

## **Operant Subjectivity**

*The International Journal of Q Methodology*

# **Participant Engagement and Data Reliability with Internet-Based Q Methodology: A Cautionary Tale**

Matthew Dairon

Shari Clare

John R. Parkins

*University of Alberta*

**Abstract:** Q methodology traditionally involves the sorting of stimuli such as textual phrases or images that are then analyzed with statistical software. Coupled with these quantitative techniques, Q methodology often involves in-depth interviews and interpretative methods. In spite of these mixed-methods strengths, scholars are turning to internet-based platforms for administering Q studies, allowing for a greater range of access to a larger pool of potential participants. In this article, we examine issues related to participant engagement and the potential impact of low-quality sorts on data reliability. These issues are particularly germane for studies utilizing online platforms for administering Q methodology studies, where the distance between researcher and participants is increased. Our analysis involves the generation of random Q sorts as a proxy for low-quality data and explores the influence of introduced low-quality data on factor loadings and interpretation. In our exploratory study, we find that the introduction of even a small number of low-quality sorts can seriously influence factor loadings; in particular, these random sorts alter the composition of Q sorts that load on less dominant “minority” factors and, ultimately, the interpretation of factors. Based on these findings, we propose an approach that allows Q methodology researchers to explore further the quality of their data to detect low-quality sorts and offer suggestions for improving participant engagement in online studies.

**Keywords:** data quality, internet surveys, online methods, participant engagement, Q-sort reliability

## **Introduction**

One of the benefits of Q methodology is a more personal approach to participant engagement. Generally, Q studies include a relatively small number of participants who are asked to prioritize and sort a set of stimuli (normally textual statements or images), then provide an explanation of their views using the provided stimuli as a guide. Sorting of the stimuli (defined as a Q sort) is typically conducted face-to-face, allowing for interaction between researchers and participants before, during and after the sorting exercise (Watts & Stenner, 2012). This direct contact with participants allows researchers to explain the Q sorting process prior to administering the sort. It also allows the researcher to gauge participants’ understanding of and engagement in the sorting process, which gives the researcher an opportunity to provide additional

Contact author: [jparkins@ualberta.ca](mailto:jparkins@ualberta.ca)

© 2017 The International Society for the Scientific Study of Subjectivity ©2017 The Authors

instruction during sorting, if required. In many Q studies, the sorting exercise concludes with an interview in which participants narrate their sort and provide a rationale for the placement of the stimuli (Watts & Stenner, 2012). This face-to-face interaction increases the richness of the data and provides researchers with critical insights they often draw upon during analysis and interpretation of the data (Gallagher & Porock, 2010).

While traditional Q methodology studies provide a richness of qualitative and quantitative information, the reliance on face-to-face contact does have drawbacks, including limits on geographic study range and increasing research costs (Reber et al., 2000). Due to these limitations, some scholars are turning to internet-based options to engage participants in Q studies. With the adoption of online methods, a growing number of scholars are gathering information from a larger pool of participants using more general recruitment methods (e.g., panels of online participants versus purposeful sampling). In doing so, researchers are expanding participation beyond a select few key informants and beginning to use Q methodology to examine the views of the general public (e.g., Davis & Michelle, 2011). These online approaches often diverge from traditional Q methodology procedures, which typically include face-to-face engagement with a smaller number of subjects selected to participate because they are considered to be highly engaged and intimately familiar with the topic in question. Thus, there are concerns that the move to online Q procedures may result in diminished data quality and analytical insight due to the increased distance between participants and researchers. There are also concerns about the inclusion of participants who may be less motivated to engage thoughtfully with the sorting process. In particular, given that sorting through a large number of stimuli is meant to be a thought-provoking process, Q scholars have expressed concerns related to online sorting completion times and whether Q sorts finished in a very short time are reliable. Furthermore, given that sorting is done independent of the researcher, there is greater potential for the participant to misunderstand the instructions for sorting or to be confused by the sorting process, which may also lead to lower quality data from online platforms. Finally, depending upon how participants are recruited into the study, participants of online studies may or may not be familiar or engaged with the topic under examination, which may influence their motivation to sort through the stimuli following the prescribed steps.

The objective in this article is to explore the potential effects of “low-quality” Q sorts on factor loading and data interpretation in order to advance understanding of how internet-based data collection methods might impact Q study findings. In the context of this study, we define “low-quality” sorts as those done by participants who do not have a complete understanding of the research question and/or sorting process or who are not fully engaged in the topic under examination. As a proxy for low-quality sorts, we generated random sorts, which we consider to approximate a haphazard sorting approach that one might observe with participants who are not engaged in or do not understand the sorting process. While random sorts may not precisely represent a disengaged participant (i.e., we have not tested whether disengagement and randomness are statistically comparable), we see the use of random sorts as a valid first step in exploring how low-quality data may influence factor loadings and factor interpretation.

In the next section, we review the literature regarding online Q methodology research. We then provide empirical evidence of how low-quality data impacts factor loadings and factor interpretation using data collected from our own online Q study. Our dataset is derived from an online survey of 105 participants who sorted statements

about energy development in Alberta, Canada. To explore the sensitivity of this dataset to low-quality sorts, we introduce a range of five to 25 random sorts, then explore implications for the identification and interpretation of factors. We conclude the article by discussing the qualitative and quantitative dimensions of Q methodology that can be preserved within web-based methods to enhance data quality.

### **Strengths and Weaknesses of Online Methods**

For researchers interested in using online methods to conduct Q studies, there are a number of software platforms available, including “Q-Assessor, WebQSort, Web-Q, QSorter, FlashQ and Hotspot” (Davis & Michelle, 2011, p. 575). As online Q-sort software has advanced, researchers seek to emulate the traditional sorting process by incorporating drag and drop capabilities, thus creating a more user-friendly interface (Davis & Michelle, 2011). For instance, software packages that offer drag and drop functionality, such as Flash Q, simulate an in-person experience by requiring participants to first examine each statement, then sort the statement into one of three “piles”: agree, neutral and disagree. After completing this initial sorting, participants are transferred to a second screen with the three piles of statements and a digital representation of the Q-sort board. Participants then move the statements from the three piles onto the Q-sort board by dragging and dropping the statements to the desired column, per the statement of instruction for sorting. As with traditional Q sorts, participants have the opportunity to rearrange the statements on the board once they have been placed. Some software packages allow researchers to track the time it takes participants to complete different aspects of the exercise. With this feature, researchers can document the duration of time it takes participants to complete both the initial and final sorts, providing a greater understanding of how the participants’ time was spent (Reber et al., 2000). Most software packages also allow researchers to include a post-sorting survey that requires participants to provide additional rationale for how the stimuli were sorted.

A major advantage of using online methods to conduct social science research is a reduction in cost associated with connecting to a more diverse public across large geographies (Dillman et al., 2009). Such access to the general public allows researchers to use Q methodology to identify under-examined points of views that challenge mainstream thinking – views that are often overshadowed by dominant perspectives. In addition to providing access to more diverse populations, web-based Q studies offer other benefits that may attract researchers. For example, in their analysis of Q-Assessor, Reber and colleagues (2000) note that administering Q sorts using web-based platforms reduces the time needed to gather and process data because the data are instantly transferred to the researcher once the participant has finished sorting. In cases where participants type in their responses, the results of the sorting do not have to be transcribed, which helps reduce error and saves time and money. Reber et al. (2000) also note that, with web-based Q sort, there is the potential for faster response times, as some participants are able to complete the sorting immediately after the invitation has been sent to the study participant.

Although web-based Q sorts do offer several benefits, there are also disadvantages unique to this approach. Unlike traditional Q-sort methods, web-based sorting does not easily allow participants to view all the statements or stimuli at once (Reber et al., 2000). As a result, scholars such as Watts and Stenner (2012) question whether web-based platforms achieve the level of comparative evaluation needed for individuals to effectively complete the sorting process. Another weakness of web-based Q sorting

involves software compatibility with certain web browsers like Firefox or Explorer (Reber et al., 2000). Similarly, if participants are not comfortable with computers or new software, they may find it difficult to complete the sort or may avoid it altogether (Reber et al., 2000). Depending on how participants are selected for a study, the differences in participants' willingness and ability to participate in a study may influence which groups or points of view are represented in a p set (Davis & Michelle, 2011).

Furthermore, a major drawback of Q studies using online methods is that researchers rarely conduct post-sorting interviews with participants, particularly for studies with large p sets drawn from large geographies. This separation between researcher and participant may seriously limit the ability of the researcher to understand the data, because interpretation often requires the researcher to understand subtle nuances in the subjectivity of participants' opinions. These subtleties may not be immediately apparent through statistical procedures alone (Brown, 1989; Brown, 2015). Also, researchers are not present to ensure that sorting is being done correctly and with minimal frustration. The increased distance between participant and researcher makes it difficult to assess participant engagement, leading to concerns about the quality of data obtained through online methods.

To address concerns about web-based Q sorts resulting in low-quality data, scholars working with such datasets suggest several additional steps in the data analysis process, including reliability or replicability tests (Ramlo, 2016a). These tests consist of collecting preliminary web-based Q sorts to determine the range of time needed to adequately and reliably complete the sort. With this information, the researcher can eliminate Q sorts that do not fall within the range of what he or she considers acceptable based on the complexity of the subject and number of statements. Similarly, researchers can analyze sorting completion times to identify potential outliers who would require additional attention (Pruneddu, 2015). While time to completion may be used to generally assess data quality in online Q studies, this metric alone may not be the most reliable means for assessing the reliability of sorts. In the analysis below, we examine completion times for Q sorts administered online and analyze how randomly generated Q sorts (as an extreme representation of low-quality sorts) influence factor loadings and interpretation.

## Methods

To address our study objectives, we examined Q sort data collected online using custom-built software. This dataset consists of responses from 105 participants from across Alberta, Canada, and is part of a larger web-based Q study dataset including participant responses from two additional Canadian provinces: New Brunswick and Ontario. (To access datasets and additional study information, refer to the online material by Hempel et al., 2015.) The intent of collecting data from a larger number of participants over a greater geographic extent than is typical of a Q study is twofold. First, the research team wished to explore and compare the general public's views on energy development in Canada. Q methodology traditionally focuses on members of the public "who have a defined viewpoint to express and, even more importantly, whose viewpoint *matters* in relation to the subject matter" (Watts & Stenner, 2012, pp. 70-71, original emphasis). Given this focus on small samples sizes, Q studies may exclude the views of the general public, or lay people, in favor of sampling the views of key informants. Although the general public may be relatively uninformed on technical issues such as energy development, their views are often critical to political debates and

public policy. Consequently, public perspectives on energy development in Canada are an important aspect of our research.

The second reason for administering a web-based Q sort was to test the capabilities of this platform to reliably gather data from the general public. The study included 48 general statements about energy development and the role of the energy sector in Canadian society (Parkins et al., 2015). The statements included a wide range of energy forms and technologies and, therefore, were considered appropriate for exploring the diversity of views on energy development in Canada.

Corporate Research Associates, Inc., a Canadian-based polling firm, recruited participants through a panel of online subjects. With attention to quotas for age, gender and education, the polling firm ensured our participants were representative of the demographics in each of the three provinces included in the study. Furthermore, the polling firm provided small incentives to participants taking part in the study. These incentives involved the accumulation of points to purchase items from the firm's catalogue. Due to software compatibility issues, existing drag-and-drop Q-sort software was not used; instead, the polling firm created custom software designed with similarities to FlashQ. For the online Q sort, participants were first presented with instructions on how to complete the exercise. They were then directed to sort the 48 statements, first by dragging and dropping them into three piles (agree, disagree or neutral), then by dragging them onto the digital Q-sort board. Following completion of the sort, participants were asked to answer a set of post-sorting questions to provide further insights into the reasons for why the participants chose the statements that were most agreeable and disagreeable. The post-sorting questions also involved demographic details, such as level of education, household income and postal code.

To test the impact of low-quality sorts on data structure and the interpretation of each factor's meaning, we introduced random Q sorts into the participant-derived dataset. The low-quality sorts were created to approximate the haphazard sorting of participants who are not engaged in or do not completely understand the sorting process. This was done by first assigning a number ranging from 1 to 48 to each of the spaces within the quasi-normal distribution, which included 11 columns ranging from strongly agree to strongly disagree (1, 2, 4, 6, 7, 8, 7, 6, 4, 2, 1). For example, number 1 corresponded to the single space under the +5 strongly agree column, while numbers 2 and 3 corresponded to the two spaces under the +4 "most agree" column, and so on. We created Q sorts for 25 "participants" (identified as R1 through R25) by first randomly assigning a number from 1 to 48 to each of the statements, creating a numerical reference to each statement. Forty-eight different random numbers were then generated and arranged within a digital representation of the Q-sort board. The rank of each of the 48 random numbers was then determined, providing a reference to the numbered statements. This process of randomly generating and ranking numbers within a digital Q-sort board was completed 25 times, creating 25 unique Q sorts.

To test the sensitivity of the factor analysis results on the introduction of low-quality sorts, we introduced a different number of randomly generated sorts to the participant-derived data set. Five was the smallest number of random sorts introduced to the participant dataset. We tested the effects of introducing five random sorts four different times, each time using a different combination of random sorts (R1-R5, R6-R10, R11-R15 and R16-R20). We also tested the effects of introducing 10 random sorts (R1-R10), as well as the effects of adding all 25 random sorts. For each of the six random data combinations, we extracted a three-, four-, five- and six-factor solution analysis, for a

total of 24 different factor analyses. All data analyses were completed using PQMethod software (Schmolck, 2014). Consistent with Q methodology analytical options (Watts & Stenner, 2012), we conducted principal component analysis with Varimax rotation.

To further explore how random Q sorts influence factor loadings of nonrandom sorts, we completed a more detailed analysis of results taken from a five-factor solution using the participant dataset plus all 25 random sorts. We compared factor loadings for each of the 105 participants before and after the introduction of the 25 random sorts, and examined how the loading behavior changed. In particular, we were interested in how the introduction of random data influenced Factors 3, 4 and 5, which we consider to represent minority discourses that are often distinct from the more dominant discourses typically represented by Factors 1 and 2. We also examined how the introduction of random data influenced the loading behavior of participant-derived sorts as a function of “time to completion.” This was done to determine whether the introduction of random data could provide any insights into the reliability of sorts that were clear outliers with respect to the average time to completion for our study.

It should be noted here that these randomly generated sorts are considered to be an extreme representation of data generated by an individual who is disengaged or does not fully understand the sorting process. In other words, we have assumed that if an individual puts little or no effort into the sorting process, the outcome of the sort will approximate randomness. Given the need for extensive analysis (beyond the scope of this article), we have not tested the assumption that random sorts are statistically comparable to sorts generated by disengaged participants or by those who do not completely understand the sorting process. Rather, the intent here is to test the sensitivity of a participant-derived dataset to Q sorts not obtained using the strict quality control methods typical of in-person Q sorting and to explore how the introduction of such data may impact factor loadings and interpretation.

## Results

### The Effects of Random Sorts on Data Structure

Factor loading values provide insights into how statistically significant Q sorts contribute to the definition of a given factor (Brown, 1980); thus, by assessing factor loading values, we can gain an understanding of how factor interpretation may be affected by the introduction of randomly generated sorts. In all but two of the 24 different analyses conducted, randomly generated Q sorts significantly loaded on at least one factor (Table 1). As a greater number of randomly generated Q sorts were introduced, an increasing number of the random sorts significantly loaded on one or more factors. Similarly, as more factors were extracted, a greater number of randomly generated Q sorts significantly loaded on one or more factor.

**Table 1. Number of Random Sorts Significantly Loading on Each Factor, as a Function of Number of Factors Extracted and Number of Randomly Generated Q Sorts Introduced into Factor Analysis**

Number of Factors Extracted	Number of random sorts introduced and number loading on each factor					
	5 (R1 – R5) <sup>a</sup>	5 (R6 – R10) <sup>a</sup>	5 (R11 – R15) <sup>a</sup>	5 (R16 – R20) <sup>a</sup>	10 (R1 – R10) <sup>a</sup>	25 (R1 – R25) <sup>a</sup>
3	0	2	2	1	2	4
4	0	1	2	3	2	6
5	1	3	2	1	4	11
6	3	3	4	2	5	13

<sup>a</sup> Identifies which randomly generated Q sorts from the dataset of 25 randomly generated Q sorts were introduced. Random sorts were introduced into a participant-derived dataset (n=105) collected as part of a study examining energy discourses in Canada.

### Using Random Sorts to Examine Reliability of Participant Data

When the loading patterns for our participant-derived Q sorts were examined by completion time category, interesting patterns were apparent (Table 2).

**Table 2. Results of Five-Factor PCA Using Participant-Derived Dataset (n=105), with Number and Proportion of Significant Loadings for Each Factor Presented by Completion Time Category**

Completion category	Factor					Significant Sorts per Category	Significant Sorts (%)
	One	Two	Three	Four	Five		
7-10 Minutes	2	0	1	0	1	4 / 8	50
11-20 Minutes	12	6	1	0	2	21 / 38	55
21-30 Minutes	18	4	1	1	2	26 / 32	81
31-40 Minutes	1	0	0	1	0	2 / 5	40
41-50 Minutes	2	0	0	1	0	3 / 6	50
>51 Minutes	6	4	1	0	0	11 / 16	69
Total # of sorts	41	14	4	3	5	67 / 105	64

**Table 3. Results of Five-Factor PCA Using Participant-Derived Dataset plus 25 Random Sorts (n=130), with Number and Proportion of Significant Loadings for Nonrandom Sorts by Completion Time and Number of Random Sorts Significantly Loading on Each Factor**

Completion Category	Factor					Significant Sorts per Category	Significant Sorts (%)
	One	Two	Three	Four	Five		
7-10 Minutes	3	1	1	1	0	6 / 8	75
11-20 Minutes	18	7	0	0	1	26 / 38	68
21-30 Minutes	21	5	0	0	0	26 / 32	81
31-40 Minutes	1	0	0	0	0	1 / 5	20
41-50 Minutes	5	0	0	0	0	5 / 6	83
>51 Minutes	6	4	0	0	0	10 / 16	63
Random Q sorts	0	1	5	3	2	11 / 25	44
Nonrandom sorts	54	17	1	1	1	74 / 105	70

First, the majority of participants (70 of 105) took 11 to 30 minutes to complete the Q sorting. Forty-seven of these sorts significantly loaded on one of the five factors extracted for the analysis, the majority of which (40 sorts) loaded on one of the dominant factors (i.e., Factors 1 or 2). The remaining seven sorts loaded on the minority factors (i.e., Factors 3, 4 or 5), with four of these sorts clearly defining the fifth factor. The remaining 20 significant sorts were distributed throughout the completion time categories, with the majority (11 of 20) coming from the >51 minute category. Eight sorts came from the shortest completion time category (7 to 10 minutes), half of which significantly loaded on one of the five factors, with two of the sorts loading on the first factor and two loading on the minority factors. While there were initial concerns that the “fast” sorts (7 to 10 minutes) were low quality and should be removed from the dataset, there was little justification for eliminating sorts on the basis of time to completion alone. For example, if time to completion was used as the criteria for eliminating sorts, we faced the question of what the appropriate cut-off time should be. Further, our assumption was that “fast” sorts were potentially unreliable, but there were also questions about the reliability of “long” sorts, with completion times exceeding 51 minutes. Given the difficulty of using an arbitrary time to completion as the only criterion for assessing reliability, we sought to use random data to further evaluate the reliability of our participant data. Additionally, including random data with

participant-derived data provided insights into how lower reliability data might influence factor loadings and data interpretation.

When the completion times for the nonrandom, participant-derived Q sort dataset (n=105, Table 2) were compared with factor loading patterns of the participant-derived data plus all 25 random sorts (n=130, Table 3), it was clear the addition of the random sorts caused a shift in the loading of nonrandom sorts from minority factors to more dominant factors. Just under half (11 of 25) of the random sorts significantly loaded on a factor, with the majority of these random sorts (10 of 11) loading on the minority factors (Table 3). While the number of significantly loading nonrandom sorts increased from 67 to 74, only three nonrandom sorts loaded on the minority factors, with two of these sorts being from the 7- to 10-minute completion category (Table 3). The short completion time for these nonrandom sorts and the fact that these sorts loaded on a factor with only random sorts made us question the reliability of these particular sorts.

In Table 4, we offer a more detailed analysis of the results presented in Tables 2 and 3.

**Table 4. Examples of Different Loading Behaviors of Participant-Generated Sorts Before and After Introduction of 25 Randomly Generated Sorts**

ID	Time to Completion	Non-Random					Non-Random + 25 Random				
		Factor					Factor				
		1	2	3	4	5	1	2	3	4	5
R1	Random	-	-	-	-	-	0.32	-0.06	0.00	-0.34	<b>0.48</b>
R3	Random	-	-	-	-	-	-0.15	0.13	<b>0.42</b>	0.17	0.21
R8	Random	-	-	-	-	-	-0.05	<b>-0.52</b>	-0.34	0.00	-0.31
R9	Random	-	-	-	-	-	0.17	-0.13	<b>0.40</b>	-0.11	0.04
R10	Random	-	-	-	-	-	0.21	-0.08	-0.10	<b>0.48</b>	-0.01
R12	Random	-	-	-	-	-	-0.05	0.02	<b>-0.41</b>	-0.04	0.31
R15	Random	-	-	-	-	-	0.15	-0.07	-0.05	-0.09	<b>-0.44</b>
R16	Random	-	-	-	-	-	0.21	0.07	-0.10	<b>0.62</b>	-0.18
R18	Random	-	-	-	-	-	0.17	0.07	-0.11	<b>0.43</b>	0.18
R21	Random	-	-	-	-	-	0.18	-0.06	<b>0.41</b>	-0.08	0.00
R23	Random	-	-	-	-	-	0.19	0.02	<b>0.51</b>	0.29	-0.04
154	7	0.22	0.09	0.33	-0.20	0.22	0.17	0.10	0.06	<b>0.58</b>	0.05
22	13	0.04	0.00	0.20	-0.04	-0.21	0.07	0.04	-0.02	0.10	<b>-0.40</b>
690	9	<b>0.78</b>	0.06	0.19	0.03	-0.03	<b>0.75</b>	0.05	0.02	0.24	0.01
691	11	<b>0.48</b>	0.19	0.11	0.25	0.26	<b>0.54</b>	0.18	0.01	0.12	0.14
118	12	-0.13	<b>0.64</b>	0.21	0.17	0.07	-0.02	<b>0.67</b>	0.09	0.02	-0.02
547	22	-0.27	<b>0.70</b>	0.03	0.22	-0.20	-0.14	<b>0.72</b>	0.01	-0.13	-0.26
499	57	-0.23	<b>0.63</b>	0.13	-0.07	0.19	-0.21	<b>0.64</b>	-0.07	0.10	0.00
859	10	-0.05	0.11	<b>0.71</b>	0.19	0.05	0.13	0.18	<b>0.55</b>	-0.14	0.06
476	12	0.01	0.32	<b>0.55</b>	-0.03	-0.03	0.09	<b>0.38</b>	0.29	-0.07	-0.05
697	27	-0.31	0.37	0.03	<b>0.48</b>	-0.23	-0.05	<b>0.38</b>	-0.12	-0.37	-0.37
933	45	0.14	-0.26	0.19	<b>0.62</b>	-0.05	<b>0.41</b>	-0.23	0.04	-0.20	-0.23
567	22	0.27	0.20	-0.09	0.13	<b>-0.54</b>	0.28	0.20	0.01	<b>-0.39</b>	0.00
558	26	0.32	-0.06	<b>0.73</b>	0.22	-0.03	<b>0.48</b>	0.02	<b>0.62</b>	-0.02	-0.15
721	34	0.16	0.09	-0.09	<b>0.61</b>	-0.19	<b>0.38</b>	0.09	-0.13	<b>-0.44</b>	-0.02
779	61	0.34	0.25	<b>0.39</b>	-0.07	0.09	0.32	0.28	0.35	0.06	0.23
850	8	-0.01	0.03	-0.03	0.11	<b>-0.42</b>	0.04	0.03	-0.01	-0.21	-0.20
023	13	0.12	0.28	0.03	0.26	<b>0.66</b>	0.19	0.26	0.00	0.37	0.26
583	20	0.22	0.21	0.10	0.28	<b>-0.39</b>	0.33	0.24	0.10	-0.35	-0.10
434	21	0.21	0.25	-0.08	0.30	<b>0.43</b>	0.28	0.24	-0.04	0.13	0.23
<b>Variance explained (%)</b>		24	12	5	7	3	23	10	4	4	3
<b>Eigenvalue</b>		31.9	10.7	3.8	3.8	3.3	32.6	11.5	5.3	4.6	4.1



First, we show the exact location of the factor loadings for a representative number of sorts, both before and after random sorts were introduced. Additionally, we show how nonrandom sorts were impacted by the introduction of random data. In all cases, there was a tendency for randomly generated sorts to load on nondominant factors (Table 4). For instance, when a five-factor solution was extracted, the random sorts typically loaded on the third, fourth or fifth factors (the minority viewpoints), rather than on the first two factors (the more dominant viewpoints). When randomly generated Q sorts did load on a common factor, it was associated with instances where fewer factors were extracted, and in these cases, the random sorts tended to load negatively on the dominant factors. Additionally, as the number of randomly generated Q sorts increased, there was a tendency for the nonrandom sorts to become more concentrated on the two dominant factors, while the random sorts loaded more frequently on the minority factors. Further, we observed factor loadings for sorts that only emerged after random data were introduced (e.g., ID 154 and 22), sorts that remained stable on Factors 1, 2, or 3 (ID 690, 691, 118, 547, 499, and 859), sorts that moved from one factor to another (e.g., ID 476, 697, 933, and 567), and sorts that became confounded (e.g., ID 558 and 721). We also found sorts that loaded on the minority factors that became displaced by the random sorts and no longer loaded on any of factor (ID 779, 850, 023, 583, and 434). In general, the introduction of low-quality sorts, represented by randomly generated sorts, dramatically impacted the factor loadings and data structure of participant-derived sorts.

### **The Effects of Random Sorts on Data Interpretation**

The final step in understanding how low-quality data can influence Q studies involved an examination of how the defining statements of each factor changed, and how this may have impacted the researcher's interpretation of results on the basis of the defining statements. With the introduction of 25 random sorts, we noted that all five factors experienced substantial changes in the position of statements within the factor array (Table 5). In Factor 1, 25% of statements changed position. For example, the statement "Climate change poses a grave and urgent threat to our planet" moved from position +2 to +3. For Factors 1 and 2, the impact of randomness on the location of statements in the factor array was less significant; however, the impact was much more dramatic on minority Factors 3, 4 and 5. For example, in Factor 4, the introduction of random data caused a substantial change in position +5 to -3 for the statement, "Companies must take responsibility and pay for the pollution." Table 5 shows how the defining statements (those in the strongly agree and strongly disagree categories) changed with the introduction of random data in two of the most impacted factors, 4 and 5, illustrating how low-quality data can substantively influence the interpretation of these factors.

In Factor 4, when we examine the defining statements prior to the introduction of random data, there appears to be a focus on corporate responsibility and deep resistance to further utilization of fossil fuels, coupled with support for statements about the limits of renewable energy production and the threats posed by NIMBYism (Table 5). One interpretation of this factor involves attention to general levels of support for energy structures and energy production. Renewable energy is not enough, NIMBYism is a problem, energy structures can be beautiful, and high-energy consumption is part of the good life. These statements are consistent with the identification of an energy-focused economic and high consumption of energy.

However, the statement opposing further fossil fuel seems contradictory to this interpretation of Factor 4.

**Table 5. Comparison of Defining Sorts in Factors 4 and 5 with and without Random Data Included**

---

**Factor 4: Factor array**

---

*Defining statements without random data (n=105)*

- +5: Companies must take responsibility and pay for the pollution they produce.
- +4: Renewables cannot generate enough energy to significantly reduce greenhouse gases.
- +4: Fossil fuels should not be burned...they should just be left in the ground.
- 5: NIMBYism (Not In My Back Yard) poses a real threat to becoming a cleaner and greener society.
- 4: Energy structures, like transmission towers, hydroelectric dams or wind turbines, can be beautiful.
- 4: A high level of energy consumption is part of the good life.

*Defining statements with random data (n=130)*

- +5: Our national energy resources should be used in Canada for the benefits of Canadians.
  - +4: Energy systems are interconnected with all living system.
  - +4: Large facilities and concentrated production is the smartest way to provide energy.
  - 5: All forms of energy should be more expensive.
  - 4: No energy source is perfect; there are trade-offs between economic development and environmental protection.
  - 4: Nature will be fine no matter what humans do; it is a robust, self-correcting system.
- 

**Factor 5: Factor array**

---

*Defining statements without random data (n=105)*

- +5: Canada's energy resources make it a powerful global leader.
- +4: Our energy problems will be solved in the future through new technology and innovation.
- +4: Energy from dirty sources should be more expensive for consumers.
- 5: Compared to citizens of other countries, Canadians consume an obscene amount of energy.
- 4: Companies must take responsibility and pay for the pollution they produce.
- 4: People accept energy projects when there are benefits to the local community.

*Defining statements with random data (n=130)*

- +5: Canada's prosperity is based on ample supplies of affordable energy.
  - +4: Any energy source that produces close to zero carbon emissions (wind, solar, hydroelectric, nuclear) is urgently needed.
  - +4: We should have more energy choices at the household level.
  - 5: Small and distributed energy sources are more resilient than centralized production.
  - 4: Canadians have a duty to be global leaders by reducing our own energy consumption.
  - 4: Improved power grid technology will help us manage energy better.
- 

There may be ways to identify a nuanced and coherent interpretation of this factor that can withstand scrutiny, but more likely, our sample of participants (n=105) suffers from low-quality sorts strongly impacting the factor array and interpretation of results, particularly in relation to Factors 3, 4 and 5. When randomness was introduced to Factor 4, we observed total shifts in the defining statements, resulting in the strongest sentiments being nationalistic toward energy resource development (e.g., "Our national

energy resources should be used in Canada for the benefit of Canadians,” +5) coupled with disagreement about the trade-offs of energy production between the economy and environment, along with disagreement about the need for more expensive forms of energy (e.g., “All forms of energy should be more expensive,” -5). Again, there are contradictions inherent in these defining statements for Factors 4 and 5, and the influence of low-quality sorts makes these statements and the associated interpretation of the factors more precarious. In this case, instead of “forcing” an interpretation onto these factors, we concluded these minority factors were highly susceptible to poor-quality sorts: the defining statements were more likely a function of low-quality sorts than an indication of complex and nuanced thinking about energy systems in Canada.

The results of this analysis, which found that random data substantially changed the data structure of participant-derived factor arrays, were confirmed in a separate experiment involving the introduction of random sorts into two different p-sets that utilized traditional in-person methods. These separate studies (Parkins et al., 2015) drew on the same concourse of statements as the web-based Q study, allowing for a direct comparison between the online and in-person results. Ten random Q sorts were selected from the 25 randomly generated Q-sort dataset, and these random sorts were added to the participant-derived empirical datasets. Similar to the web-based dataset tests, when randomly generated Q sorts were introduced to each of these empirical datasets, loading patterns on the minority factors changed most substantially, with a high number of random sorts loading on the minority factors. More dominant factors were also affected, as factor loading values and defining sorts for nonrandom sorts were altered with the introduction of random sorts, thereby changing how the nonrandom sorts contributed to the definition (and interpretation) of each factor.

## **Discussion**

In this article, we introduced methods of exploring the reliability of Q methodology data obtained using online platforms and explored the potential impacts of low-quality sorts on factor analysis and interpretation. This focus on data reliability addresses a growing number of questions about how Q studies may be impacted by online data collection procedures. These new online procedures allow researchers to include more diverse participants from much larger geographies, resulting in greater distances between the researcher and their study participants. In order to recruit participants across these large geographies, online Q studies may also rely on participant recruitment methods that may meet desired demographic parameters but result in participants who are less motivated to thoughtfully and carefully complete the Q sort. The online platform also prevents participants from engaging with the researcher to seek clarification when confused or frustrated, which may ultimately impact data quality. Within this context, some researchers call for closer scrutiny of online data quality, suggesting that low-quality data produced by disengaged participants can be identified through proxies such as Q sort completion times (Pruneddu, 2015).

Although we are also suspicious of short completion times, our analysis suggests that time to completion may not be the best way to determine data reliability. Published research on web-based surveys shows that variables such as a participant’s age, level of computer experience, and confidence and/or knowledge of the subject matter can all influence how quickly web-based exercises are completed (Malhotra, 2008). Web-based Q study participants may become distracted with other tasks during the sorting, thus increasing their completion time. Alternatively, participants with poor computer skills may also take longer to complete the sorting exercise. In both of these instances, an

increase in time to completion may or may not be commensurate with the level of thought or engagement that the participants put into the sort. Further, short completion times may not necessarily imply low levels of participant engagement, particularly in instances where the participant is very knowledgeable about the subject matter and/or has very distinct and well-formed opinions about the topic. Given these uncertainties, we feel that a participant's completion time alone is not a reliable indicator for predicting participant engagement and agree with Malhotra (2008) that completion time should not be the sole factor in filtering out potential unwanted responses.

Using random sorts as an extreme example of low-quality data, results from this study demonstrate that even a small number of low-quality sorts can impact data structure. Given that factor analysis seeks to find commonality within data and that even randomly generated datasets can share similarities that are statistically significant, these results are not surprising. Furthermore, principal components analysis tends to be very sensitive to outliers and influential observations, which can be misleading and influence the conclusions drawn from the data (Shi, 1997). As such, our results should give Q researchers pause, particularly because many Q researchers will use as few as two loadings to define a minority factor. In these instances, researchers need to be absolutely certain that the sorts being used to define minority factors are reliable and are meaningfully contributing to the understanding of the subjectivity under examination. Low-quality sorts in this case are likely to introduce "noise" into the interpretation, resulting in erroneous insights. Moreover, as is demonstrated clearly in this article, this noise in the dataset cannot be overcome by increasing the number of study participants.

The impact of low-quality sorts can vary depending on the significance of the loading and the number of defining sorts within an individual factor. For example, with only a few random Q sorts contaminating a dataset, the interpretation of a dominant factor would only be minimally impacted, as the contribution of each Q sort to a factor array is weighted (Brown, 1980). However, as more randomly generated Q sorts populate a dataset, the impact on study results can become significant enough to cause concern. In tests introducing 10 or 25 randomly generated Q sorts to the datasets, randomly generated sorts crowded out participant-derived sorts that had previously loaded on minority factors. In other words, instead of loading together with random sorts, the nonrandom sorts tended to drop out of the minority factors (3, 4 and 5) and instead loaded with the dominant factors (1 and 2). The shifting of participant-derived sorts left the random sorts to define the minority factors, thus substantially changing the interpretation of all factors.

The crowding out observed in these tests represents a considerable concern, as this phenomenon can affect the interpretation of factors in several ways. First, when low-quality Q sorts crowd out reliable Q sorts, the defining sorts for the dominant factors change, thereby affecting how these majority factors are interpreted. Second, as higher numbers of low-quality sorts populate a dataset, a researcher's ability to reliably identify and interpret minority factors is severely compromised. The third manner in which low-quality Q sorts are likely to affect Q-methodology findings is by diminishing the ability of researchers to come to nuanced and/or accurate conclusions. One of the reasons behind the use of web-based platforms in Q studies is to explore the possibility of unique perspectives on an issue within a general population. In our case, we were interested in learning more about the prevalence of discourses on energy development in Canada. Research of this nature reveals some discourses that are fairly well understood and represented by dominant voices related to climate concern or market-

oriented logics. It is the minority views, often expressed through Factors 3, 4 and 5, that are of most interest, often reflecting more nuanced positions on a subject. Based on our analysis, it is these minority views that are most deeply impacted by low data quality. Q methodology is praised for its ability to provide detailed analyses of diverse points of view, and based on the results of this study, it is clear that low-quality Q sorts present a serious threat to the interpretation of results for both web-based and traditional Q studies.

Attention to the impact of low-quality sorts, as noted by Ramlo (2016b) and Pruneddu (2015), is imperative to ensure high quality and accurate findings. Toward this end, we suggest that the insertion of randomly generated sorts into a participant-derived dataset offers researchers a method for further examining the structure of the dataset to identify sorts that may be problematic. For example, as illustrated in Table 3, where Q sorts with shorter completion times load on a factor with only randomly generated sorts, the researcher should give further scrutiny to these sorts to ensure that these data are robust. Additionally, to promote data reliability, researchers utilizing web-based Q-sorting platforms should adequately screen participants to include only those participants who are likely to engage in the study at the level necessary to provide reliable data. One way in which this can be achieved is by developing and applying more stringent selection criteria.

Drawing on the critique that web-based Q sorts have increased the gap between researcher and participant (Brown, 2015), another potential method of reducing the probability of having low-quality Q sorts as part of a web-based Q study is to hybridize the sorting exercise by combining the qualitative and quantitative elements of conventional Q methodology procedures with web-based platforms. One manner in which hybridization can be achieved is to utilize a web-based platform to administer the sorting exercise followed by directly connecting with participants for a post-sort interview (Davis & Michelle, 2011). If in-person interviews are not feasible, telephone or online technologies (e.g., Skype) may be used. By connecting with the participant, researchers are able to probe for further explanations or deeper meanings. In doing so, researchers gain a greater sense of the participants' understanding and points of view on the subject matter. Connecting the researcher to the participants in this manner also provides the participant with an opportunity to discuss additional statements or themes that may have influenced how they completed the Q sort, which in turn provides the researcher with a better understanding of the rationale behind the participant's response. Further, knowing that a researcher is waiting to interact with them may also encourage participants to put forward higher levels of effort and become more engaged in the sorting process. If participants are not engaged and put little effort into sorting, researchers may be better equipped to identify low-quality sorts based on the participants' responses in the post-sorting interview.

By taking steps to reduce the likelihood of including participants who will produce low-quality sorts, in addition to reducing the gap between researcher and participant, researchers who typically rely on web-based platforms to administer Q studies may see changes in how they design their studies. For instance, the introduction of web or phone based post-Q sort interviews will likely restrict the number of participants included in a web-based study, due to the associated monetary and time costs of data collection and analysis. These costs, however, are likely to reflect the original intentions of Q methodology as a mix-method approach and will most certainly ensure more accurate analysis and interpretation of results.

## Conclusion

With the growing use of web-based Q-sort software, researchers are able to examine the views of larger and more geographically diverse groups of people. However, by utilizing web-based platforms, Q researchers may sacrifice the high level of researcher oversight and connection between the researcher and participant often found in traditional Q studies and, in doing so, increase the potential for low-quality Q sorts, which will pollute study datasets. Although some researchers recognize this problem and are taking steps to identify low-quality sorts in their dataset, the full impact of how low-quality Q sorts can potentially impact study findings has received little attention. In an effort to address this gap in the literature, this study has explored the relationship between participant engagement and data quality. Our findings illustrating the degree to which random data can potentially influence factor interpretation highlight the importance of maintaining a connection between the researcher and study participants and we call on the Q community to consider carefully how online sorting software is designed. Additional tools to help researchers assess the veracity of data collected using online platforms are needed.

## References

- Brown, S. R. (1980). *Political subjectivity: Application of Q methodology in political science*. New Haven, CT: Yale University Press.
- Brown, S. R. (1989). A feeling for the organism: Understanding and interpreting political objectivity. *Operant Subjectivity*, 12(3/4), 87-97.
- Brown, S. R. (March 7, 2015). Q method listserv communication.
- Davis, C. H., & Michelle, C. (2011). Q Methodology in audience research: Bridging the qualitative/quantitative 'divide'? *Participations: Journal of Audience and Reception Studies*, 8(2), 559-593.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3<sup>rd</sup> ed.). Hoboken, N.J.: John Wiley & Sons.
- Gallagher, K., & Porock, D. (2010). The use of interviews in Q methodology: Card content analysis. *Nursing Research*, 59(4), 295-300.
- Hempel, C., Parkins, J., & Beckley, T. (2015). Identifying energy discourses in Canada with Q methodology, UAL Dataverse, V1, University of Alberta. doi:10.7939/DVN/10635
- Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, 72(5), 914-934.
- Parkins, J. R., Hempel, C., Beckley, T. M., Stedman, R. C., & Sherren, K., (2015). Identifying energy discourses in Canada with Q methodology: Moving beyond the environment versus economy debates. *Environmental Sociology*, 1(4), 304-314.
- Pruneddu, A. (April 24, 2015). Q method listserv communication.
- Ramlo, S. (2016a). Q method listserv communication. April 26.
- Ramlo, S. (2016b). Mixed method lessons learned from 80 years of Q methodology. *Journal of Mixed Methods Research*, 10(1), 28-45.
- Reber, B. H., Kaufman, S. E., & Cropp, F. (2000). Assessing Q-Assessor: A validation study of computer-based Q sorts versus paper sorts. *Operant Subjectivity*, 23(4), 192-209.
- Schmolck, P., & Atkinson, J. (2014). PQMethod, version 2.35 (Computer software and manual). Retrieved June 2014 <http://schmolck.userweb.mwn.de/qmethod/>
- Shi, L. (1997). Local influence in principal components analysis. *Biometrika*, 84(1), 175-186.

Watts, S., & Stenner, P. (2012). *Doing Q methodological research: Theory, method and interpretation*. London: SAGE Publications.