52% older.

Data collection design. The study was conducted in three distinct phases. Phase I defined indicators of quality and value, which then directed the design of instruments for data collection. Data was collected in Phases 2 and 3, following a quasi-experimental design (Campbell & Stanley, 1963), with participants evaluating the device in two situations through onsite and home trials as described below.

Onsite trials. In a randomized, post-test only design for repeated measures, the participant consumers evaluated the Lids-Off™ against a competing product on given indicators in systematic hands-on trials of both products. The competing product was an under-the-cabinet mountable device we identified as the marketplace competitor at the time the Lids-Off™ was brought to market, which we omit naming. Each user tried the Lids-Off™ and its competitor in a pre- determined, randomized sequence. They performed a standardized set of tasks, opening five different jars of varying sizes and combinations of weights, sizes, jar materials, and lid materials. In order to minimize participant learning from device to device during the trials, we randomly assigned participants to the product testing sequence. Participants gave detailed evaluative feedback on each task, using questionnaires that were provided in accessible formats. Trained observers recorded their performance on separate sheets. Additionally, as participants exited, they were interviewed for comparative evaluation of the test product against its competitor. They were also asked to assess (estimate) the product's (monetary) value and share their purchase intent. Sessions were video recorded to facilitate post-trial measurements of task completion time and other analyses.

Home trials. Participants who tried the device in their homes were asked to give evaluative feedback, comparing it to similar devices and or methods (critical competitors) that they had used or were familiar with. The duration of the trial period was six months. Longitudinal data was collected over the first two months in a series of six weekly measures that consisted of participant ratings and comments on given indicators. Additionally, participants also gave feedback at the beginning (Day 1) and after two months. Changes in participant perceptions of quality were tracked by repeating questions across questionnaires on key indicators. Changes in participants' acceptance of the product and its value were tracked by asking participants to share their purchase intent, first at the onsite interview and again at the interview at the end of two months of home trial. Further, by letting participants voluntarily use the device with no obligation for formal feedback during the remaining months, and by questioning them later, we measured the extent of participants' use or non-use of the product during that period. Finally, at the end of the study, participants were offered the opportunity to purchase the product in exchange for part of the compensation due to them. This opportunity to purchase was a quantifiable measure par excellence for assessing the real value of the product to the user.

Indicators. Rather than deriving indicators solely based on theory, we identified indicators through actual empirical observations of consumers performing device-related tasks (opening jars). This was done in Phase I. Seven consumers who customarily used a variety of methods to accomplish the jar opening task were interviewed in their homes and were observed and videotaped as they performed the task. The video-recordings were submitted to task analyses by a team of designers and clinicians, who extracted problem statements pointing to indicators. These designer perceptions, together with consumer perceptions, served as criteria when defining final indicators. We tabulated and distributed them in a two-dimensional matrix, with the universal design (UD) guidelines as one dimension and the T²RERC device evaluation criteria as the other, both discussed earlier. This Indicator Matrix gave us a framework in which to map specific indicators of product quality and value reflecting designer and consumer perspectives. This matrix was refined over the course of the three case studies. As a tool for organizing indicators for efficacy assessment it marks an outcome from the overall experience.

The final set of quality indicators included effectiveness and efficiency measures, usability measures (i.e., ease of use, comfort, operability and learnability), and product-specific measures such as durability. Value indicators addressed the relevance and or benefits to users, including: (a) satisfaction and benefits perceived from actual use, (b) device use or abandonment, (c) purchase intent, and (d) response to purchase opportunity. The indicators, identified thus, generated instruments that guided the next two phases.

Instruments. We distributed the indicators appropriately over instruments for measuring consumer-perceived quality and value both under controlled conditions (onsite trials) and under free and natural conditions at home over an extended period. Besides protocols and scripts for trial-administration, the onsite trial instruments included: (a) two separate questionnaires for consumers to record evaluations of the device and its marketplace competitor, (b) exit interviews with consumers to elicit comparison of the two devices, and (c) two separate questionnaires for observers to record objective assessments of consumer trials of the two devices. Home trial instruments included (a) initial questionnaire on Day 1 to capture consumer's first impressions and learnability data; (b) weekly questionnaires for consumers to record evaluations of the device against other known alternatives; (c) comprehensive consumer questionnaire at the end of two months (EOT); and (d) two telephone interviews, one halfway through the home trial and the other at the end of the study. Purchase intent and value questions were part of all instruments through onsite and home trials, and of the home trial telephone interviews. We followed this up with an actual offer of the device for purchase at the very end of the study, in order to assess product acceptance and how much the consumer really valued the product. Consumers also answered questions about frequency of use and abandonment at this time.

Data analyses. Both descriptive and inferential analytical techniques were used as appropriate. We used percentages for description and supplementary content analyses to interpret narrated consumer comments, purchase intent and use/abandonment data. ANOVA and paired t-tests were performed for statistical inference for comparative analyses between Lids-Off™ and its onsite competitor as well as for weekly trend analyses of home trial data.

As for judgment standards, we found no previous benchmarks for 'acceptable levels of impact,' theoretical or practical. In other words, how good the results on quality and value have to be in order for Lids-Off™ to be considered a 'worthy' transfer? In a sense, Lids-Off™ was selected for our pilot study because it was a best-case scenario; we let the results speak and enlighten us about such a standard – if and how far the product of such a scenario can go to achieve its potential.

Results

We present findings focused on select key variables common among the Lids-Off™ study and the other two cases. Also, we focus on descriptive and qualitative data in order to provide in-depth views of each context and to explain how the three cases differ. Some analyses reported in Stone et al. (2009) are excluded here.

Sample Size and Attrition

The original sample of 50 completed onsite trials of the Lids-Off™ and continued to home trials. Three participants dropped out subsequently leaving a sample population of 47 and a dropout rate of 6% – the lowest of all three studies.

Indicators and the Indicator Matrix: Two Related Outcomes

The indicators of quality and value for evaluating the Lids-Off™ were derived in Phase I by observing consumers as they performed jar opening tasks. Table 2 shows the number of indicators distributed along two dimensions of an Indicator Matrix, with designer and consumer perspective criteria for

Table 2
Indicator Distribution by Designer and Consumer Perspectives in the Lids-Off™ Study

		T²RERC CRITERIA FOR DEVICE EVALUATION*											
		Effectiveness	Durability	Portability	Operability	Maintenance/ Repairability	Comfort	Reliability	Safety	Learnability	Affordability	Other	Total
UNIVERSAL DESIGN PRINCIPLES**	Equitable	18	3		8	3	8	3	3		3	8	57
	Flexible	27		12	27	6	22		3		2		99
	Simple and intuitive				5		1			1			7
	Perceptive information				4								4
	Error tolerant	2			9		4		14		5	1	35
	Low physical effort	8			22		5					1	36
	Appropriate size and space				13	2	1						16
	Total	55	3	12	88	11	41	3	20	1	10	10	254

*Lane et al. (1997)
** The Center for Universal Design (2002)

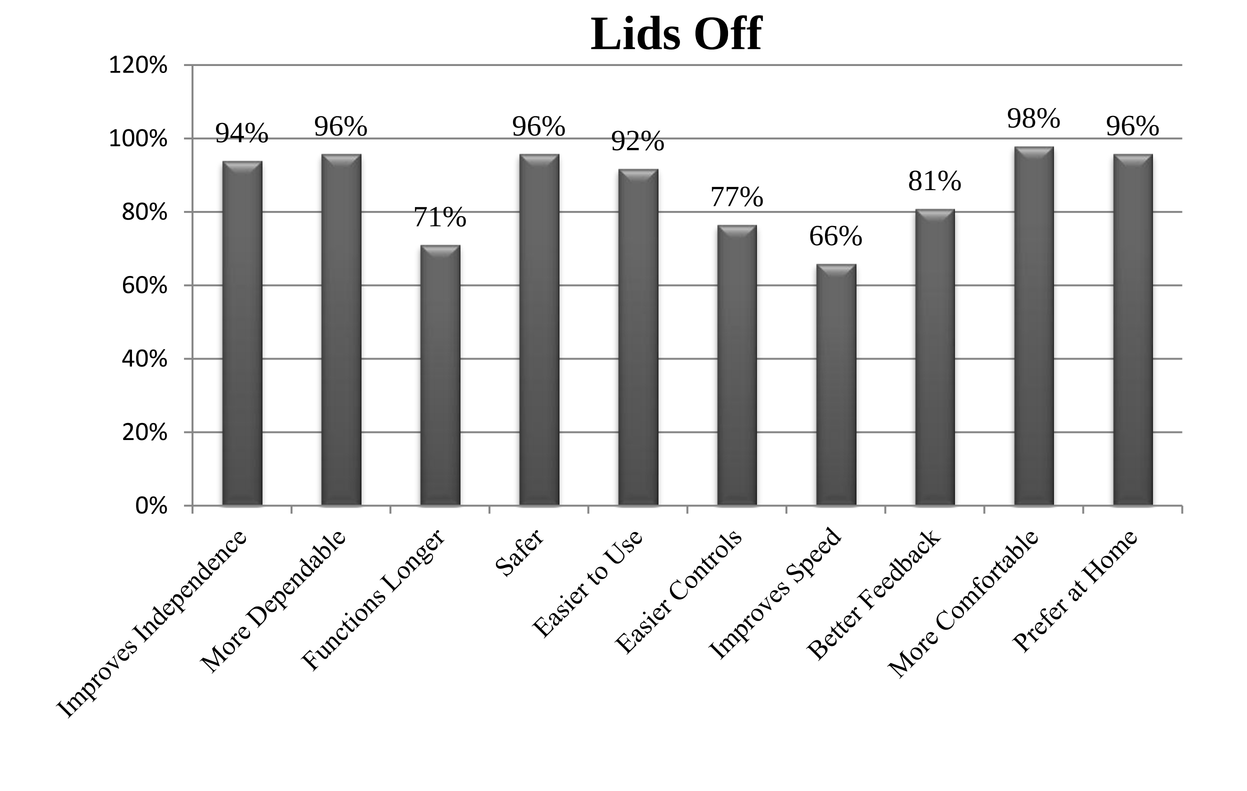

Figure 2. Percentage of participants who preferred Lids-Off™ over its competitor.

the two dimensions. The clustering pattern (distribution) of the indicators across the cells was a rough guide to generating items for instruments and observation protocols for the onsite and home trials with proper weights. For example, effectiveness, operability, and comfort demanded greater weight.

We reiterate that as the front runner of the study series and pilot experience, the Lids-Off™ case was an opportunity to improve and develop the methodological components. The Indicator Matrix in Table 2 developed and refined as a framework for organizing and consolidating efficacy indicators was a useful, although secondary, outcome from the studies; it provided the master template for subsequent studies.

User Assessments of Product Quality

We present below assessments of the product's quality by the participants.

Compared to Marketplace Competitor (Onsite Trials)

As described earlier, participants evaluated Lids-Off™ at the onsite trials against a power-assisted jar opener, mountable under the cabinet. Participants opened five food jars that included a variety of jar and lid types, using each device in the determined sequence. At the exit interview after the trials, participants' comparative evaluations of 'device-versus-competitor' were captured. Figure 2 shows these results on eight key indicators. For each indicator shown on the X-axis, the corresponding column in the figure shows the percentage of participant-users who preferred Lids-Off™ over its competitor.

As the figure shows, the percentages overwhelmingly favor (66%-98%) Lids-Off™ on most indicators. Consumers judged the device superior to the other product in improving functional independence.

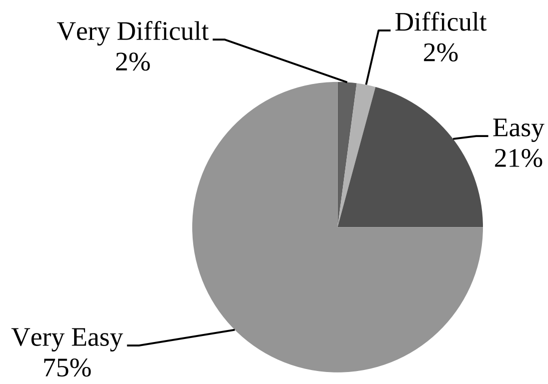

Figure 3. Ease of use of Lids-Off™ operating instructions: User perception on Day 1.

Participants generally deemed it safer, easier, more comfortable to use, and more dependable. Many found it faster and they preferred it for use at home.

Compared to Critical Competitors at Home

The following results include participant assessments over the home-trial period including Day 1 and after two months of use.

On Day 1. Learnability of a product (instructions, setting up) is an important indicator for disability populations challenged for independent operation. We measured this on Day 1 after Lids-Off™ was set up at home. (We point out that participants did have some, but not total, learning from onsite trial). Figure 3 presents user evaluations of its intuitiveness and its learnability through the instructions manual. This simple circle graph shows the percentage distribution of people who rated the manual of operating instructions easy or difficult on a five-point

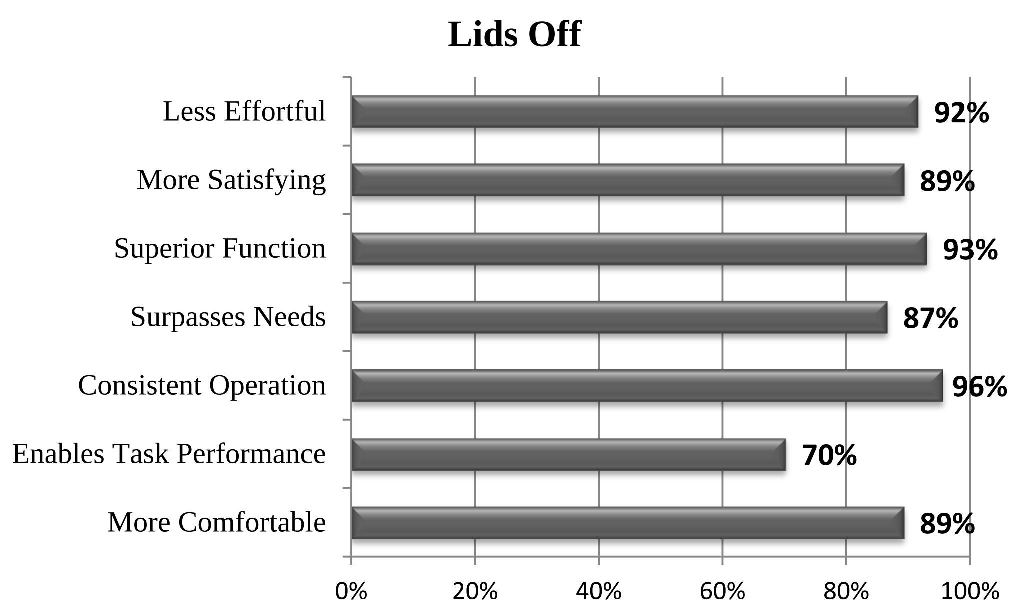

Figure 4. User judgment of Lids-Off™ quality relative to competitors (at two months).

scale. The legend shows the five specific scale points color coded, while the graph shows the corresponding percentage of people who rated Lids-Off™ at levels of easiness from very easy to very difficult.

As we can see, learning to operate Lids-Off™ operation was not very challenging. Learning was considered easy to very easy by 96%, with only 4% considering it difficult or very difficult. We point out that the instructions manual was simplified by its physical design, which was relatively straightforward and intuitive–unlike the other two devices that had complex interfaces to contend with for operation.

After two months of home use. Figure 4 shows user perceptions of Lids-Off™ quality at the end of two months of home trial. For seven selected indicators, it presents the percentage of participants that had 'positive' perceptions, i.e., those who rated at the higher end of the five-point scale (e.g., 4 or 5).

The figure shows positive ratings to be consistently high, ranging from 70% to 96% across all indicators. Participants showed a clear liking for Lids-Off™ over other alternatives they had used or known. They found it consistent in operation, functionally superior, less effortful, more comfortable and more satisfying. As many as 87% of the participants acknowledged that it surpassed their needs.

Over the home-trial period. Figure 5 compares user perceptions between the beginning (Day1) and after two months of home trials on three key indicators. These are 'independent use' (user can operate device without assistance), 'ease of use' of the device, and 'device for me' (device fits user needs). When measured as before-and-after changes in user perceptions over the seven-week trial period, they measure impact of the devices on users' functional capabilities. Perceptions are presented in the graph as percentages of positive ratings, i.e., 4 or 5 on the five- point rating scale. The X-axis shows the three indicators, with paired columns of percentages of positive ratings for each indicator, one for Day 1 and the other after two months of use.

Figure 5 shows overwhelmingly high percentages (84% to 98%) in terms of user perception of quality of Lids-Off™, and consistently so from beginning to end. The apparently small increase or change is explained by higher beginning levels. The highest change (9%) was in accepting the device as 'a fit to their needs.' Figure 6 corroborates the above results, through the

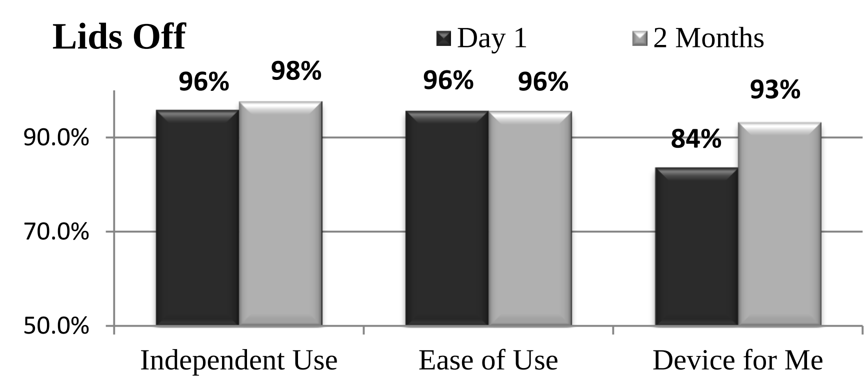

Figure 5. Change in user perception of Lids-Off™ quality over two months.

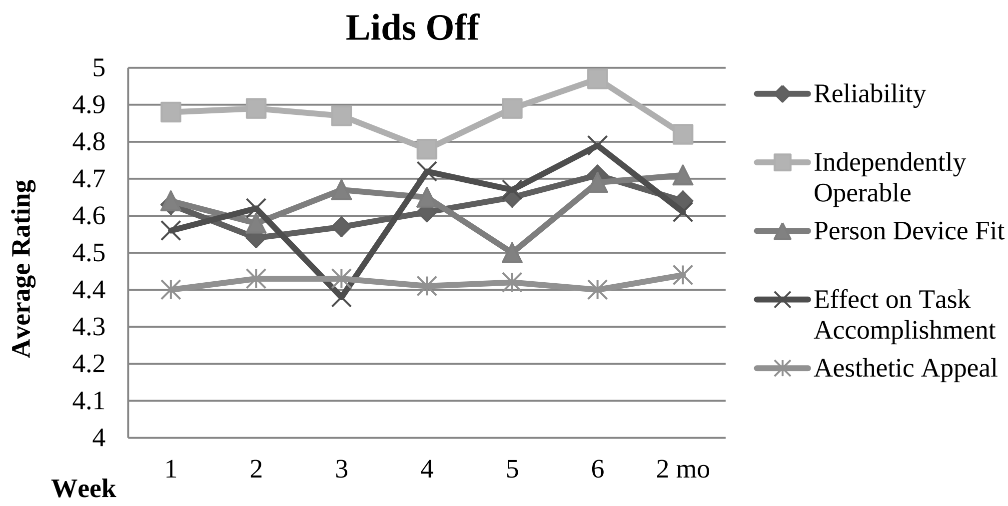

Figure 6. User perception of Lids-Off™'s quality over home trial: Weekly trend.

weekly trend of user perceptions of quality over the home trial period. The graphs in the figure trace the mean ratings on five important indicators – reliability, independent operability, person-device fit, task accomplishment, and aesthetic appeal. Ratings on Lids-Off™ were generally high (4.5 to 5.0) from beginning to end, with slight increases on all indicators.

Product Value to User: Acceptance, Use and Purchase

The foregoing section reported Lids Off™ as a success in terms of user satisfaction relative to its merit. To what extent did they consider it relevant to their needs? What was the level of acceptance of the product for own use? How much did they value it as a result of their experience?

Product Acceptance

Product acceptance was high for Lids-Off™. Only three participants dropped out of the study during the home trials, and they did so for reasons unrelated to device usability or value. Additionally, participant comments at the end of two months of home use (see Table 3) point to satisfaction with features that are key to usability (reliability, ease of use) and their dissatisfaction is limited to storage and accommodation (size and cord length)

Table 3
Participant Perceptions of Device Usability: After Two Months of Home Use

Feature most liked by participants	*n*
Ease of use	24
Portability	5
Aesthetic appeal	5
Reliability	10
Features least liked by participants	*n*
Incompatible with working/storing environment; is too big for narrow counter spaces;	7
cord is too short for plugged-in storage on counter spaces	3

Table 4
Use and Abandonment of Lids-Off™ by Study Participants

Frequency of use	*n*	%
Every Time	23	49
Most of the Time	13	28
Some of the Time	7	15
Rarely	2	4
Very Rarely	2	4
Never	NA	NA
Total- end of home trial	47	100
Drop outs/Missing	3	6

concerns.

To explain their preferences, participants added comments such as: "Lightweight, easy to move," "Worked every time," and "Attractive machine." On the other hand, participant comments related to features least liked included: "Electrical cord too short. Limits area where it can be used," "Short cord. A bit cumbersome," and "It's too big; takes up too much space. I can't keep it plugged in at all times because it's too big for my counter," "Its large size. Cord does not plug into Lids-Off™ itself."

Product Use

The use and abandonment data from Table 4 below attest to the acceptance of Lids-Off™ by its users, and corroborate the foregoing results. During the latter four months of home trial period when participants had no obligation to give feedback to the study, they continued to use the device as shown in this table. Only 8% (four people) used it rarely, and no one said 'never.' The reasons mentioned for rare use included: (a) "Did not work well for me;" (b) inconvenience with cord; (c) disability status fluctuation; (d) getting help in opening jars; (e) "small kitchen; hard to move device around;" and (f) "I really don't know."

Product Purchase

As mentioned earlier, the purchase

Table 5
Purchase Intent vs. Purchase Decision by Participants in Lids-Off™ Study

Situation	Question	Would buy the product		Total
		n	**%**	
Onsite Trial Exit Interview	*Which one would you buy? – Product or its Competitor?*	46	92	50
Mid-Home Trial (2 months)	*Likely to trade part of study compensation to buy product?*	35	70	50
Mid-Home Trial (2 months)	*Would you buy your original device again?*	29	58	50
End of Study (6 months)	Actual Decision to Buy	37	79	47

opportunity we posed to consumers at the very end of the study as part of the compensation due to them was intended to assess the value of the product to the user. We fixed the purchase price at half its market value, a fair price for a 'used' device. Table 5 tracks participants' purchase trend over the course of the study – from intent to actual decision.

As shown in Table 5, 50 people started the Lids-Off™ study, 47 completed it, of whom 37 chose to purchase the product, giving up part ($15-half the retail price) of the total compensation ($200) due to them. The Lids-Off™ seems to have offered 'value' to nearly 80% of users that completed the study.

Case Two: The Point Smart Mouse Driver Software

Method

Procedures

The procedures described under the Lids-Off™ study guided the study of efficacy of the Point Smart software as well. Within the intended uniformity however, contextual adaptations of procedures introduced some variations as described below.

Sampling

Point Smart was aimed at consumers with limited hand control and/or low vision. As with Lids-Off™, sampling was purposive. The priority was to maximize consumer experience and to assemble a sample population with a variety of functional needs that demanded the use of a better mouse for navigating the computer screen. The target disability population was more heterogeneous and smaller. The final sample size was 32. Disabilities ranged from arthritis to diabetes to spinal cord injury; one individual had an added difficulty of having no voice. Table 6 shows the sample distribution by functional limitation, as well as by age and gender.

As can be seen, twice as many participants had motor difficulties as had visual problems. Participants' limitations required them to use mouseware accessories to access the Point Smart software on the computer, therefore requiring complex hardware interfaces as well. Eighteen individuals used a standard mouse; others used a trackball (*n*=9), touchpad (*n*=2), joy stick (*n*=1), pen mouse (*n*=1) and Dynabeam/Dynavox (*n*=1). The logistics of enabling complex hardware interfaces were therefore unique to the Point Smart study. There were more female participants than male participants in the sample. The age range was 18-70 and median age 49.

Data Collection Design

Following the design for Lids-Off™, data was collected in Phases 2 and 3, after identifying indicators and building instruments in Phase 1. The basic quasi-experimental design was

Table 6
Point Smart Study Sample Distribution by Functional Limitation, Age, and Gender

Functional Limitation	*n*	%	Age Group	*n*	%	Gender	*n*	%
Hand control	26	81	18 – 24	5	16	Male	13	41
Visual difficulties	15	47	25 – 34	3	9	Female	19	59
			35 – 44	5	16			
			45 – 54	7	22			
			55 – 70	12	37			
Total	32		Total	32	100	Total	32	100

followed with slight modifications.

Onsite trials. One variation in design was dispensing with random assignment of participants to the product testing sequence, which did not make sense in the case of Point Smart. All participants came with prior knowledge of Microsoft®'s mouse software, which made it the competing software by default. There was new learning on the Point Smart software but not on the Microsoft® mouse software. All participants thus tried the Microsoft® software first and then the Point Smart. We planned an additional measurement of participant performance with the Point Smart at the end of the home trial. Assuming possible learning only on the Point Smart, we took the difference between first (Microsoft®) and third (Point Smart) performances as a measure of comparative efficacy, while the difference between the second and the third (both Point Smart) gave us an absolute measure of efficacy based on pre-post gains. Another variation was the use of an additional instrument for the measurements. The Compass Assessment Software designed by Koester Performance Research (2002), a software program that was also brought to market by Info Grip, measures eight point-and-click skills of computer interaction necessary for tasks such as text composition, web navigation and electronic communication, and configuring and customizing tests for the user. Its speed and accuracy data gave us the needed time-per-task data, which dispensed the need for video recording of the onsite trials. A third variation regarded the set-up of onsite trial sessions. Several logistical provisions became necessary in order to accommodate different disability groups with motor and sensory (visual and communication) impairments. Participants came with their own accessible mouse hardware (foot-operated, pen mouse, head mouse, augmentative communication devices, and others). Computer settings had to be customized for onsite clinical trials and also pre-determined and prescribed for later home trials. A clinician expert worked with us at these trials to configure support systems and assist observing researchers. All the same, as with Lids-Off™, participants: (a) performed the same standard tasks–web browsing, emailing, and simple word processing–using each software program; (b) participants gave feedback on questionnaires and on exit interviews; and meanwhile (c) observers recorded their performances as well.

Home trials. As in the case of Lids-Off™, participants performed tasks of their choice using Point Smart and gave weekly feedback on questionnaires on their use of the software. They also gave feedback on Day 1 and at two months. Quality and Value questions were repeated across questionnaires and interviews. Unlike the Lids-Off™ case however, Phase 3 lasted only four months, given the tedious nature of the tasks and participants' energy levels. Monitoring and tracking logistics was complex. Interacting with the Point Smart software through special mouse accessories at home required special support systems and equipment. Computer platform and system compatibility was a concern, and occasional technical assistance by Info Grip became necessary. Compatibility with computer hardware (for example, a laptop) and assistive/adaptive mouse hardware such as foot operated mouse, and others was also an issue.

Indicators and instruments. Procedures were the same as with Lids-Off™. In Phase One, six persons with disability were interviewed at home for extraction of indicators. Onsite trials and over-the-home trials followed the same design as for Lids-Off™, with corresponding consumer questionnaires, observer questionnaires, and interviews. Participants also had the opportunity to purchase the software at the end, in exchange for part of the monetary compensation due to them.

Data analyses. Both descriptive and inferential techniques were used. Additionally, individual by individual analyses were initiated and are underway (Stone et al., 2009), not reported in this article.

As we had no previous benchmarks for measuring 'acceptable levels of impact,' we were guided by the results of our pilot study on Lids-Off™ in interpreting from the results whether Point Smart was a 'worthy' transfer. In a sense, Lids-Off™ set the practical standard as to the heights to which a transfer can reach in achieving quality and value for the consumer.

Results

We reiterate that findings are focused on select key variables that the Point Smart study has in common with the other two cases. Additionally, we will focus on descriptive and qualitative data in order to capture context-specific information that can explain its differences from the other two cases.

Sample Size and Attrition

Of the 32 participants who initiated the study as per the previous table, only 25 completed the home trials. This is a drop-out of 72%, which is more than the attrition in the Lids-Off™ study. Several withdrawals were due to incompatible computer hardware. We next present findings on participant assessment of the quality of the Point Smart based on onsite trials and home use.

User Assessments of Product Quality

We present below assessments of Point Smart's quality by the participants.

Compared to Marketplace Competitor (Onsite Trials)

As mentioned earlier, participants compared Point Smart with the Microsoft® mouse software at onsite trials, performing the given set of standardized tasks. The exit interview after the trials captured the participants' comparative evaluations of Point Smart vis-a-vis Microsoft® software. Figure 7 shows these results on the same eight key indicators selected for Lids-Off™ reporting. The indicators are shown on the X axis, and the corresponding columns show the percentage

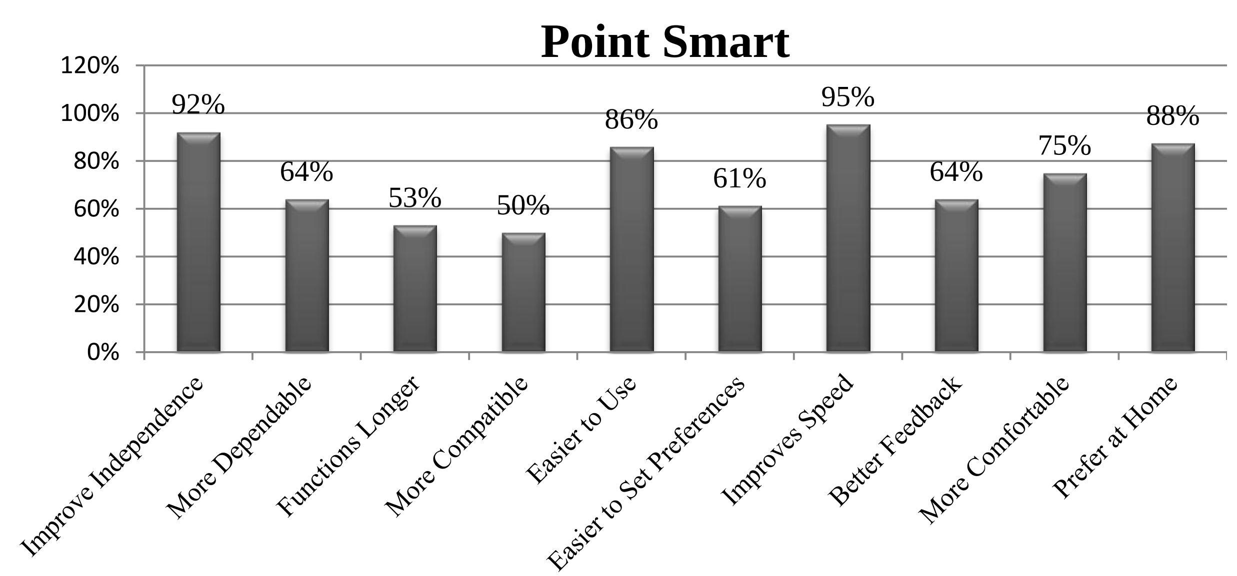

Figure 7. Percentage of participants who preferred Point Smart over its competitor.

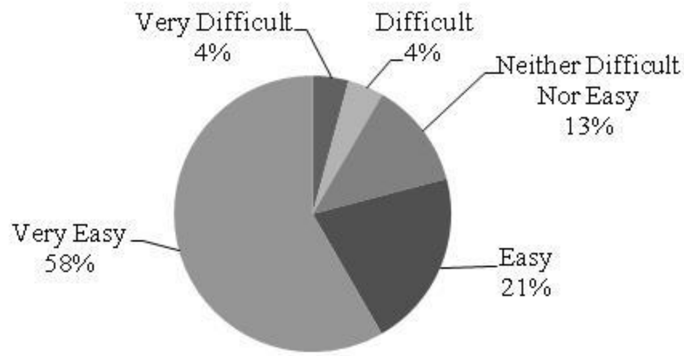

Figure 8. Ease of use of Point Smart's operating instructions: User perceptions on Day 1.

of participant users who preferred the Point Smart over its competitor.

Figure 7 shows that over half (50%–95%) of the participants preferred the Point Smart over its competitor on these indicators, a positive result in favor of Point Smart. A major proportion (95%) rated Point Smart superior to its competitor Microsoft on speed improvement. This is a meaningful finding for the Point Smart user population, considering their need of for multiple accessory interfaces for functional independence, which potentially inhibit speed.

Compared to Critical Competitors at Home

The following results include participant assessments over the home-trial period,

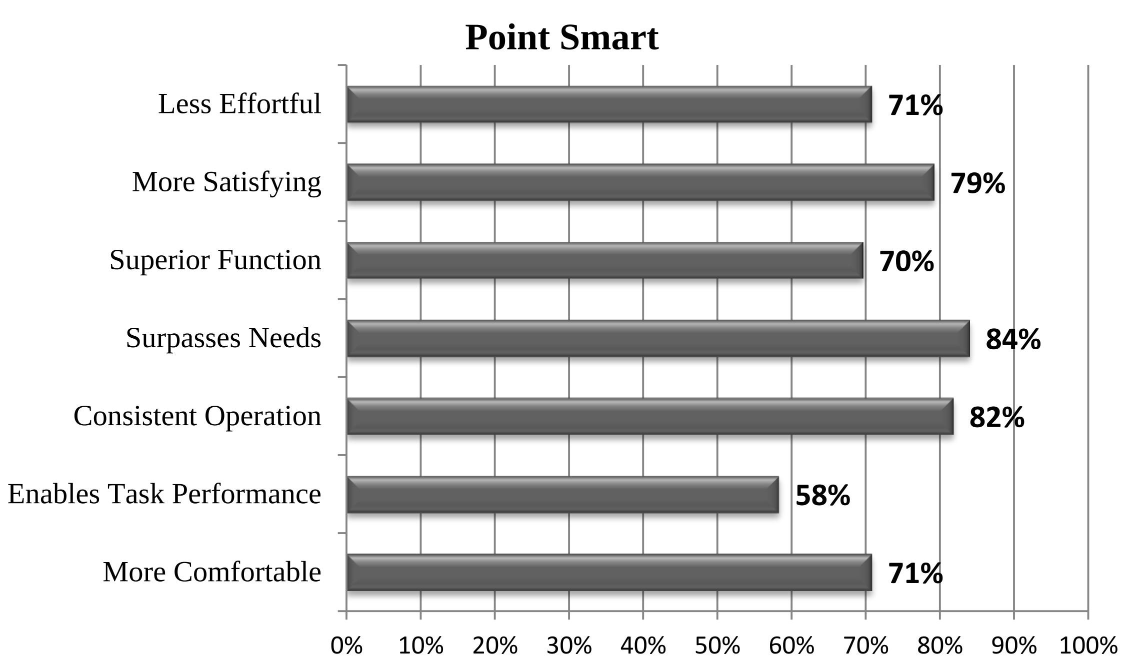

Figure 9. Point Smart's quality relative to alternatives: User assessment at two months.

including Day 1 and after two months of use.

On Day 1. Figure 8 captures Point Smart's learnability and intuitiveness soon after it was set up at home. The circle graph shows participants' evaluations of the operating instructions manual on ease of use, distributing percentages of people who rated the manual easy or difficult on a five-point scale. The legend shows color codes for the five specific points of the scale, while the graph shows the corresponding percentage of ratings at levels of easiness from 'very easy' to 'very difficult.'

More than three-quarters of the sample (79%) considered the Point Smart 'easy' to learn. On the other hand, 8% found it difficult to very difficult. These results are not surprising, considering that Info Grip provided a standard manual of instructions, downloadable from its website, for use by all of its customers. Ironically, this defeated the purpose for those users who could not navigate computer screens for downloading tasks, for which they were seeking Point Smart in the first place. Stabilizing the mouse cursor on the screen was a challenge for this population, and basic tasks like placing it on a desired icon, or a word or a letter needed help. Alternative (print or other) versions would have been more appropriate.

After two months of home use. Figure 9 shows user assessments of Point Smart's quality at the end of two months of home trial. For the seven indicators, it presents the percentage of positive perceptions, i.e., ratings at the higher end of the five point rating scale (e.g., 4 or 5).

Figure 9 shows moderate ratings ranging from 58% to 84% on the indicators presented. Notably, Point Smart earned the satisfaction of as many as 84% of its target sample who acknowledged it as 'surpassing their needs.' Interestingly though, less than 50% were willing to buy the Point Smart at this point. In open comments, many reported frustration with unresolved technical problems, inconsistent performance, and hardware compatibility issues.

Over the home trial period. Figure 10 compares user assessments of Point Smart between the beginning (Day 1) and after two months of home trials, on three key indicators–'independent use' (i.e., user can operate device without assistance), 'ease of use' of the device, and 'device for me' (device fits user needs). Viewed as before-and-after changes in user

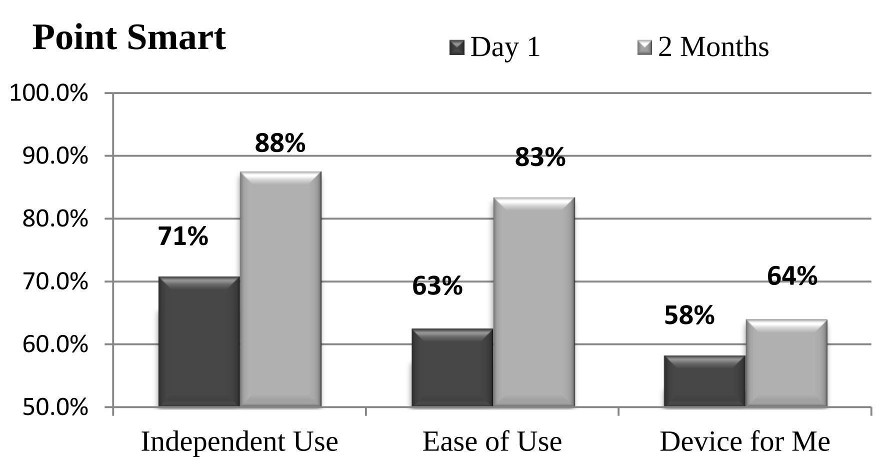

Figure 10. Change in user perception of Point Smart's quality over two months.

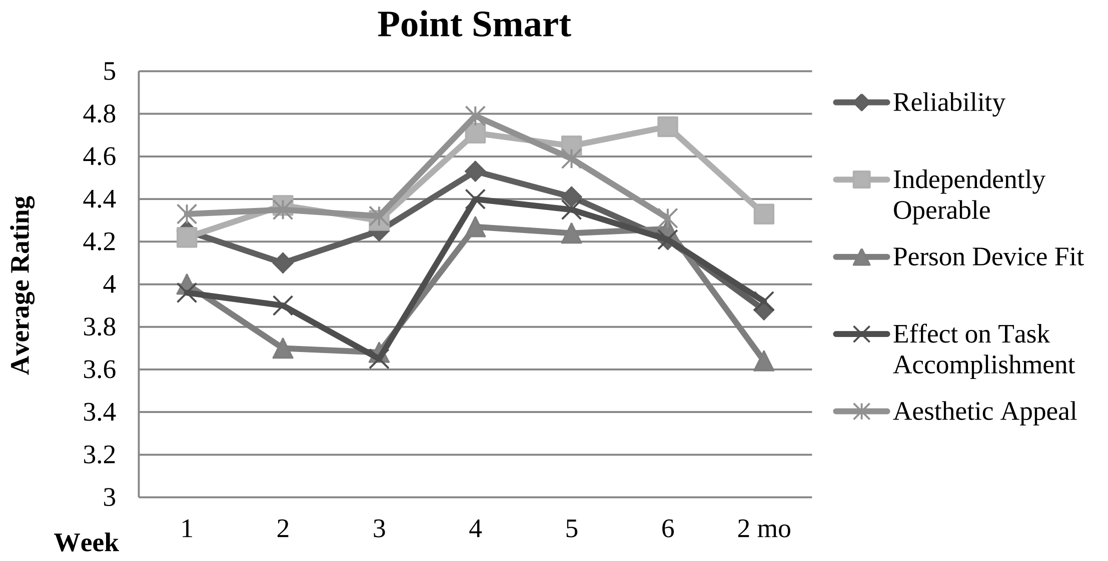

Figure 11. Participant perception of Point Smart's quality over home trial: Weekly trend.

perceptions over the seven-week trial period, they measured device impact on users' functional capabilities. Percentages presented in the graph are of positive ratings, i.e., 4 or 5 on the five-point rating scale. The X axis shows the three indicators, with paired columns of percentages of positive ratings for each indicator, one for Day 1 and the other after 2 months of use.

As shown in Figure 10, ratings were not too low on Day 1 (58%-71 %) and they increased on all three counts. Gains were 20% (from 63 to 83%) on 'ease of use' and 17% (from 71 to 88%) on 'independent use.' Note however, that only 64% elected to characterize it as a 'device for me' in the end, with only an 8% change during the trial period. Point Smart did not reach a high level of acceptance by users after two months of home use.

Figure 11 shows the weekly trend of user perceptions of quality over the home trial period and corroborates the above results. The graphs trace mean ratings on five important indicators–reliability, independent operability, person-device fit, task accomplishment, and aesthetic appeal.

Ratings started out high (4.0 to 4.5) on all indicators. Most showed an initial increase but declined after Week 4. Note in particular the oscillating ratings on person-device fit, which dropped to 3.5 in the end, while ratings were relatively higher on independent operability.

Product Value to User: Acceptance, Use, and Purchase

The following results address participants' acceptance, use, and purchase of Point Smart.

Product Acceptance

Dropouts and the reasons for them fairly indicate a product's acceptance during home trial. Seven out of the 32 initial participants (22%) dropped out of the Point Smart study during the home trials. The reasons were partly related to hardware interface issues and partly to do with their dissatisfaction with Point Smart's usability. On one hand, it was incompatible with some computer platforms or mouse hardware (touch pad, pen mouse). On the other hand, features that made it uniquely accessible–such as 'button gravity' and 'automatic direction control'–did not always work reliably.

Participant perceptions of Point Smart's usability in responses to the end-of-two-months questionnaire corroborate foregoing difficulties and disappointments. On one hand they pointed out three features that they liked most: wrap-around (n=12); large, bright pointer (n=8); and drag-and-drop (n= 4). Participants' comments that attest to these preferences included: "Wrap around, animated pointer and gravity;" "wrap around function and the intelligent cursor positioning;" and "the larger pointer and the bright green color of the pointer."

On the other hand, participants pointed out five features that they liked the least: the gravity feature (*n*=3); the wrap-around feature (*n*=3); the automatic direction control and enable button gravity feature (*n*=1); automatic cursor positioning (*n*=1); and increased crashes (comment that occurred frequently throughout the questionnaire). Comments that attest to these perceptions included: "Crashing software, not letting me use the tablet and trackball at the same time;" "Automatic direction control and enable button gravity function;"and "Automatic cursor positioning and speed control."

Although the foregoing issues do not explain cases of earlier dropouts, they shed light on general problems and probable sources of dissatisfaction with the device. Point Smart's acceptance level was not overwhelmingly high. Consistency of operation was an issue, and there was discrepancy between promise and delivery of features advertised; showed it did not reach its potential and made it less acceptable than expected.

Product Use

The foregoing explains the data presented in Table 7 on use and abandonment of Point Smart.

These data points relate to voluntary use in the last phase of home trials, when participants had no obligation to give feedback to the study. About 73% continued to use the device every time or most of the time. About 14% used it rarely or 'never.' Comments related to rare use included: "Wasn't working;" "Not sure why it wasn't working." Considering the reported difficulties, these results are not surprising.

Product Purchase

The purchase opportunity posed to consumers at the end of the study asked them to exchange part of the compensation ($50 out of $150) due to them. Table 8 presents

Table 7
Use and Abandonment of Point Smart by Study Participants

Frequency of Use	*n*	%
Every Time	11	50
Most of the Time	5	23
Some of the Time	3	14
Rarely	NA	NA
Very Rarely	1	5
Never	2	9
Total- end of home trial	22	100
Drop outs/Missing	10	31

Table 8
Purchase Intent vs. Purchase Decision by Participants in the Point Smart Study

Situation	Question	Would buy the product		Total
		n	*%*	
Onsite Trial Exit Interview	*Which one would you buy? – Product or its Competitor?*	23	72	32
Mid-Home Trial (2 months)	*Likely to trade part of study compensation to buy product?*	12	48	25
Mid-Home Trial (2 months)	*Would you buy your original device again?*	11	44	25
End of Study (6 months)	Actual Decision to Buy	7	28	25

the number participants who considered the product to be of value or relevance to them; it shows the trend of purchase intent over the course of the study and their actual decisions of whether to purchase at the end.

As shown in Table 8, 32 people started the Point Smart study. Purchase intent dropped from 72% to 48% (nearly half) by midway through the home trial. While 25 people completed the study, only seven of them chose to purchase the product in the end. This represents acknowledgement of real value only by 28% of users that completed the study, or 22% of the entire sample.

Case Three: The Kelvin Interactive Thermostat

Method

Procedures

The Kelvin efficacy study followed essentially the same procedures described under the Lids-Off™ except for contextual adaptations that were made as in the case of the Point Smart.

Sampling

Kelvin study participants were visually impaired and included individuals with low vision (legally blind) and those who were totally blind. Table 9 below shows the sample distribution by functional limitation, as well as by age and gender. Sampling was purposive, with the priority placed on maximizing consumer experience and the variety of functional needs that demanded the use of non-visual sensory interaction for accessibility. The sample size was 48 and included legally and totally blind individuals who reported that they were in charge of operating thermostats in their residence. There were more female participants than male participants. The age range was 25-86 and the median age was 58. The mix of younger and older persons around the median age was fairly even.

Data Collection Design

Data was collected in Phases 2 and 3 following the basic procedure, after identifying indicators and building instruments in Phase 1. The basic quasi-experimental design was followed and modifications were minimal.

Table 9
Kelvin Study Sample Distribution by Functional Limitation, Age, and Gender

Functional Limitation	*n*	%	Age Group	*n*	%	Gender	*n*	%
Low Vision / Legally		58			4			35
Blind	28		25 – 34	2		Male	17	
Totally Blind	20	42	35 – 44	4	8	Female	31	65
			45 – 54	14	29			
			55 – 64	12	25			
			65 – 74	9	19			
			75 – 86	7	15			
Total	48	100	Total	48	100		48	100

Onsite trials. A talking thermostat with functionality and features similar to Kelvin was the marketplace competitor selected for onsite performance comparisons. Targeted to blind users, it was designed to talk to users although it could not receive their voice input as Kelvin did. Onsite trial participants performed five specific tasks using each thermostat's command functions/features: reading room temperature, changing the temperature setting, setting the time, setting the day, and programming the device for weekday and weekend temperatures. Participants were randomly assigned to the product testing sequence. They gave feedback on questionnaires as well as through the exit interview. Observers recorded their performance as well. Video recording was not necessary for the Kelvin study since time was recorded through direct observation using stopwatches. The trial logistics had their own complexities due to each individual's need to 'understand' if not totally 'learn' the programming feature of each thermostat before performing the tasks.

Home trials. The home trial lasted six months. Participants used the Kelvin thermostat either for air conditioning or heating, depending on the time of the year each started the use. Participants performed the needed tasks of their choice using Kelvin, completed weekly feedback questionnaires on their use of the software. They also gave feedback on Day 1 and at two months. Quality and value questions were repeated across questionnaires and interviews. Monitoring and tracking logistics was complex. The Kelvin thermostat needed installation expertise assisting the consumers as home furnaces and circuitry needs varied. Skilled external technical assistance became necessary, introducing delays in individual home trial start dates.

Indicators and instruments. Six persons with disabilities were interviewed at home in Phase One for extraction of indicators. Onsite trials and home trials used consumer questionnaires, observer questionnaires, as well as interviews generated by these indicators and structured after the other two studies. Participants also had the opportunity to purchase the thermostat at the end in exchange for part of the compensation due to them for participating in the study.

Data Analyses

Both descriptive and inferential techniques were used. In the absence of previous benchmarks for 'acceptable levels of impact,' we were again guided by the results of our pilot study on Lids-Off™ in order to judge whether Kelvin was a 'worthy' transfer or not. For our purposes, Lids-Off™ had set the practical standard as to the heights to which a transfer can go in achieving quality and value for the consumer.

Results

As with the preceding cases, this section will address the select key variables common to the three cases. Also, we will focus on descriptive and qualitative data in order to capture context specific information that can explain differences among the three cases.

Sample Size and Attrition

Though participants had interface issues with the use of Kelvin, these did not ensue from participant use of accessories, but rather from device incompatibility with users' furnaces. Of the 48 participants who initiated the study (see Table 9), only 25 completed home trials. This represents a 48% drop out, which is almost half the initial sample, representing the highest attrition of all three studies. Interestingly, 11 of the 48 who finished the clinical trials did not even start the home trials, for various reasons (installation not authorized by residential management, previous knowledge of device, incompatible furnaces such as electric or wood burning, and personal reason). Of the 37 who started the home trials, technical quality and usability reasons lost 12 people within two months. In all, 25 people completed home trial.

User Assessments of Product Quality

The following sections present assessments of Kelvin's quality by the participants.

Compared to Marketplace Competitor (Onsite Trials)

We recall that participants gave comparative evaluations of device-versus- competitor at the exit interview of the onsite trials. Figure 12 shows these results for Kelvin using the eight reference indicators earlier presented with the other two studies.

The columns in Figure 12 show the percentage of participants who judged Kelvin superior to its competitor on the corresponding indicators along the X-axis. Kelvin fared rather poorly on all indicators,

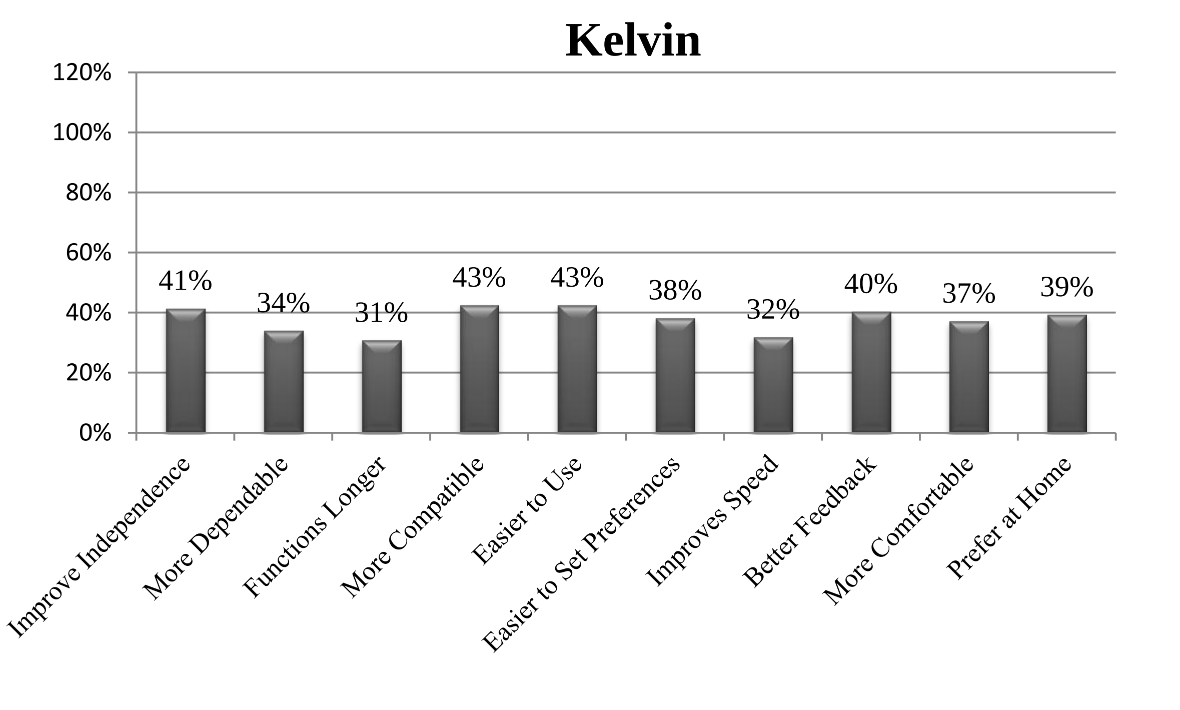

Figure 12. Percentage of participants who preferred Kelvin over its competitor.

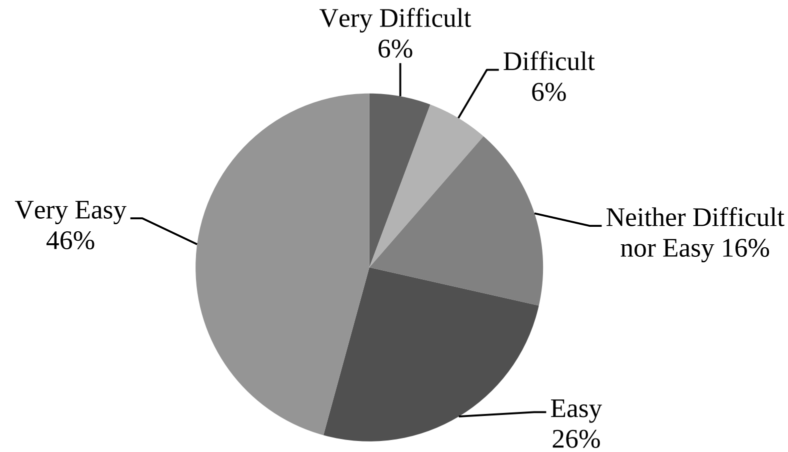

Figure 13. Ease of use of Kelvin's operating instructions: User perception on Day 1.

with less than 50% of the participants (32%-43%) acknowledging it to be superior to the other product on any indicator.

Compared to Critical Competitors at Home

The following results include participant assessments over the home-trial period including Day 1 and after two months of use.

On Day 1. Participants received audio manuals (on cassettes and CDs) as part of Kelvin's installation for trials at home. Considering the importance of Kelvin's programmable feature to blind users living alone, learnability of these manuals was critical to its use, which we measured on Day 1 after the installation. Figure 13 presents user evaluations of Kelvin's intuitiveness and learnability based on the instruction manuals. This simple circle graph shows the percentage distribution of people who rated the manual of operating instructions easy or difficult on a five-point scale. The legend shows the five specific scale points color coded, while the graph shows the corresponding percentage of people who rated Kelvin at levels from 'very easy' to 'very difficult.'

As we can see, Kelvin was considered easy to very easy to learn by 72%. However, as much as 12% of participants considered it 'difficult' or 'very difficult.' Participant comments did not speak highly to its learnability, and pointed to the instructions being difficult to learn from. This is not surprising because Kelvin's target users were persons with blindness that greatly depended on non-visual manuals. Kelvin came with standard print versions, and made CD and audio instructions available only upon request. Large print manuals and Braille versions preferred by some were not an option (later supplied by the study at request).

After two months of home use. Figure 14 shows user perceptions of Kelvin's quality at the end of two months of home trial. For the seven selected indicators, it presents the percentage of positive perceptions by participants, i.e., participant ratings at the higher end of the five-point rating scale (e.g., 4 or 5).

Figure 14 shows moderately positive ratings for Kelvin (59%-77%) in relation to alternative devices they had known and used. In particular, 77% judged it consistent in operation, 70% found it 'enabled task performance' and 73% found it 'functionally

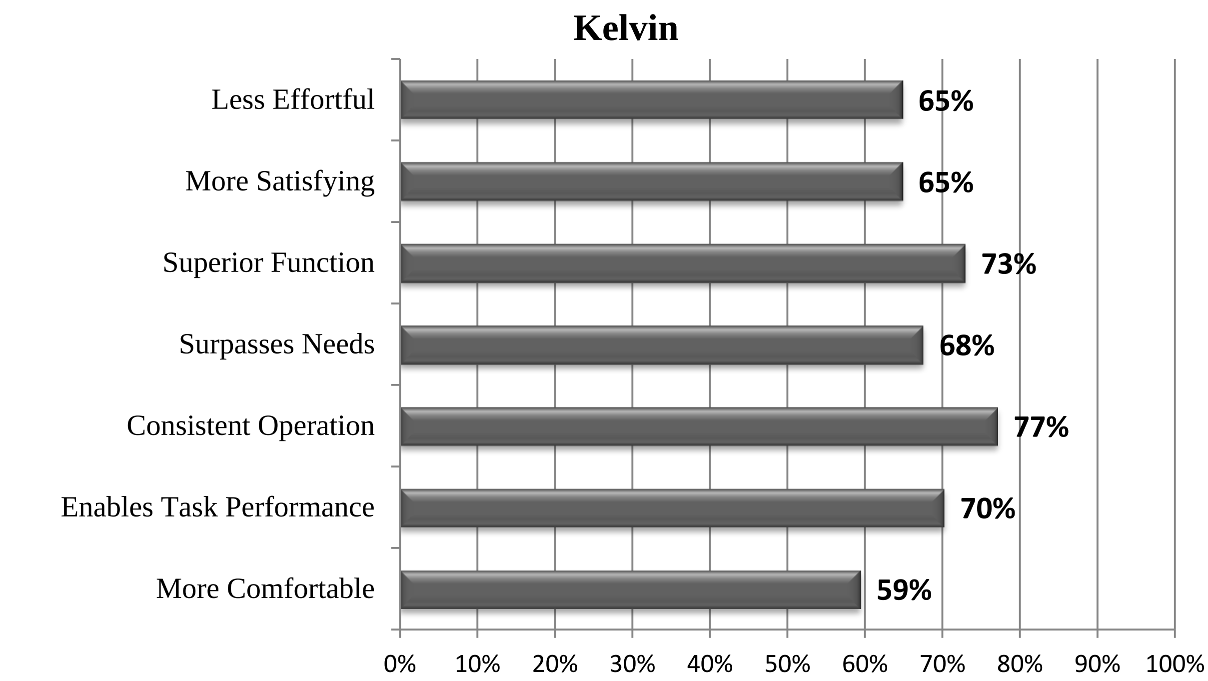

Figure 14. User perception of Kelvin's quality relative to competitors (at two months).

superior.' This is an intriguing finding in light of lower ratings on other counts, considering that these are blind users. One thing that uniquely distinguishes the Kelvin from equivalent devices in the market is its voice input recognition feature.

Over the home trial period. Figure 15 compares user perceptions between the beginning (Day 1) and the end of two months of home trials on three key indicators: 'independent use' (user can operate device without assistance), 'ease of use' of the device, and 'device for me' (device fits user needs). These are before-and-after changes in user perceptions over the seven-week trial period and a measure of Kelvin's impact on users' functional capabilities. Perceptions are presented in the graph as percentages of positive ratings, i.e., scores of 4 or 5 on the five-point rating scale. The X-axis shows the three indicators, with paired columns of percentages of positive ratings for each indicator, one for Day 1 and the other after two months of use.

Figure 15 shows mixed results for Kelvin. Ratings started out reasonably high on all three indicators (70%-80%). On Day 1, 78% recognized its potential to impact their independent functioning, with 9% more joining them at the end. However, the number of people who thought it was easy to use actually decreased during the period (from 80% to 70%). Also, there was no difference regarding it being a good fit (device for me) before and after the period (70% both times). In light of the high percentage (89%) that found it enabled independent use it is not surprising that as many as 70% continued to accept the device as a fit, despite declining perception in its ease of use.

Figure 16 summarizes the weekly trend of participant ratings on the five key indicators common to the three studies. Ratings were moderately high from beginning to end ranging between 4.0 and 4.5. They did drop slightly on all indicators towards the end, while the decreasing trend on person-device fit reversed itself by the end. This

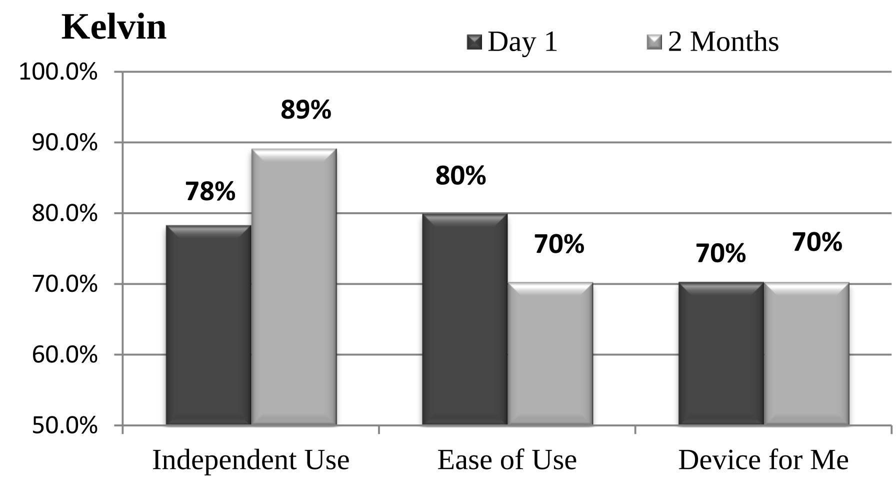

Figure 15. Change of user perception of Kelvin's quality over two months of home trial.

corroborates findings from the previous section.

Product Value to User: Acceptance, Use and Purchase

The following results address the participants' acceptance, use, and purchase of the Kelvin thermostat.

Product Acceptance

The foregoing section reported mixed findings relative to Kelvin's success in terms of user satisfaction. On one hand, as many as 23 people (48%) dropped out of the home trials, but on the other hand, early drop outs (n=11) were for reasons other than dissatisfaction with quality (logistical/installation issues). The question of acceptance and use addresses the remaining

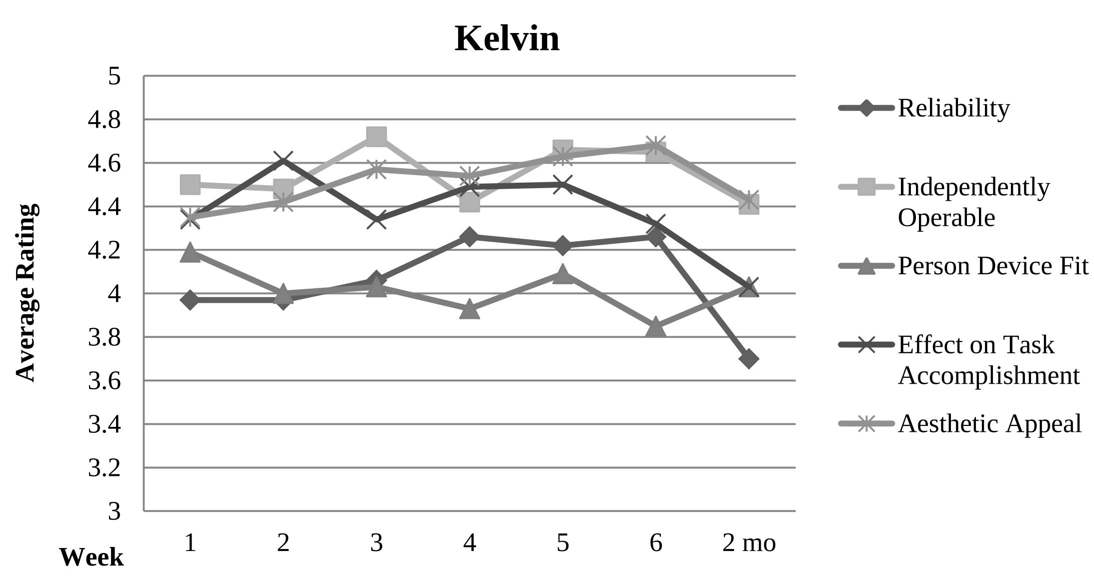

Figure16. Participant perception of Kelvin's Quality over home trial: Weekly trend.

Table 10
Participant Perceptions of Kelvin's Usability: End-of-2-Month Responses

Most liked features	*n*	%
The clear voice commands	16	64
Person's ability to check the temperature and time	12	48
Least liked features		
The talking feature was too sensitive	6	
Some of the buttons were too small or difficult to operate for other reasons	7	

37, including 12 who dropped out for quality reasons, as we shall see later, and 13 who completed the study. To what extent did they consider Kelvin relevant to their needs and acceptable for their own use?

Kelvin did not fare very well on product acceptance. Both technical quality and usability issues surfaced as early as the onsite trials and continued into the home trials: buttons were hard to push and were too small and inaccessible for blind user reading. Ironically, the voice activation feature was both a positive and a negative feature. It would respond to the voice of the user, but it would also annoyingly respond to any or all noise in the environment.

Table 10 captures participant comments about their likes and dislikes of the 25 persons at the end of the home trials.

As noted, acceptance varied among the 25 participants. The device worked for some but not for others. Its key features (voice interaction and temperature setting) satisfied 12-16 persons while frustrating 6-7 others who also had difficulty operating it. Those who reported their preferred features offered comments such as: "It does reflect vocally what the settings are. Like the voice activation;" "Independence it provides. Like the availability of voice, repeats things if one

Table 11
Purchase Intent vs. Purchase Decision by Participants in Kelvin Study

Situation	Question	Total	Would buy the product *n*	Would buy the product %
Onsite Trial Exit Interview	*Which one would you buy? – Product or its Competitor?*	48	18	38
Mid-Home Trial (2 months)	*Likely to trade part of study compensation to buy product?*	37	18	49
Mid-Home Trial (2 months)	*Would you buy your original device again?*	37	9	24
			Bought the product *n*	Bought the product %
End of Study (6 months)	*Actual Decision to Buy*	25	12	48

could not understand;" "Gives you room temp and time;" and "The voice is very good. The program works well."

Comments corresponding to frustrating features included: "When it talks unprompted. The programming buttons are all the same shape, make them different shape;" "It keeps going off when you are talking to someone and it doesn't always do what you tell it to do;" and "Programming; buttons are too small and the sensitivity."

Product Use

During the final four months of the home-trial period (when use was voluntary) it was redundant to ask the use and abandonment question, "How often did you use?" There was no new programming and the 'using' activities were minimal–checking temperature and reporting malfunction if any.

Product Purchase

As with the other two studies, a purchase opportunity was posed to the 25 participants at the very end of the study for purchasing Kelvin giving up $65 (half the market value) from the compensation amount ($150) owed to them. Table 11 presents the trend of participants over the course of the study–in terms of purchase intent and actual purchase decision at the end.

As shown in Table 11, 48 people started the study, 25 completed it. However, only 12 of them chose to purchase the product. This represents only 25% of the total sample. Interestingly, this also represents about half (48%) the people who tried it out to the end, to whom Kelvin seems to have offered 'value.' Interestingly, they were divided in their reasons for acceptance/rejection. Table 12 below summarizes the reasons why participants did (n=12) or did not (n=13) buy Kelvin as per the telephone interview at the end of the study.

Table 12 lists the reasons acknowledged by the 12 participants who bought Kelvin and by the 13 who did not. Their comments

Table 12
Reasons for Buying/Not Buying Kelvin

Is this the reason you bought Kelvin?	*n*	*Is this the reason you did not buy Kelvin?*	*n*
Had confidence in Kelvin's ability to perform accurately.	10	Had no confidence in its ability to perform accurately.	7
Had confidence in the Kelvin's ability to perform consistently.	10	Had no confidence in its ability to perform consistently.	5
Kelvin's voice was easy to understand.	11	Didn't like the voice	1
Found Kelvin programming reliable.	9	Programming too difficult.	5
Frequently used the hands free feature of Kelvin	7	Controls were too difficult to understand.	4
Buttons and controls were easy to locate.	11	Buttons were too hard to push.	1
Maintenance of Kelvin was simple and easy.	8	Pushing the buttons was too painful on fingers.	0
The display screen was easy to read	1	Did not trust its safety in the house	2
Total who bought	12	Total who did not buy	13

corroborate the earlier findings (see Table 10) on acceptance of Kelvin. Beyond the comments, participants volunteered additional reasons beyond what is listed in Table 12. One person reported that a family member did not like the thermostat; some participants claimed they either needed the money or that the price was too much (*n*=4); some reported that the thermostat talked too much, was too sensitive, or that its voice went off at will (e.g., multiple sounds in the house triggered it; *n*=3); some pointed to its inconsistency, saying that it wouldn't maintain temperature settings (e.g., while set at 68 degrees, the temp rose to 70), that it worked inconsistently, that it was totally inaccurate, that its clock kept gaining time, or that it failed to respond to voice commands (*n*=3). Some commented on the buttons, pointing out that they were too small with small print, that they required too much additional AT to read, or that they were too difficult to manipulate because of eyesight (*n*=2).

Summary and Discussion

We summarize and discuss below findings from the three case studies of efficacy assessment presented in the previous section. The methodology used for evaluating quality and value was uniform and systematic for all three products. Results, however, varied with respect to consumer satisfaction and acceptance of the products, as tied to product functionality and features.

Lids-Off™ was liked by an overwhelming number of participants, and it received high ratings on all indicators of technical quality and usability from beginning to end. At the onsite trial, it was clearly rated superior to its marketplace competitor. At home trials, most found it intuitive and learnable. After two months of home use, most gave it high ratings and considered it to be consistent, comfortable, and effortless to operate; most found it both satisfying and said that it surpassed their needs. Seventy percent (70%) considered it enabling. Over 90% embraced the device as a fit for their needs. The product was a success in terms of quality, relative to both market place and critical competitors. In evaluative terms it showed merit. Additionally, it showed worth or value to its consumers. The study had the lowest dropout rate. Users accepted it as a fit for their needs. Most (92%) used it voluntarily during the optional feedback period during home trials. Nearly three-quarters (74%) chose to buy it at the end. Technical quality or usability was rarely mentioned as a factor by those who chose not to purchase it. Money was an issue in isolated cases, but overall the product seems to have been considered cost-effective. Lids-Off™ was a success in that it showed both merit (quality) and worth (value) for this disability population.

The Point Smart software was less successful than Lids-Off™, with mixed results on efficacy. In onsite trials it was preferred to its competitor (Microsoft), although not as overwhelmingly as Lids-Off™. It held great promise and was preferred to its competitor (88%). At home, it was fairly learnable, with just 8% finding it to be difficult. Initially, a good number (70% to 84%) found it consistent in operation, functionally superior, less effortful, more comfortable, and more satisfying than other alternatives. It even ran close to Lids-Off™ regarding 'surpassing needs' of the disability population in question. But it was less an 'enabler' than either Lids-Off™ or the Kelvin, and notably, only 64% embraced it as a fit for their needs. Rating trends declined after four weeks on several usability indicators and on person-device fit. Thus, Point Smart showed dubious merit and its initially positive user perceptions suggest the product's underachieved potential. The product also showed dubious worth or value. Over two months of home use, there was a decline (from 72% to 50%) in participants' willingness to buy the product. During the optional home-use period, more people (14%) abandoned its voluntary use than did participants in the Lids-Off™ study. Interest in the product declined, with only 22% buying the product at the end.

In terms of cost effectiveness, it is difficult to relate the low purchase numbers to the software's affordability because a confounding factor was its vulnerability of duplication from the trial CD version. At any rate, user comments that supported the declining ratings and declining purchase intent suggested that the effectiveness did not outweigh the cost, at least for those for whom the product worked. In conclusion, although Point Smart was considered to be more effective than its competitor at onsite trials, the home trials clearly showed it did not reach the height of its potential in terms of merit and worth. It was not effective enough to be valuable to most participants.

The Kelvin thermostat was also less successful than Lids-Off™ and showed mixed results. Unlike Lids-Off™ and Point Smart, it was not a big success at the onsite trials. Less than one-third (15%-34%) of participants regarded Kelvin as more favorable than its formidable competitor, based on all indicators. At home, it was less learnable due to inaccessible instructions manual, as with Point Smart. Interestingly however, usability ratings shifted upward by the end of two months, with over two-thirds of participants favorably disposed to Kelvin's use. They reported that it surpassed their needs and rated it highly based on usability indicators, with over 80% attesting to its consistency of operation. It was even perceived to be as enabling (70%) as Lids-Off™. However, trends in perceptions from beginning to end were mixed, rising to 89% from 71% on its independent use while falling to 70% from 80% on ease of use. In all, 70% steadily embraced Kelvin from beginning to end as a 'person-device fit.' This compares favorably with results for Point Smart (64%), which suggests that Kelvin did work for more persons in its sample. While Kelvin was 'less effective' than its competing product (onsite trials), it was effective for 70% of those who persisted with it at home. One thing that uniquely distinguishes Kelvin from equivalent devices in the market is its voice input recognition feature.

User purchase behavior was interesting in the case of Kelvin. Only 25% bought it at the end. The drop-out rate was highest for Kelvin due to usability issues and malfunctioning units, but almost half (48%) of the remaining people bought the device, suggesting that it was valued by those for whom it worked. This did not happen in the case of Point Smart, where only 28% of the remaining participants bought it. Both Kelvin and Point Smart were less affordable than Lids-Off™ in terms of absolute dollar value, but more consumers decided to buy Kelvin as compared to Point Smart. This suggests that Kelvin's effectiveness outweighed its cost for more people. It seemed more 'needed' and 'valued.' In conclusion, although Kelvin was not 'more effective' than the chosen competitor, it appealed to a good proportion over the home trial in absolute terms and was valued by about half of participants. Its merit and worth did not reach the heights of the Lids-Off™, but it fared slightly better than Point Smart.

In summary, Lids-Off™ came out successfully both on quality and value counts, whereas Point Smart and Kelvin were less so on both counts. Neither Point Smart nor Kelvin reached their potentials in terms of quality and in terms of acceptance by the user group studied. Point Smart started out well but its perceived quality and value declined in users' eyes over the study period. Kelvin started with unfavorable user perceptions, but it was more appreciated in real-life trials. It was perceived as promising, however only by a limited few who valued it. What can we conclude about their efficacy? What factors explain their apparent lack of success with the participant group as a whole? What does this say about the effectiveness of the T²RERC intervention? What are the lessons to the intervention process?

Similar Method, Unique Contexts

At this point in evaluating the three devices, context becomes important. Despite that similar methods were used, contextual differences among the three cases make it difficult to generalize across them. First, product uniqueness and individual corporate realities affected the degree of the T²RERC's intervention and the company's use of the intervention. Second, logistics affected the implementation of the efficacy study itself although same methods guided them. These points are considered below.

Differences in Design Challenges

Each of the three products was unique in design because of the functional needs of the different populations they targeted. While the T²RERC intervened for an 'inclusive' redesign of each of the three prototypes, the three products initially targeted different markets. Lids-Off™ is a home appliance targeted to mainstream buyers, while Point Smart and Kelvin more directly targeted persons with disabilities. As AT products, the last two had more challenging accessibility issues with which to contend. These stemmed from complexity involved in operating them and dependency on hardware and system interfaces. They did not reach the same height, either on quality or on value, as Lids-Off™, whose clear championship in this respect and successful sales volume lend support to an effective intervention by the T²RERC in its development.

Recognizing that AT outcomes are functions of person-device compatibility, it may be argued that a subject-by-subject analysis of the findings is a more valid way of inferring products' benefits to users, rather than evaluating products based on analysis of group data as we did. Such analyses might shed a different light on these results, and we are currently analyzing for differential effectiveness based on functional needs. However, user comments suggest that technical issues and software operability were more of a problem than device incompatibility. Besides, our onsite trials design in this study did permit direct observation of individual performance, and home trials permitted individual tracking of each consumer's experience with the product use. The general frustration reported by participants, our informal observation of the context of product use, as well as the history of the product development reveal that there is more to the difference in impact than appears on the surface.

Differences in the T²RERC Intervention

None of the three prototypes originally targeted the disability market exclusively before the T²RERC intervened. However, both the Lids-Off™ and the Kelvin got the benefit of the full systematic evaluation input from the T²RERC, from the design stage through successive prototype evaluations. Meanwhile the Point Smart case was an exception to our typical intervention. As described earlier, support to Info Grip came at a much later stage of development. Opportunities for timely capture of input for its design were missed. Practical constraints further hastened the product to market before all technical refinements were fully in place. Support was thus not ideal for Point Smart. The repercussions of this difference in evaluative input showed its effect on the levels of quality level and acceptance of Point Smart. Users recognized promise, were impressed with its usability, but were frustrated at the barriers to its full use.

Installation issues, instructional manual quality, and hardware and software compatibility issues made technical assistance crucial for Point Smart. While Info Grip has been very receptive to feedback from the efficacy study and is bringing out its next version of Point Smart, Black and Decker has brought a line of products into the market and requested our continued support.

Differences in Information Use by Developer

Although very similar support was provided both to Kelvin and Lids-Off™ during development, there was a difference in how the two companies used our evaluation information. Whereas Lids-Off™ took all key recommendations, the post-commercialized Kelvin did not incorporate some key features identified in the focus groups, including contrasting or light-up buttons, backlit displays, enlarged lettering on digital displays and switches, and a carbon monoxide detector among others. This was a difference in the use of the evaluative information provided to the two companies. Also, as pointed out earlier in the background section, the production of Kelvin was outsourced and there were quality control issues in the production processes. Kelvin needed technical support during home trials due to malfunctioning units that resulted from production flaws. Such differences were important factors in the final outcome of how each product impacted user perception of quality, and consequently its acceptance.

Differences in Study Implementation

Iterating case studies represent 'real-world' formative evaluations. They can be very valuable for developing best practices in research methods by illuminating how methods need tailoring to contexts. Device and user individualities dictated variations in test protocols in the case of the three efficacy studies. As mentioned earlier, a clinician expert had to work with the Point Smart study participants at the onsite trials, pretesting and configuring the device with each mouse type, and guiding home trial set-ups. Pre-screening tests on computers were needed in recruiting participants for the Point Smart in order to identify true functional limitation. Additionally, unforeseen complications with product operation had logistical implications for the Point Smart and the Kelvin home trials, thus requiring frequent technical support by the respective companies. In contrast to these two studies, the Lids-Off™ was an almost seamless study.

Conclusions and Lessons

In light of the foregoing, conclusions are more straightforward about the relative efficacy of the three products than they are about the effectiveness of the T^{2}RERC's intervention. It is easy to see that Lids-Off™ was a success in terms of its benefit to its end users whereas Point Smart and Kelvin were only partially beneficial. As for the T^{2}RERC's transfer process, the Lids-Off™ case lends evidence to its effectiveness, and one could argue that it would have been just as effective in the other two cases, had those contextual difficulties been surmountable to the point of being 'best-case scenarios.' One could also argue however, that realities are more often far from being best-case scenarios, and there is need to further improve the T^{2}RERC process so it responds to such challenging realities. Indeed, the contrasting cases in this study hold lessons that might lead to improving the transfer process and shedding light on future technology transfer.

Developing Products for Optimal User Benefits: Lessons and Implications

The study of efficacy of its transferred products was a response of the T^{2}RERC to the issue raised in its 2003 conference about advancing the state of the science and practice of technology transfer through continued study of its model. While previous evidence on successful transfers attested to the merit of the model, the product efficacy studies sought evidence of the model's worth in terms of benefits from its outcomes to end users. In discussion here is the extent to which the studies provided such evidence and in what ways the experience was an enlightened step toward advancing theory and practice of technology transfer.

Technology transfer has long been present in business and industry practice as part of New Product Development (NPD) through Stage-Gate and similar models (Kahn et al., 2005). In academic circles, interest in technology transfer stems from a desire, at least in theory, to link research to NPD through university technology transfer offices that act as bridges to the marketplace. Policy makers have increasing expectations in terms of linkages to new product development from the research projects they approve for funding. Linked to return on investment, there is a growing recognition of the need for knowledge translation (Canadian Institutes of Health Research, 2004; Sudsawad, 2007) resulting in an awareness for the need for transdisciplinary or Mode 2 research (Gibbons et al., 1994; Nowotny, Scott, & Gibbons, 2001) as well as attempts at its operationalization (MacLean, MacIntosh, & Grant, 2002; Savory, 2006). In this context, academic- industry partnerships have been recognized as important for advancement of theory and practice in technology transfer, and paradigms have been attempted (Arvantis, Kubli, & Woerter, 2008; Renault, Cope, Dix, & Hersey, 2008; Sharif & Baark, 2008; Vaajakallio, Vehmas, Keinonen, & Mattelmaki, 2008). Current thinking seems to point to the wisdom of academic and industry collaborations involving joint research and development work.

In light of the above, this article deliberately uses an academic framework (the CIPP model) to integrate and interpret the T^2RERC experience with product efficacy assessment. In effect, it layers an academic perspective over the business model (the Stage-Gate model for NPD) that guided the T^2RERC in its product development support. This should allow for imperfections in both models–one theoretical and the other practical–to surface as repercussions from the case studies, with lessons for the academic and industry partners who try to deliver new products of quality and value.

Our experience through the challenges from these contextual differences led to three important lessons. They go beyond the T^2RERC and the partnering companies to include academic researchers or knowledge brokers and their corporate partners, and they clarify questions about the realities of collaborative models.

Lesson 1: Consumer Input

Consumer input is fundamental to ensuring the quality and value of a product in development. The timing of the input is key–it should be captured prior to (re)design, during prototype improvement, and at the end of the development process. All three product developers recognized the value of the consumer input in shaping their product after our feedback from the efficacy study, if not earlier.

Lesson 2: Product Quality

A business partner's (or company's) commitment to product quality is as important for success as the academic research partner's (T^2RERC's in this case). Both Kelvin and Lids-Off™ received standard evaluative support from us, but the product developers used the information differently. Kelvin's diminished value for the consumer can be explained by its omission of important features as well as by its choice to outsource operations, thereby investing less on quality assurance and production control.

Lesson 3: Customer Support

It is difficult to achieve product value without adequate post-commercialization support to the consumer in the use of the product. Involving the consumer in development may yield the desired product; commitment to quality by both partners may enhance its appeal and value; but unless a manufacturer or vendor renders the product viable for use, consumers will be unable to certify and accept it as right for their needs. As mentioned earlier, both Point Smart and Kelvin were complex to install and operate. Consumer learning and appreciation of these products depended heavily on the availability and effectiveness of accessible versions of instructional manuals. This is a lesson to both partners–the (academic partner/broker), T^2RERC, should address this during development of the new product; and the company should build this support into its marketing plans.

In summary, the differing case contexts partly explain differential findings in the efficacy of the three products. Lids-Off™ encountered the optimal conditions for achieving desired product quality and value levels, i.e., the T^2RERC's systematic and timely evaluative support and Black and Decker's incorporation of the recommended functions and features into the product. Kelvin, which did poorly on quality and value, was a case of complete and timely input by the T^2RERC but limited corporate commitment to quality and product support. Point Smart was the least valued by its users in spite of its perceived potential, and it was also the case with the least optimal conditions under which to achieve its potential. The case study suggests that while the T^2RERC successfully brought a new product of quality and value to the market place, the corporate partner had an equally significant role in achieving this outcome. In this sense, the intervention into the prototype is in fact a joint effort between the academic-researcher/knowledge-broker (the T^2RERC) and the business partner. Effectiveness and impact cannot be achieved without equal commitment.

Implications for Practice

Academic and corporate partners are each stakeholders in a collaborative product development process, and the above lessons hold implications for them both. First, *involving consumers during (rather than after) the development process is important.* Corporate requests for support need to be timely. On the other hand, academic support teams should work within the company's product development schedules and deadlines.

Second, *evaluation information is only as good as follow-up decisions to improve product.* It should be recognized that the academic role is to enlighten through evaluation, but improvement decisions are a direct corporate concern. Also, commitment to improving quality includes minimizing production flaws through *maintaining control over operations flow.* Practical constraints can make a huge difference in the final design of the product, and smaller companies face a bigger disadvantage than larger companies in commercializing their products with the quality that the product deserves. The Kelvin thermostat might well have suffered the consequences of outsourcing by bidding. Included among the user dissatisfaction comments is the poor quality material that diminishes accessibility to the touch-dependent blind user.

Third, accessibility is key to an AT product's usability, and the importance of post-commercialization product support cannot be minimized. Those responsible for the technology transfer intervention (the T^2RERC

/academic broker) should make sure that product manuals are part of their evaluations, so that the product is learnable and can be independently put into operation by consumers. Accessible manuals and technical support to user cannot be overlooked as something obvious that production will take care of. As examples, Point Smart's instructions were web-based, and the user needed Point Smart to access them. Similarly, large print and Braille version options would make the Kelvin thermostat more accessible to blind users.

In general, the efficacy studies suggest that academic-corporate collaborations have great potential for developing products of quality and value, provided there is appropriate use of evaluation as a tool for achieving this. Evaluations are important, not only for what we learn from them, but also for what we learn about them. In this article, we have used the CIPP systems approach as the framework for analyzing how adequately evaluation was utilized in the development and commercialization of the three devices studied. In theory, this approach should maximize evaluation's potential for achieving optimal benefits to product users. The cases illustrate the value of each step in this approach. All the same, through these cases, we have also come to realize the challenges of translating theory into practice. Challenges to this task posed by business-world realities are often greater than the academic world realizes. While there is awareness of the need to make mainstream products more inclusive, it is yet to be recognized that this has implications for effort both by industry and academia. Each of these two sectors has developed its own specialized knowledge and expertise, but unfortunately each has done so mostly in isolation from the other. It is time that the two worked hand-in-hand to develop working frameworks, offering models that do exactly what models should – represent reality. Perhaps this is the best lesson that we have learned from the efficacy studies.

It is important to note that such efficacy studies are realistically only performed once all of the prior research and development outputs are achieved and all transfer and commercialization are accomplished. One cannot know how a new product will meet the needs of intended customers until they use the product and compare their experience with it to other products/methods for accomplishing the same tasks. The study shows that one can optimize the effort to meet customer needs by integrating the relevant design and consumer criteria from the earliest stages of development. Beginning with that end in mind is the best means to ensure that new products do indeed contribute to the quality of life for persons with disabilities.

Acknowledgement

This study was supported by funding from the National Institute on Disability and Rehabilitation Research (NIDRR) of the U.S. Department of Education (USDE) under grant number H133E030025. The opinions contained in this presentation are those of the authors and do not necessarily reflect those of the U.S. Department of Education.

The authors acknowledge the support and collaboration of past and present colleagues from the University at Buffalo's Center for Assistive Technology, who have lent their expertise and contributed in different ways at various stages of the efficacy study series– among them Katie Beaver, Kate Wagner, Sumana Silverheels, Christine Oddo, and Asha Subramaniam. Whether they conducted the onsite trials and observed consumers, tracked them during the home trials, or helped with data entry and analyses, their services and efforts were invaluable to this four-year study.

References

Action Talking Products, LLC. (2008). *KELVIN Talking Thermostat*. Retrieved May 29, 2008, from http://www.actiontalkingproducts.com/

Arthanat, S., Stone, V. I., & Usiak, D. J. (in press). The aftermarket payoff from consumer involvement in new product design. *Technology and Disability*.

Arvanitis, S., Kubli, U., & Woerter, M. (2008). University-industry knowledge and technology transfer in Switzerland: What university scientists think about co-operation with private enterprises. *Research Policy, 37*, 1865-1883.

Babbar, S., Behara, R., & White, E. (2002). Mapping product usability. *International Journal of Operations and Product Management, 22*, 1071-1089.

Batavia, A. I., & Hammer, G. S. (1990). Toward the development of consumer-based criteria for the evaluation of assistive devices. *Journal of Rehabilitation Research and Development, 27*, 425-436.

Bauer, S. M. (2003). Demand pull technology transfer applied to the field of assistive technology. *Journal of Technology Transfer, 28,* 285-303.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Washington, DC: American Educational Research Association.

Canadian Institutes of Health Research. (2004). *Knowledge translation strategy 2004-2009: Innovation in action*. Retrieved January 6, 2009, from http://www.cihr-irsc.gc.ca/e/26574.html

Center for Universal Design. (2007). *About universal design*. Retrieved June 22, 2007, from http://design.ncsu.edu/cud/

Design for All Foundation. (n.d.). *Design for all: What is?* Retrieved June 22, 2007, from http://www.designforall.org/en/downloads/dossier-DfA-Fd-ang.pdf

Dumas, J. S., & Redish, J. C. (1994). *A practical guide to usability testing*. Norwood, NJ: Ablex.

Dzida, W. (1995). Standards for user-interfaces. *Computer Standards & Interfaces, 17*(1), 89-97.

Fuhrer, M. J. (2002). *Observer's report on research for the 2002 RERC summative program review* (Final Draft). Washington, DC: U.S. Department of Education, National Institute on Disability and Rehabilitation Research.

Gibbons, M., Limoges, C., Nowotny, H., Trow, M., Schwartzman, S., & Scott, P. (1994). *The new production of knowledge: the dynamics of science and research in contemporary societies*. Thousand Oaks, CA: Sage.

Green, W. S., & Jordan, P. W. (1999). *Human factors in product design*. London: Taylor and Francis.

Han, S. H., Yun, M. H., Kwahk, J., & Hong, S. W. (2001). Usability of consumer electronic products. *International Journal of Industrial Ergonomics, 28*, 143-151.

Infogrip, Inc. (2003). *PointSmart*. Retrieved May 29, 2008, from http://www.infogrip.com/product_view.asp?RecordNumber=988

International Standards Organization. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) – Part II guidance on usability* (ISO CD 9241-11). Geneva: International Organization for Standardization.

Jennings, L. (2006, August 8). Mason professor dubs products for the elderly "Nana" technology. *The Mason Gazette*. Retrieved June 22, 2006, from http://gazette.gmu.edu/articles/8787/

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards* (2nd ed.), Thousand Oaks, CA: Sage.

Jordan, P. W. (1998). *An introduction to usability*. London: Taylor and Francis.

Kahn, K. B., Castellion, G., & Griffin, A. (Eds.) (2005). *The PDMA handbook of new product development* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Koester, H. H., (2004-05). *Compass Assessment*

Software (Version 1.0.0) [Computer software]. Ann Arbor, MI: Koester Performance Research.

Lane J. P., Leahy, J. A., Bauer, S. M. & Stone, V.I. (2008, June). *From mind to market: differentiating research, development and production processes to achieve outcomes*. Workshop presented at the Rehabilitation Engineering and Assistive Technology Society of North America (RESNA) Annual Conference, Arlington, VA.

Lane J. P., Usiak D. J., Stone V. I., & Scherer M. J. (1997). The voice of the customer: Consumers define the ideal battery charger. *Assistive Technology, 9*, 130-139.

Lane, J. P. (1999). Understanding technology transfer. *Assistive Technology, 11*(1), 5-19.

Lane, J. P. (2003). The state of the science in technology transfer: Implications for the field of assistive technology. *Journal of Technology Transfer, 28*, 333-354.

Leahy J. A. (2003). Paths to market for supply push technology transfer. *Journal of Technology Transfer, 28,* 305-317.

Leahy J. A. (2005, June). *Participatory development: Importance of primary market research in identifying market trends*. Paper presented at the Rehabilitation Engineering and Assistive Technology Society of North America's (RESNA) Conference, Atlanta, GA.

MacLean, D., MacIntosh, R., & Grant, S. (2002). Mode 2 management research. British *Journal of Management, 13*, 189-207.

Malhotra, N. K. (1999), *Marketing research: An applied orientation* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Microsoft. (2009a). Dictionary. *Efficacy*. Retrieved March 29, 2010, from http://encarta.msn.com/encnet/features/dictionary/DictionaryResults.aspx?lextype=3&search=efficacy

Microsoft. (2009b). Thesaurus. *Efficacy*. Retrieved March 29, 2010, from http://encarta.msn.com/encnet/features/dictionary/DictionaryResults.aspx?lextype=2&search=efficacy

Nowotny, H., Scott, P., & Gibbons, M. (2001). *Rethinking science: knowledge and the public in an age of uncertainty*. London: Polity Press.

Pirkl, J. J. (1991). Transgenerational design: A design strategy whose time has arrived. *Design Management Journal, 2*(4), 55-60.

Popovic, V. (1999). Product evaluation methods and their importance in designing interactive artifacts. In W. S. Green & P. W. Jordan (Eds.), *Human factors in product design* (pp. 26-35). London: Taylor and Francis.

Renault, C. S., Cope, J., Dix, M., & Hersey, K. (2008). A new technology transfer paradigm: How state universities can collaborate with industry in the USA. *Industry and Higher Education, 22,* 99-104.

Rich, C., Sidner, C., Lesh, N., Garland, A., Booth, S., & Chimani, M. (2006). DiamondHelp: A new interactive design for networked home appliances. *Personal and Ubiquitous Computing, 10*(2-3). Retrieved on January 16, 2009, from http://www.springerlink.com/content/j8mv467586n32611/

Rouse, W. B. (1991). *Design for success*. New York: John Wiley and Sons.

Savory, C. (2006). Translating knowledge to build technological competence. *Management Decision, 44*, 1052-1075.

Scriven M. (1973). The methodology of evaluation. In B. R. Worthen & J. R. Sanders (Eds.), *Educational evaluation: Theory and practice* (pp. 60-104). Belmont, CA: Wadsworth.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.) Newbury Park, CA: Sage.

Sharif, N., & Baark, E. (2008). Mobilizing technology transfer from university to industry: The experience of Hong Kong universities. *Journal of Technology Management in China. 3*(1), 47-65.

Stone, V.I., Bauer, S.M., Lane, J.P., Usiak, D.J., Khan, Z., & Prabhu, C. (1998). Wheelchair tiedowns: Ideal features and existing products. *Technology and Disability, 9*(3), 159-178.

Stone, V. I. (2003). Systematic technology transfer: A case study in assistive technology. *Journal of Technology Transfer, 28*, 319-332.

Stone V. I. (2005, October). *Assessing efficacy of assistive technology transfers: A validation of the T²RERC's technology transfer model.* Paper presented at the Annual Meeting of the American Evaluation Association, Toronto, Ontario, Canada.

Stone V. I., Lockett, M., & Usiak, D. J. (Eds.) (2009). *A resource guide to evaluation in the context of new product development.* Unpublished document available online at http://t2rerc.buffalo.edu/

Story, M., Mueller, J., & Montoya-Weiss, M. (2002). *Evaluating the universal design performance of products*. Retrieved March 29, 2010, from http://www.design.ncsu.edu/cud/pubs_p/pudperformproduct.htm.

Stufflebeam, D. L. (1973). Excerpts from "Evaluation as enlightenment for decision making." In B. R. Worthen & J. R. Sanders (Eds.), *Educational evaluation: Theory and practice* (pp.128-150) Belmont, CA: Wadsworth.

Stufflebeam, D. L. (2001). Evaluation models. *New directions for evaluation, 89*. San Francisco: Jossey-Bass.

Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco: Jossey-Bass.

Sudsawad, P. (2007). *Knowledge translation: Introduction to models, strategies and measures.* Austin, TX: Southwest Educational Development Laboratory, National Center for the Dissemination of Disability Research.

Trace Center. (2003). *Designing a more usable world–For all.* Madison, WI: University of Wisconsin-Madison. Retrieved June 2, 2003, from http://trace.wisc.edu/world/

Vaajakallio K., Vehmas K., Keinonen T., & Mattelmaki T. (2008, September). *Design research as industry-university collaboration.* Paper presented to the EuroMot 2008 Conference, Nice, France.

Worthen B. R., & Sanders, J. R. (1973). *Educational evaluation: Theory and practice.* Belmont, CA: Wadsworth.

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practical guidelines* (2nd ed.). New York: Longman.