Q-Methodology as a Research and Design Tool for HCI

Katie O'Leary
The Information School | DUB
University of Washington
Seattle, WA 98195 USA
kathlo@uw.edu

Jacob O. Wobbrock
The Information School | DUB
University of Washington
Seattle, WA 98195 USA
wobbrock@uw.edu

Eve A. Riskin
Electrical Engineering
University of Washington
Seattle, WA 98195 USA
riskin@uw.edu

ABSTRACT

A "discount" version of Q-methodology for HCI, called "HCI-Q," can be used in iterative design cycles to explore, from the point of view of users and other stakeholders, what makes technologies *personally significant*. Initially, designers critically reflect on their own assumptions about how a design may affect social and individual behavior. Then, designers use these assumptions as stimuli to elicit other people's points of view. This process of critical self-reflection and evaluation helps the designer to assess the fit between a design and its intended social context of use. To demonstrate the utility of HCI-Q for research and design, we use HCI-Q to explore stakeholders' responses to a prototype Alternative and Augmentative Communication (AAC) application called *Vid2Speech*. We show that our adaptation of Q-methodology is useful for revealing the structure of consensus and conflict among stakeholder perspectives, helping to situate design within the context of relevant value tensions and norms.

Author Keywords

Qualitative methods; quantitative methods; user studies; design methodology; psychology; personal significance.

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User interfaces—*evaluation/methodology, theory and methods*.

INTRODUCTION

The usefulness of technologies is often characterized in terms of their function, productivity, effectiveness, and efficiency. But usefulness and other attributes of technologies (such as aesthetics) can also be characterized and experienced in terms of *personal significance*. By "personal significance," we mean the ways that technologies affect people's experiences of human identity and human relationship: social connection, intimacy, competence, agency, community, belonging, solidarity, and so on. And yet, we find a lack of empirical methods

specifically suited to understanding the personal significance of technologies for the purposes of design. Phenomenology in HCI [25] has served as a philosophical approach to understanding people's personal experiences of technologies, but it can be difficult to operationalize due to the inherently elusive nature of "experience." Many traditional HCI design and evaluation methods, such as usability testing [18], cooperative design [9], and contextual inquiry [11], *can* be used for investigating users' notions of personal significance, but these methods were centrally developed with workplaces in mind and do not as readily yield insights into human identity and human relationship. Interviews and ethnographies certainly investigate what is personally significant to users, but these methods produce volumes of unstructured or semi-structured data that can be difficult to translate into concrete design requirements [20].

To address these limitations when attempting to understand and design for personal significance, we have discovered and tailored *Q-methodology* [21] for use in HCI. Q-methodology, originating in psychology in the 1930s, is promising for capturing the personal significance of technologies in a systematic and analyzable way that can speak directly to design requirements. Furthermore, Q-methodology is a small-sample technique that can lend statistical validity to the qualitative interpretation of subjective data, making it useful for the small samples typical of HCI studies (<50 people).

We demonstrate an approach to using Q-methodology that maximizes its potential for informing the design of personally significant technologies, technologies for which mere productivity is not the primary concern. In our approach, called "HCI-Q," designers represent their designs as a set of statements that stakeholders sort according to personal significance. Rather than being facts about a design's utilitarian functions, or statements about a design's subjective value gathered from intensive and often time-consuming interviews (which is the approach in traditional Q-methodology), HCI Q-set statements are the designer's *a priori* hypotheses about how a design might affect individual and social behavior. Stakeholders reveal their attitudes towards the personal and social implications of a design by reacting to the design-as-statements. Such revelations can help designers to situate design ideas within the consensus and conflicting views among stakeholders

early in the design process, or to evaluate an existing system.

To demonstrate the value of HCI-Q, we use it to explore stakeholders' responses to a prototype Alternative and Augmentative Communication (AAC) application, *Vid2Speech* [19]. Our findings indicate that HCI-Q has several advantages for research and design. HCI-Q facilitates critical reflection on design, making explicit the designer's *a priori* assumptions. HCI-Q engages stakeholders, quantifies their perspectives on a design, and leverages statistical methods to reveal the structure of consensus and conflict among those perspectives. By engaging stakeholders in the evaluation of designs and providing tools for quantifying the data, HCI-Q helps designers place constraints on design according to statistically valid and qualitatively rich perspectives of personal significance. Both critical reflection and evaluation keep the focus of design on the experiences of the user. These advantages of HCI-Q will become clear in our account of using HCI-Q with Vid2Speech, below.

The contributions of this work are: (1) the exposition of Q-methodology and its application to HCI problems; (2) our modifications to the Q-methodology process for the purposes of HCI research and design, called HCI-Q; and, (3) empirical results and reflections from our application of HCI-Q to a novel AAC application, Vid2Speech.

RELATED WORK
There are many approaches to gathering design-relevant information from the users' perspectives and for linking data collection to design-specification. Approaches such as Participatory Design (PD) [11], and Value Sensitive Design (VSD) [7], and unstructured methods such as interviews, ethnographies, and Cultural Probes [8], are useful for gathering evidence of the significance to users of different personal and social implications of a design. Embracing the users' subjective perspectives, producing mutual understanding between designers and stakeholders, and minimizing the researcher's authority are important features that HCI-Q shares in common with these other methods.

However, in contrast to unstructured qualitative methods, HCI-Q is well-suited to quantifying subjective data, and for providing statistical support for design requirements. Much like Repertory Grid Technique (RGT) [10], HCI-Q is highly structured and efficient, yet highly sensitive to human values, attitudes, and beliefs. Unlike RGT, which requires parallel design of multiple prototypes, HCI-Q can be used to focus on a single design artifact or design *idea*. It can also be used to complement other methods. For example, HCI-Q can be used within the VSD approach to: (1) explicitly engage designers in critical reflection on the underlying properties of their designs that support or hinder human values; and (2) generate an instrument for empirical investigations of stakeholder responses to technologies. It can also be used early on within a PD approach to get a

snapshot of the structure, substance, and strength of conflict and consensus among participant-designers.

In addition to methods and approaches that explicitly engage stakeholders' perspectives, there are various methods for engaging designers in critical reflection on the functions and futures of designs, such as scenario-based design [4], value scenarios [17], future workshops [12], alternative nows [16], and design noir [5]. HCI-Q shares many of the advantages of scenario-based methods that Carroll [4] described: it "evokes reflection," "coordinates design action," "affords multiple views of an interaction," and "promotes communication among stakeholders." The difference between scenario-based methods and HCI-Q is that HCI-Q does not focus generally on scenarios of uses, tasks, or possible futures. Rather, it focuses specifically on the *personal and social implications* (both desired and feared) of a design. Instead of detailed narratives, HCI-Q focuses on short subjective statements of opinion about how a design will affect personal and social behavior. In contrast to scenario-based design, HCI-Q ultimately gives authoritative control to stakeholders by using the short descriptions as stimuli for engaging stakeholders in decisions about what social and personal properties make designs meaningful and desirable.

Q-methodology has had application in the social, behavioral, and health sciences [3,6]. It has already been used without modification for the purposes of understanding people's attitudes towards technologies (*e.g.*, [1,2,15]) but never explicitly for the purposes of *design*—that is, for generating design ideas and for providing rigorous backing for design decisions. Two efforts that use Q-methodology to investigate social computing have hinted at its potential for informing design and evaluation [1,2], but they have not streamlined their data gathering instruments or data gathering procedures for that purpose, as HCI-Q has. These traditional approaches to using Q-methodology are less practical than HCI-Q because they can be time consuming to implement and lack a direct and deliberate connection between the results of data analysis and constraints for design.

Our optimization of Q-methodology for HCI makes it (1) more effective for exploring the personal significance of technologies, (2) faster to implement, and (3) directly relevant to providing constraints for design.

A BRIEF OVERVIEW OF Q-METHODOLOGY
Although a comprehensive treatment of Q-methodology is beyond the current scope, readers can find traditional accounts elsewhere [14,21,23,24], and especially Brown's *Political Subjectivity* [3], which has informed much of our overview. Here, we cover the basics with an eye towards enabling the HCI researcher to make informed use of Q. Then we describe HCI-Q in detail.

Apparatus and Procedures
Q-methodology uses a set of stimuli, called a "Q-set," to elicit people's points of view. The stimuli usually consist of

a set of statements, gathered from interviews and relevant literature, about a social, psychological, political or other phenomenon of interest [3]. Participants sort the statements according to what is personally significant to them, and rank the statements by arranging them in a table from Most Disagree (-4) to Most Agree (+4), with statements of relative personal *in*significance placed in the middle (0) (Figure 1). This sorting procedure is called Q-sorting, and results in a "Q-sort" for each participant. The table that captures the Q-sort is called a "forced distribution" because participants are forced to distribute the statements in groups in a normal distribution around a zero-point, with a standard deviation (Figure 1). Brown [3] suggests that for studies that involve people who are expected to be relatively uninformed about the subject matter most participants will not have strong opinions about most statements. Therefore, more room should be provided in the middle around the zero-point of the distribution. But participants who are experts are likely to agree or disagree with most statements, and so a flatter distribution is more appropriate. Ultimately, using a forced distribution requires participants to weigh their own sentiments carefully, as they cannot simply render a binary decision and stack everything up at the endpoints.
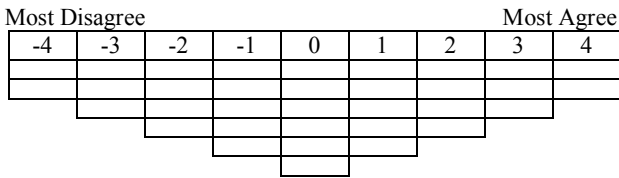
Most Disagree								Most Agree
-4	-3	-2	-1	0	1	2	3	4

Figure 1. Forced distribution for the Q-sorting procedure.

There are at least three additional reasons for using a forced distribution rather than a free-sorting procedure. The first reason is pragmatic: with a relatively modest set of 33 statements and a distribution range of -4 to +4, there are nearly 11,000 times as many ways to arrange the statements as there are people in the world [3]; therefore, rather than being limiting, the forced distribution provides some structure to a potentially overwhelming procedure. The second reason is statistical: the forced distribution produces Q-sorts with equivalent means, yielding the same normalized distribution of data points in each Q-sort, which aids in the statistical comparison of Q-sorts during data analysis. The third reason is phenomenological: it models the tendency of characteristics that are *least* like us (or statements that we most disagree with) to be just as psychologically significant as characteristics that are *most* like us [3]. Therefore, statements placed at the extremes of the Q-sort distribution are most salient for a person and are duly given the most statistical weight during analysis.

Statistical procedures for data analysis

Statistical analysis of Q-sorts, called factor analysis, is automatically computed by freely available online software packages. Our brief description here demonstrates the purpose and value of factor analysis in Q-methodology.

First, the strength of relationship between each pair of Q-sorts is calculated using Pearson's product-moment correlation. This first analysis results in a large correlation matrix, for example, in a study with 20 participants, 190 coefficients will be calculated [23]. Factor analysis reduces this data by clustering Q-sorts that are statistically similar. This "factor extraction" is achieved by a common method such as principal component analysis or a centroid method [13]. "Factor rotation" is then performed *objectively* to maximize the correlation of each Q-sort with a single factor, or *subjectively* to maximize the importance of a particularly significant perspective (*e.g.*, to make one Q-sort define a single factor). Factors that emerge reveal clusters of highly correlated Q-sorts. Each factor represents a unique *point of view* among all the other factors. This procedure will be illustrated for our study of Vid2Speech near the end of this paper.

A factor represents a cluster of Q-sorts that have attributed similar values (-4 to +4) to each statement in the Q-set. Each factor is defined by an averaged sum of its Q-sorts. Some Q-sorts contribute more to defining a factor than others; they have a higher "loading" on that factor. Factor loadings of Q-sorts are used to determine factor scores for each statement in the Q-set. Factor scores are normalized z-scores that indicate, for each factor, the position relative to the mean assigned to each statement (*i.e.*, relative to the zero point representing the "neutral" position in the Q-sorting procedure). Factor scores can be mapped to the range of real numbers expressed in the Q-sort distribution (*i.e.,*-4,-3,…,+3,+4) to create a representation of a "model" Q-sort for that factor, called a factor array (Figure 2).

Model Q-Sort for Factor A

-4	-3	-2	-1	0	1	2	3	4
24	22	23	16	7	2	4	5	1
27	30	25	20	14	3	9	10	19
	33	29	21	17	6	11	18	
		36	26	31	8	15		
			28	34	12			
			32	35	13			
				37				
				38				

Model Q-Sort for Factor B

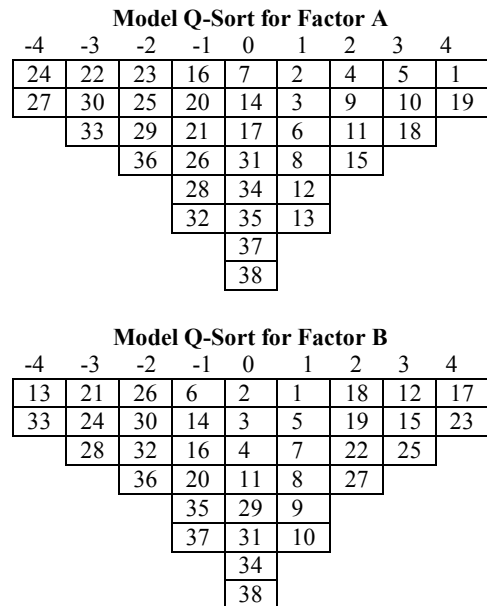| -4	-3	-2	-1	0	1	2	3	4
13	21	26	6	2	1	18	12	17
33	24	30	14	3	5	19	15	23
	28	32	16	4	7	22	25	
		36	20	11	8	27		
			35	29	9			
			37	31	10			
				34				
				38				

Figure 2. Model Q-Sorts (factor arrays) for Factors A and B.

The relative ranking of statements in each model Q-sort provides a rich source of information regarding the tensions, contradictions, and convictions expressed within

and among each of the factors. The factor scores calculated for each model Q-sort are compared statistically to reveal controversial and consensual statements between factors, at significance levels of $p<.01$ and $p<.05$. For example, in Table 1, statement 13 is a *controversial statement* for factors A and B, because it has been attributed opposing values in each model Q-sort (1 and -4, respectively). Consensus statements, like statement 16 in Table 1, are statements to which both factors have the same degree of agreement or disagreement. Table 2 shows how the statistical data shown in Table 1 directly relates to and effectively highlights opinions that are particularly controversial among participants.

Statement No.	Factor A	z-score	Factor B	z-score
13*	1	0.63	-4	-1.70
15+	2	1.16	3	1.22
16+	-1	-0.61	-1	-0.51
25*	-2	-0.94	3	1.36

Table 1. Factor z-scores for some consensual (+) and controversial (*) statements, at p<.05, for Factors A and B.

No.	Statement
13*	Personalized videos should be useful not only for communication, but also for cooperation.
15+	Personalized videos should give caregivers the ability to customize the AAC intervention as the child grows.
16+	Caregivers should be able to share personalized videos.
25*	I am concerned that personalized videos will cause a problem of invasion of privacy.

Table 2. Examples of controversial (highlighted in red) and consensual statements for Factors A and B.

Typically, after the participants complete the Q-sort procedure, they are asked to give an explanation for the statements with which they most agree and most disagree. Any demographic or other information relevant to the analysis of data is also collected at this time. This qualitative data is used to better understand the points of view that define each factor.

As we will see, in HCI-Q, factors clarify for the designer how the *a priori* assumptions about a design should be organized into design priorities from the perspective of users and other stakeholders. Herein lies the power of HCI-Q for capturing personal significance—what matters to users—and doing so in a robust, quantifiable way. Then, from a number of participants, the number of actual, underlying perspectives can be elicited through factor analysis. Often there are actually many fewer underlying perspectives that can characterize users than there are users themselves.

A Note about Statistics in Q-Methodology

Factors that emerge from analysis of Q-sorts in Q-methodology are not generalizable to a population of *people*, as in inferential statistics; rather, they are generalizable to a population of *statements*. In Q-methodology, the *statements* that participants sort into the

forced distribution (*see* Figure 1) are sampled for representativeness of a phenomenon, and the participants themselves are the variables. Factor analysis clarifies and reveals the subjective structure of the "universe" from which the statements were drawn, not the proportion of a population that adheres to a particular view.

Stephenson [21,22], the inventor of Q-methodology, outlined three rules for ensuring statistical soundness in a Q-methodology research design that pertain to sampling population of the statements in the Q-set: (1) homogeneity in kind; (2) heterogeneity among items of a kind; and (3) subjectivity. Homogeneity in kind is achieved by ensuring that each statement within the Q-set is sampled from the same "universe" of speech, literature, or other source of subjective statements about the phenomenon. Heterogeneity among items of a kind is achieved by ensuring that each statement is unique and that the Q-set expresses a balanced account of a phenomenon. Finally, the third rule of subjectivity is the principle that enables Q-methodology to elicit personal significance: statements must *not* be factual because facts cannot elicit subjective agreement or disagreement; they prohibit the ordinal ranking of subjective statements on which Q-methodology depends (*see* Figure 1). Q-methodology is most appropriately used when it is used to examine subjective experiences, attitudes, and opinions—things of great importance where personal significance of technologies is concerned.

AN HCI-TAILORED APPROACH TO Q-METHODOLOGY

HCI-Q uses the Q-set as a design methodology and the Q-sorting procedure as an evaluation methodology (Figure 3). These HCI-specific uses of the apparatus and procedure of Q-methodology optimize its value and utility for HCI contexts, especially iterative design.
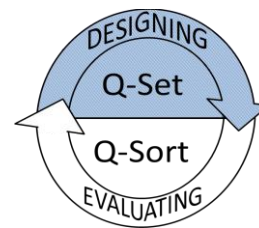

Figure 3. Using HCI-Q in an iterative design cycle.

HCI-Q is a discount approach to using Q-methodology for exploring the significance to users of positive and negative implications of design artifacts and ideas (Table 3). The HCI-specific customizations of Q-set craft and Q-sort technique are described next.

	Q	HCI-Q
Q-set sample	Interviews, literature, discourse.	Design, design idea, discussion about a design, description of a design.
Q-set structure	Experimental design, or unstructured.	Structured, equal parts positive and negative statements (≥ 15).
Participants	Purposive, $n<50$	Users and other stakeholders.

Q-sort	Make values explicit.	Users critically evaluate the design-as-statements.
Factors	Reveal patterns of subjectivity.	Reveal patterns of conflict and consensus among users directly relevant to design constraints.

Table 3. Our HCI-specific customization of Q-methodology.

Q-Set Craft as Design Methodology

The Q-set is the set of statements that is used as stimuli to elicit user perspectives. Within HCI-Q, the Q-set statements are the designer's *assumptions* about the personal and social implications of a technology that are likely to be perceived, negatively or positively, as particularly significant by stakeholders. Crafting the Q-set helps designers become aware of the assumptions that shape their interpretation of the problem they are addressing and the technological solution they are designing. Second, the assumptions that are surfaced during the Q-set craft are written in short, concise statements that can be communicated to design teams and other stakeholders. The result of Q-set craft is a set of subjective statements about a design that the designer submits to the users and other stakeholders for evaluation.

Traditionally, crafting the Q-set is the most time consuming part of implementing a Q-methodology study because it is common practice to craft statements based on analyses of interviews and relevant literature (*see* Table 3). Interviews, in particular, can be time-consuming to conduct, transcribe, code, and test for inter-coder reliability (*e.g.*, [1]).

To make it faster and to maximize its usefulness for design, HCI Q-set statements are gathered from analyses of designs. Designers reflect on the personal and social implications (both desired and feared) that are assumed to, or could conceivably follow from, properties of the design. For example, as it will be seen in the next section, we evaluated our design of a personalized video-based AAC system for its benefits for communication, socialization, literacy, and belonging as well as its risks of violating privacy, distracting attention, and creating value conflicts between clinicians and other caregivers. This critical reflection generates short Q-set statements of positive and negative opinions about the design or design idea.

The emphasis on both the positive and negative implications of a design is crucial in two respects. Firstly, design methodologies that encourage designers to consider the "darker side" of their designs [5,17] can catalyze critical and realistic reflection on the consequences that designs have in people's lives. Secondly, balancing the Q-set with an equal proportion of positive and negative statements (with a minimum of 15 statements each) is an effective way to avoid biasing the Q-set. Moreover, narrowing the scope of the design "universe" to a single key feature of a design that provokes social or collaborative uses, such as personalized video, helps to ensure a balanced, representative, and complete Q-set.

Q-Sort Technique as Evaluation Methodology

After the designer crafts the Q-set, stakeholders sort the statements therein. During the Q-sort technique, stakeholders reflect on and respond to the designer's assumptions about the implications of a design, which can help mitigate designer "blindness" to pitfalls and yield what stakeholders consider to be the opportunities of a design. Unlike open-ended responses to interview questions about desired or feared outcomes of a design, Q-sorts are compact, holistic, and quantifiable evaluations of designs that are directly related to specific design features. Attitudes towards designs-as-statements reflect back on that design, catalyzing innovation and iteration.

Next, when considering factor analysis, we find that an analogy helps. Just as a prism is used to separate a beam of light into a spectrum of its constituent colors, factor analysis of Q-sorts reveals the underlying spectrum of perspectives that are held by a group of stakeholders (Figure 4). Because the factors are quantitatively and qualitatively unique, they define areas of consensus *within* clusters and areas of tension *between* clusters. Furthermore, one of the advantages of HCI-Q is that it allows the designer or researcher to focus on and view clusters of data relative to extreme or unique points of view. For example, in the evaluation of a technology design, a factor defined by the Q-sort of a stakeholder who has a disproportionate amount of power in the social, political, or cultural context of use, may heed special attention and reveal the ways in which that outlier perspective relates to others in the group.
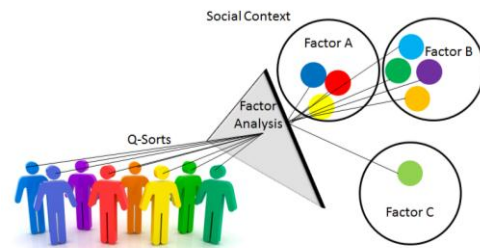

Figure 4. Underlying views emerging from factor analysis.

In summary, the HCI-specific customizations of Q-set craft and Q-sort technique, called "HCI-Q," can help HCI researchers and designers to be reflective practitioners and to study the subjective experiences of technologies from the point of view of the users and other stakeholders. HCI-Q reveals structures of subjective experience, of consensus and conflict, through the application of factor analysis to multiple points of view. Even when prototyping is not complete, designers can use HCI-Q to make explicit their own assumptions about what a design should do and invite stakeholders to assess the fit of those assumptions with their experiences and expectations of technologies in everyday life. The following sections of this paper provide an example of how to implement a Q-methodology study using our HCI-Q approach.

HCI-Q EVALUATION OF VID2SPEECH

Vid2Speech [19], built for Android tablets, is a prototype Augmentative and Alternative Communication (AAC) system for children with complex communication needs that uses personalized video to enhance graphical symbols of words. Unlike typical static symbols for AAC, personalized videos capture the movement, emotion, and context of actions and abstract concepts like "jump," "wait," "love," "tired," and "hungry." Caregivers can use the tablet's built-in front- and rear-facing cameras to capture new videos in daily life whenever an opportunity arises. Children use the videos for communication by tapping them, triggering speech output.

The goal of our use of HCI-Q with Vid2Speech was to understand whether the social and personal outcomes that we designed the system to facilitate were actually perceived by stakeholders as useful, important, and valuable. Furthermore, we were interested in understanding what negative implications of the design were perceived as most personally significant to stakeholders.

Participants

Fourteen AAC stakeholders were recruited by email from the United States, Ireland, and Canada. Our sample consisted of 3 parents of children who use AAC, 6 speech language pathologists who work with children either in schools or in private practice, 3 university faculty who specialize in AAC, and 2 teachers with students who use AAC. All participants were women. No participants had ever used Vid2Speech or video-based AAC.

Children with complex communication needs who are the primary intended users of Vid2Speech were not recruited because the HCI Q-set instrument is not an appropriate method for gathering their perspectives. The intended child users of this system are 5-12 years old and preliterate. In future work, we will use a complimentary method better suited to children with complex communication needs for gathering their perspectives. The current study was focused on caregivers.

Apparatus

We used Q-Assessor[1], a free online tool, to host our Q-set and deploy it to participants. We used a Q-set sample of 38 statements derived from our procedure of critical reflection on our assumptions about the social and personal outcomes facilitated by personalized video (see next subsection). We copied the Q-set statements into the Q-Assessor software and we provided participants with a link to our statements for sorting them online. We found Q-Assessor inadequate for data analysis because it does not provide a means for discarding non-significant factors before factor rotation, or for choosing different factor extraction methods. Therefore, we manually input the Q-sorts from Q-Assessor into PQMethod[2], an MS-DOS-based data-analysis program for Q-methodology. We derived our factors and examples from PQMethod.

Q-Set Procedure by Vid2Speech's Designers

As designers of Vid2Speech, we reflected on our assumptions about how we intended personalized video to affect the behavior of both caregivers and children using the system. Our aspirations for the system to provide a means for social inclusion, personal fulfillment, empowerment, and vocabulary expansion were tempered with our fears that the system may violate privacy, instigate power struggles for control over video representations, limit children's language development, or cause distraction from social exchanges. Our initial reflection on the design ideas for Vid2Speech resulted in a sample of 43 statements, which we evaluated based on the Q-set criteria for statistical soundness: homogeneity in kind, heterogeneity among items of a kind, and subjectivity. We rejected five statements based on those criteria, leaving a balanced set of 38 statements, 19 positive and 19 negative. Although we created the statements ourselves, our participants sorted them, meaning our resulting data cannot be said to be "from us" any more than an experimenter who designs an experiment's tasks can be said to have performed those tasks. Table 4 illustrates the design ideas for personalized video that we focused upon for this exercise.

Q-Sorting Procedure by Stakeholders

Participant stakeholders were given a brief description of the methods for capturing videos and using videos for communication with the Vid2Speech application. Participants were instructed to think of a child familiar to them who uses AAC as they considered each statement during the Q-sorting procedure.

The Q-Assessor software presented participants with one Q-set statement at a time and allowed them to drag-and-drop each statement into one of three piles: agree, disagree, or uncertain (*see* Figure 5). Statements were presented randomly. After participants completed the initial sorting stage, they were given the choice to review their piles and make changes or to continue. During the second sorting stage, participants were asked to put the two statements with which they most agreed and most disagreed in the designated boxes in the table (*see* Figure 6). Then participants could drag-and-drop the statements as they pleased into the remaining boxes.

After completing the Q-sorting procedure, participants were asked to give reasons for the two items with which they most agreed and most disagreed. They were also asked to give the age and current AAC strategies of the child AAC user they were thinking of as they sorted the statements. The survey took participants 20-30 minutes to complete.

---

[1] http://q-assessor.com/

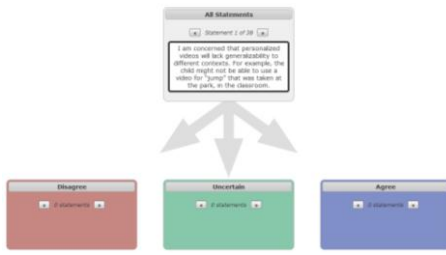[2] http://www.lrz.de/~schmolck/qmethod/

Figure 5. Q-Assessor drag-and-drop interface for initial sorting of statements into piles.
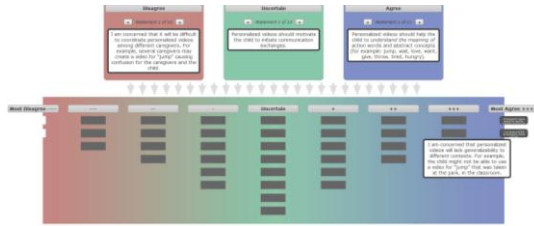

Figure 6. Q-Assessor interface for final sorting of statements in the forced distribution.

Analysis

We used PQMethod for factor extraction and rotation. One of the advantages of PQ Method is that it provides the correlation matrix of Q-sorts upon which the factor analysis is performed. With access to the correlation matrix, the user can verify that the factor analysis actually reflects the relationships between Q-sorts. For example, in our study, the Q-sorts of P7 and P8 had a highly significant correlation, and upon examination, we saw that they both loaded highly on the same factor. This gave us confidence that our factor analysis was reflecting true patterns in our data. We chose principal component analysis (PCA) for factor extraction and the *varimax* method for factor rotation. The reason we chose PCA is that it gives an explanation of the statistical variance explained by each factor that it extracts. This information is useful for deciding *objectively* how many factors to rotate. We had no theoretical or other reason for choosing to rotate factors subjectively, therefore PCA and *varimax* were best suited for our aims.

Unrotated Factor Matrix				
Factors	**1**	**2**	**3**	**4**
Sorts				
1	0.3694	0.4420	-0.3502	0.4447
2	0.6133	-0.0897	-0.4316	-0.2618
3	0.6769	-0.4846	0.1029	-0.1690
4	0.3290	0.3933	0.3462	-0.0881
5	0.6917	-0.4719	0.0543	-0.1234
6	0.3010	-0.3577	-0.2607	0.5858
7	0.8726	-0.0721	-0.1103	0.0340
8	0.8799	-0.0130	-0.0117	-0.1708
9	0.6367	0.5718	-0.2881	0.1122
10	-0.0751	0.7298	0.3588	0.0162
11	0.6497	0.1469	0.3397	-0.0086
12	0.7412	-0.1436	0.3397	-0.0086

13	0.4324	0.5486	-0.3038	-0.4363
14	0.6978	0.0907	0.3038	0.3004
Eigenvalues	5.2403	2.1599	1.1878	1.0019
% Pct. Var. Expl.	37%	15%	8%	7%

Figure 7. The first four factors extracted using PCA, with percentages of variance explained.

PCA automatically extracted eight factors, and we chose to rotate only the first two factors because they each explained the greatest amount of variance in the data, together over 50% of the total (Figure 7). After *varimax* rotation, PQMethod automatically flagged the Q-sorts that defined each factor (Figure 8) and then produced a complete analysis report, including the factor scores, factor arrays (*i.e.,* model Q-Sorts), and the controversial and consensual statements. Factor A was defined by 8 Q-sorts and Factor B by 6 Q-sorts. No Q-sorts were discarded in the analysis, meaning there were no confounds; all sorts loaded significantly on only one of the two factors.

		Factors	
		A	**B**
	1	0.0759	0.5710*
	2	0.5662*	0.2521
	3	0.8311*	-0.0476
	4	0.0677	0.5083*
	5	0.8368*	-0.0290
Sorts	**6**	0.4456*	-0.1414
	7	0.7760*	0.4056
	8	0.7505*	0.4594
	9	0.2324	0.8236*
	10	-0.4537	0.5766*
	11	0.4706	0.4715*
	12	0.7031*	0.2749
	13	0.0721	0.6948*
	14	0.5412*	0.4498

Figure 8. The two rotated factors, A and B, with defining Q-sorts indicated with an *.

RESULTS

The two-factor mathematical outcome described above is merely an index to a rich body of qualitative information. The factors represent values and concerns that characterize two quantitatively and qualitatively different aggregate viewpoints on the uses of personalized video for AAC. The following qualitative analysis of each factor is informed by the arrangement of statements in the model Q-sort for each factor, and by the qualitative responses from each participant about their reasons for ranking the statements with which they most agreed and most disagreed. The factors reveal two different evaluations of the significance of personalized video, one positive and the other skeptical.

Factor A: Personalization is Empowering

Factor A was defined by 8 people: 3 parents, 1 teacher, 3 speech language pathologists, and 1 researcher. The model Q-sort representing the perspective of Factor A revealed an overwhelmingly positive evaluation of personalized video. The most significant aspect of personalized video for these

stakeholders was its potential for empowering children with more effective means for expressive communication. This is reflected in participants' degree of agreement (from -4 to +4) with statements about personalized videos reflecting the unique culture and language of the child's family (+4), and the special interests and preferences of the child (+3). P12 exemplified this attitude: "Kids want to communicate about what is important to them—pictures of themselves, their family, their favorite toys, books, or movies."

Furthermore, these stakeholders felt that personalized videos could provide context to communication partners, who often struggle to understand AAC users' communication bids, resulting in communication breakdowns and lost opportunities for socialization. For example, P5 said, "It's usually the people around my child (myself included) that often fail to understand. The kids are incredible and make do with the limited communication options they have available." Showing further support for their conviction that personalized videos could be useful for establishing common ground with communication partners, stakeholders defining Factor A strongly disagreed with the statement suggesting caregivers would have difficulty understanding the videos (-3). On the contrary, by helping communication partners to understand the child's expressive communication, personalized videos were considered highly significant for their potential ability to help communication partners to relate to the child (+3). As P6 put it, "Communication should always be two-way. Caregivers can learn as much about the people they care for, in order to connect, respect, and build trust."

Stakeholders defining Factor A were not concerned that the children would have trouble recognizing the videos (0); rather, they felt strongly that personalized videos would help children understand the meaning of action words and abstract concepts (+4). P3 explained the strengths of personalized video for enhancing the meaning of action words: "Jumping involves smiling, grimacing, panting... [With video] the whole experience is represented." P2 put it another way: "Personalized videos are from the real world. It's the most direct way to show the concept."

Finally, Factor A was characterized by an optimistic attitude towards caregivers' abilities to capture and create relevant personalized videos. They strongly disagreed with the idea that personalized videos would be limiting for children due to lack of oversight by AAC experts (like speech language pathologists) (-3). P14, a speech language pathologist, identified parents as the true AAC experts, saying, "I don't think 'experts' have the best input to provide when deciding the content of what a child might like to communicate." Participants also strongly disagreed with the suggestion that personalized videos would be too difficult to create (-3) and suggested, in the words of P3, that "taking a video clip is a very straightforward process that most people use already on their phone or camera." The contrast between Factor A and Factor B could not be starker, as described next.

Factor B: Personalization is Burdensome

Factor B was defined by 6 people: 2 researchers, 3 speech language pathologists, and 1 teacher. The model Q-sort representing Factor B's perspective revealed a more skeptical stance towards personalized video. The most significant aspect of personalized video for these stakeholders was the potential burden on parents. They most agreed with the statement about personalization being too time-consuming for caregivers to create (+4). P11 sums up this attitude when she says, "In my experience with AAC users and their families, parents have very little time for technology management. While parents may have great intentions of creating videos, they are very busy managing the child's overall care." P9 echoed P11's sentiment, saying, "This fabulous idea to me as a therapist exhausts me as a parent." Stakeholders who aligned with Factor B agreed significantly with Factor A ($p<.05$) that personalized video could be valuable for reflecting the special interests of children (+2); however, they did not feel that personalization was sustainable over the long term. According to Factor B's view, personalized video is impractical, not only because of the parental burden it imposes, but also because video raises privacy concerns (+3). P13 explained that, "Privacy is a major concern because children who need to use the communication devices often do not have the ability to self-monitor for privacy when using the device."

Factor B also voiced concerns about the feasibility of video for use as a communication tool. P13 compared personalized videos to static symbols: "The richness of personalized videos… make it more vague what is actually being represented. For instance, the focus of a video of a kid eating snack at the table may be either the act of eating or the act of staying at an activity for a period of time." Because of concerns raised about the practicality of videos for representation, Factor B most agreed with the idea that personalized videos should be used to enhance any static AAC symbol set (such as Picture Communication Symbols) (+4) with which the child was already familiar. Personalized videos were considered appropriate as a supplementary method to static symbols and a tool for learning those symbols. P4 stated, "Many children can and should quickly outgrow the need for these kinds of videos."

Finally, unlike Factor A, Factor B expressed strong disagreement with the suggestion that personalized video could be useful as a tool for capturing evidence of the child's language development and facilitating cooperation among caregiver's (-4). P10 summed up this perspective by saying, "Caregivers should have ways of monitoring progress, but I believe this should be separate from a person's communication device. The device is designed for the AAC user, not the caregivers." Thus, Factor B represents an important counterpoint to Factor A.

DISCUSSION

The two factors that emerged resulted in a rich evaluation of the design ideas for Vid2Speech. The two factors

represent prototypical views of personalized video that can guide future directions for development of and research about personalized video in AAC. The participants' explanations for the statements in the Q-set with which they most agreed and most disagreed provided important insights that helped with the interpretation of the quantitative analysis and the two-factor solution. Table 4 below shows the results of the stakeholder evaluations as they relate to specific design ideas. Factor analysis helped us to uncover constraints on our design by revealing conflicting viewpoints about 2 design ideas, affirmation for 3 ideas, and rejection of 1 idea.

Design Ideas for *Vid2Speech*	Evaluation
Enable caregivers to capture videos in real time.	Contention
Pair videos with static images.	Yes
Facilitate expressive communication with speech output.	Contention
Facilitate receptive communication with video.	Yes
Support socializing with video.	Yes
Support sharing videos and coordinating care among caregivers with web site, email, Bluetooth or other social media features.	No

Table 4. Design ideas for Vid2Speech and their evaluation by stakeholders.

Our original assumptions about how to characterize the usefulness of personalized video were challenged by stakeholders. For example, one of the more surprising discoveries from our analysis was that speech output may *not* be the most significant feature of personalized videos from the perspective of stakeholders. Speech output *per se* was not particularly salient to either factor (1,0). This discovery was surprising because we had considered *lack of speech* to be the one of the main problems that we were designing for. This finding has helped us to view a key design feature in a new light, and to consider other social and personal implications deemed more significant by participants. For example, in their qualitative explanations, participants attributed high personal significance to helping children to establish common ground with communication partners, and supporting children to learn more abstract forms of representation. Moreover, Factor B's concerns about privacy helped us to see the seemingly socially appealing feature of sharing and coordinating videos as potentially hazardous from the perspective of users who may be particularly vulnerable to breaches of privacy.

Beyond providing constraints for design, the data from HCI-Q helped us to discover new opportunities for research and design that we had not previously considered. For example, one area of significant consensus ($p<.05$) was disagreement with the idea that "AAC experts" need to provide oversight for capturing relevant videos (-3,-2). Our original notion of experts included speech language pathologists and AAC researchers. Qualitative analysis of participant responses revealed that parents are considered the experts when it comes to personalizing AAC and that

we need to embrace them as key informants in our design process.

Lessons from Using HCI-Q
HCI-Q offers several advantages as a tool for research and design in HCI. It provides a snapshot of the structure, substance, and strength of conflict and consensus among the stakeholders of a technology. First, HCI-Q provides structure by reducing a large sample ($n\leq50$) of perspectives to a small number (≤7) of mutually exclusive clusters (called factors). Second, HCI-Q provides insight into the substance of the conflicts and consensus by identifying the specific social and personal implications that are controversial or agreeable. Third, HCI-Q provides information about the strength of conflicts and consensus by providing the statistical significance of the differences or similarities in opinion, and the ordinal value (from -4 to +4) assigned by each factor to each design constraint or goal. Beyond these three advantages of quantification, HCI-Q also provides a small amount (~75-100 words per person) of structured qualitative data about personal significance directly relevant to the interpretation of the factors.

Because of its focus on *personal significance*, HCI-Q is not very suitable for evaluating productivity features or the technical (rather than *social*) usability of technologies. It is most suitable in HCI contexts wherein the usefulness of a technology is likely to be characterized in terms of the particular social and personal judgments of its stakeholders, especially when those judgments are likely to be in tension due to unequal power relationships or competing priorities. The ease with which HCI-Q gathers responses online, analyzes them automatically, and reduces them for simplicity was encouraging for prospective use in time-sensitive and resource-constrained iterative design cycles. Moreover, we were encouraged that our results provided constraints on a design *idea* with which none of our participants were initially familiar. The ability to evaluate designs at the idea stage is promising for future uses of HCI-Q for pre-prototyping explorations, the most important outcome of which might be the ability of stakeholders to define the role of technologies in their social and personal lives.

FUTURE WORK
We plan to create a tailored program for implementing HCI-Q that will obviate the need for two separate software programs for data collection and analysis. With an HCI-Q specific online tool, we will explore how the tool can be optimized so that designers with little familiarity with statistics can use HCI-Q most effectively. Future work will also explore the most effective way to introduce HCI-Q into a PD approach, for example, after participant observation wherein the designer's assumptions are formed, but before focus groups and interviews wherein the results of HCI-Q can be used to build mutual understanding.

CONCLUSION
We have demonstrated the utility and value of HCI-Q, a new method adaptation for understanding the personal

significance of technologies from the perspectives of users. We showed that HCI-Q has the advantage of reducing large amounts of data into a few rich and informative clusters, called "factors." This data reduction is achieved with quantitative techniques that are computed by freely available software programs, and gives statistical support to the qualitative interpretation of each factor. Data analysis provides constraints on design thereby revealing areas of consensus and conflict among stakeholders, and aids the discovery of new opportunities for research and design. It is our hope that HCI-Q can become a useful method for both research and design in human-computer interaction.

ACKNOWLEDGMENTS

REFERENCES

1. Anandarajan, M., Paravastu, N. and Simmers, C. A. Perceptions of personal web usage in the workplace: A Q-Methodology approach. *Cyberpsychology and Behavior, 9,* 3 (2006), 325-335.

2. Best, M.L., Wornyo, E., Smyth, T.N. and Etherton, J. Uses of mobile phones in post-conflict Liberia. In *Proc. ICTD 2009*, IEEE Press (2009), 468-477.

3. Brown, S. R. *Political subjectivity: Applications of Q Methodology in political science*. Yale University Press, New Haven, 1980.

4. Carroll, J. M. Five reasons for scenario-based design. *Interacting with Computers, 13,* 1 (2000), 43-60.

5. Dunne, A. and Raby, F. *Design noir: The secret life of electronic objects*. August Media, Boston, MA, 2001.

6. Dziopa, F. and Ahern, K. A systematic literature review of the applications of Q-Technique and its methodology. *Methodology: European J. of Research Methods for the Behavioral and Social Sciences, 7*, 2 (2011), 39-55.

7. Friedman, B. Value sensitive design. In W.S. Bainbridge (ed.), *Encyclopedia of Human-Computer Interaction*, Berkshire Publishing Group, Great Barrington, MA, (2004), 769-774.

8. Gaver, W. W., Boucher, A., Pennington, S., & Walker, B. Cultural probes and the value of uncertainty. *interactions*, *11*,5, (2004), 53-56.

9. Greenbaum, J. and Kyng, M. *Design at work: Cooperative design of computer systems*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1991.

10. Hassenzahl, M., & Wessler, R. Capturing design space from a user perspective: The repertory grid technique revisited. *International Journal of Human- Computer Interaction*, *3-4*,12, (2000), 441–459.

11. Holtzblatt, K. and Jones, S. Contextual inquiry: A participatory technique for system design. In *Participatory Design: Principles and Practices*, D. Schuler and A. Namioka (Eds). Lawrence Erlbaum, Hillsdale, NJ, (1993), 177-210.

12. Kensing, F. and Madsen, K.H. Generating visions: Future workshops and metaphorical design. In J. Greenbaum and M. Kyng (eds.), *Design at Work: Cooperative Design of Computer Systems.* Lawrence Erlbaum, Hillsdale, NJ, (1991), 155-168.

13. Kim, J., and Mueller, C. W. *Factor analysis: Statistical methods and practical issues*. Sage Publications, Newbury Park, CA, 1978.

14. McKeown, B. and Thomas, D. *Q Methodology*. Sage Publications, Newbury Park, CA, 1988.

15. Meloche, J. A. Q Methodology as a research methodology for human computer interaction. In *Proc. OzCHI 1999*, Charles Sturt University, (1999), 149-152.

16. Moggridge, B. Futures and alternative nows. In *Designing Interactions*. M.I.T. Press, Cambridge, MA, (2007), 587-611.

17. Nathan, L.P., Klasnja, P.V. and Friedman, B. Value scenarios: A technique for envisioning systemic effects of new technologies. In *Ext. Abstracts CHI 2007*, ACM Press (2007), 2585-2590.

18. Nielsen, J. *Usability engineering*. Academic Press, San Diego, CA, 1993.

19. O'Leary, K., Delahunt, C., Dowden, P., Darmansya, I., Heng, J., Riskin, E.A., Ladner, R.E. and Wobbrock, J.O. Design goals for a system for enhancing AAC with personalized video. *Proc. ASSETS 2012*, ACM Press, (2012), 223-224.

20. Shapiro, D. The limits of ethnography: Combining social sciences for CSCW. In *Proc. CSCW 1994*, ACM Press (1994), 417-428.

21. Stephenson, W. *The study of behavior: Q-Technique and its methodology*. Univ. of Chicago Press, Chicago, IL, (1953).

22. Stephenson, W. Correlating persons instead of tests. *Character and Personality*, 4, (1935), 17-24.

23. Watts, S. and Stenner, P. *Doing Q Methodological research: Theory, method and interpretation*. Sage, Los Angeles, CA, 2012.

24. Watts, S. and Stenner, P. Doing Q methodology: theory, method and interpretation. *Qualitative Research in Psychology, 2,* 1 (2005), 67-91.

25. Winograd, T., & Flores, F. *Understanding computers and cognition: A new foundation for design*. Addison-Wesley, Reading, MA, 1987.