Positive Reactions to a Q Sort for Personality Assessment

Chris D. Fluckinger
Michelle R. H. Brodke
Bowling Green State University

Abstract: *Experimental research comparing participant reactions to the Q- sort experience with other measurement methods is lacking. To address this gap in the literature, we collected data from 370 student participants who completed a 50-item personality instrument in both Q-sort and Likert formats. After completion of each instrument, participants provided ratings for a series of assessment reactions, including perceived control, liking of personality statements, and visual appeal, as well as qualitative feedback regarding their perceptions of the assessment experience. Results indicate that the Q sort was perceived more favorably regarding visual appeal. The Q sort was also rated as less limiting in allowing participants to express their personality. In addition, the qualitative feedback indicated a general preference for the Q sort. Taken together, these results provide insight into reasons behind participant reactions to the Q-sort procedure, particularly in relation to Likert measures.*

Introduction

This investigation focuses on participant reactions to the process of Q sorting. It is important to note that this is not a direct application of Q methodology to explore a specific area of study. Instead, this study is an exploration of how the experience of completing a Q sort is different from that of another common measurement method: in this case, a Likert rating scale. One key factor in the design, which also distinguishes this study from a more traditional application of Q methodology, is that the content (items) was held constant between the two methods, so that participant differences in reactions could be interpreted as due to the method as opposed to the content.

Q methodologists have argued that completing a Q sort results in meaningfully different information from study participants compared to information obtained from normative approaches (such as Likert scaling). Stenner (2011) observes that "A participant completing an

Contact author: cdfluck@bgsu.edu
Operant Subjectivity: The International Journal of Q Methodology, 2013, 36(4): 335-341

attitude scale or personality measure is, from the psychologist's point of view, a 'subject' in the full sense of being passively subjected to measurement" (p. 200). Stenner contrasts this perspective with the Q-sort experience: "Participants are not passive subjects but *genuinely active participants* who operate on a set of items from an explicitly self-referential (what *I* believe from *my* perspective) point of view" (p. 201). In essence, completion of a Q sort involves different levels of engagement in assigning numbers to items as well as different beliefs about the extent to which the measure as a whole captures unique or individual aspects of the person. In other words, completing a Likert-based personality assessment involves the simple process of providing ratings for a series of statements. However, when completing a Q sort, participants can see a holistic portrait of their personalities as they actively create it during the assessment.

Few studies that investigate assessment reactions directly compare methods. One exception is McKeown (2001), which found that Likert scales produced more extreme scores than Q sorts. However, participant judgments and reactions that could account for these differences were not directly assessed. Additional work revealing how participants react to completing a Q sort compared to a normative scale can provide insight about the mechanisms through which Q sorts reveal the diversity of meanings participants attribute to a set of items. The present work seeks to address this need.

Research Design

Participants were university students recruited from undergraduate psychology classes to take part in a study, described as involving personality assessment methods in a job-application context, in exchange for extra credit. In job applications, personality measures are commonly deployed. Data collection occurred in a campus computer lab with a lab proctor present to explain study procedures and answer questions. Assessment order was counterbalanced, so that half of the participants experienced the Q sort first and half experienced the Likert first. The Q sort was performed with FlashQ software (Hackert & Braehle, 2006). For the Likert measure, we employed the SurveyMonkey (surveymonkey.com) platform, which is commonly used in social science data collection. The assessments were calibrated to look as similar as possible in terms of color schemes, font sizes, and other aesthetics.

Participants completed the assessments under simulated-applicant instructions, which required them to imagine that they were completing the assessments as a hurdle in applying for a job. Upon entry into the lab, each participant was provided with a packet including informed consent and instructions specific to the simulated job-applicant context. Participants were seated at computers, and the necessary web links to

each personality measure were provided in the packet in one or the other order for completion. Assessment reactions (both ratings and qualitative feedback) were recorded directly after completion of each measure.

Six assessment reaction items, which can be viewed in Table 1 (next page), were included to investigate direct features of the measure that might be influenced by the method of assessment. Quantitative reactions were measured with a 5-point Likert scale (*strongly disagree*, *strongly agree* anchors). We acknowledge the initial irony of using a Likert scale as evidence of between-method differences; however, our goal is to not to paint the Likert as an invalid measurement method, even if our results indicate lower user favorability to the technique in the simulated job-application setting. The Likert scale was deemed the clearest way to quantify the feedback as an initial exploration of Q-sort reactions, although an alternative method could be an avenue for future research. Qualitative feedback was also requested. After completing a measure and responding to the reaction items for that assessment, participants were encouraged by the lab proctor, and provided with space in the packet, to describe their experience with the assessment.

The personality items included in this study were taken from the 50-item International Personality Item Pool (ipip.ori.org), which is often used in research involving the 'Big 5' factor model of personality, which covers openness to experience, conscientiousness, extraversion, agreeableness, and emotional stability (Lim & Ployhart, 2006). To accommodate these 50 items, the Q distribution was constructed in a quasi-normal fashion of 3, 5, 6, 7, 8, 7, 6, 5 and 3 boxes in 9 columns.

For the analysis strategy, we chose to run a 2 x 2 x 6 mixed-model multivariate analysis of variance (MANOVA), which allows for inspection of potential order effects (Q sort first or Likert first). In this model, there is one between-subjects factor (order of assessment) and two within-subjects factors: target of rating (Likert or Q sort) and type of reaction (6 assessment reactions items). This analysis allows for interpretation of individual factors while controlling for the effect of the other factors, as well as interaction effects. This analysis is based on previous experimental designs that required mixed-model MANOVAs (Szeto, Straker, & O'Sullivan, 2005). We were most interested in the target of rating x type of reaction interaction. If significant, we then interpreted individual reactions via paired-samples *t*-tests ($p < .05$).

Results

Quantitative Findings

The MANOVA produced two main effects: target of rating (Wilks' lambda = .95; $p < .01$; $\eta^2 = .02$) and type of reaction (Wilks' lambda = .77; $p < .01$; $\eta^2 = .23$), indicating that the Likert and Q-sort means (when aggregated

across reactions items) differed, and that mean assessment reactions differed when aggregating across assessment types, respectively. Order of assessment did not interact with either of these variables. However, the hypothesized two-way interaction was significant and produced a large effect size (Wilks' lambda = .53; $p < .01$; $\eta^2 = .47$), indicating that the size of differences between the reactions was dependent on whether the rating was for the Q sort or Likert. We also note the presence of a three-way interaction with a small effect size (Wilks' lambda = .96; $p < .05$; $\eta^2 = .04$). To understand this interaction, six 2 x 2, mixed-model ANOVAs (1 for each reaction) were performed with order of assessment as the between-subjects factor and differences between a given reaction for the Likert and Q sort as the within-subjects factor. In one case, the interaction was significant (*This assessment provides enough control to choose responses*), as participants viewed the Likert measure as providing more control to choose responses than the Q sort ($M = 3.98$ and $M = 3.56$ for Likert and Q sort, respectively), but only when the Likert is presented first. When the Q sort is presented first, order did not appear to matter ($M = 3.66$ and $M = 3.57$ for Likert and Q sort, respectively). Given the small effect size for the three-way interaction, and that it is driven by one reaction, we next interpret main effects for the other reactions without regard to order of assessment.

Table 1: Within-subjects Tests for Reactions Items

Variable	*Likert*		*Q sort*			
	M	*SD*	*M*	*SD*	*t*	*d*
The assessment limits my ability to express my personality	3.29	1.17	3.04	1.03	-3.74**	.20
I understand how this assessment would be scored	3.36	1.15	3.22	1.05	-2.31*	.12
This assessment took a reasonable amount of time	3.49	1.24	3.55	.99	1.02	—
I liked the personality statements in the assessment	3.56	.92	3.66	.79	2.16*	.11
This assessment is visually appealing	2.72	1.01	3.91	.89	11.56**	.97

Note: * p < .05, ** p < .01, N = 370

Results for main effects of the reactions items can be found in Table 1. Of the items, four of five show differences between the assessments, with the exception of reasonableness of time to complete the assessment. Effect sizes are reported for each reaction that produced a significant *t*-statistic. The most prominent finding is that the Q sort was viewed as more visually appealing than the Likert measure (a large effect; $d = .97$), despite the fact that care was taken to reduce superficial differences between the measures through similar, use of color and graphics. Another finding involved participants indicating that the Likert

measure limited ability to express personality (a small effect; $d = .20$).

Two additional findings involving assessment features are worth noting. First, there were no differences in perceptions about assessments taking a reasonable amount of time, even though respondents took approximately 10 more minutes on average to complete the Q sort. This indicates that given the other perceived advantages of the Q sort, participants do not mind the fact that it can take longer to administer. Second, participants liked the personality items slightly more when administered via Q sort. This finding is notable, because each assessment included identical items.

Qualitative Findings

The qualitative feedback provided a rich source of information involving reactions to the two assessments. Counts were performed by the first author, and questionable cases were resolved with the second author when necessary. Generally, participants were more likely to provide feedback for the Q sort (132 comments provided for the Q sort and 82 for the Likert), and the feedback provided was typically more detailed and specific for the Q sort ($M = 22$ words per comment for the Q sort and $M = 13$ words per comment for the Likert). In fact, a substantial portion of the feedback for the Likert focused on general opinions regarding personality assessment (46 comments for the Likert and 21 for the Q sort) and less on the process of choosing responses (whereas the Q-sort feedback included 63 mentions of the relative ordering of responses, for example). In sifting through this feedback, we observed general patterns and themes. Below, we present the three most prominent themes.

Theme 1: Q sort was more engaging and/or enjoyable. The clearest theme to emerge is that when participants contrasted the two measures, most showed a preference for the Q sort. Of the 45 comments that directly compared measures, 38 preferred the Q sort. Reasons provided for this preference often focused on the interactive nature of the Q sort, which seemed to create an increased amount of engagement, as well as increased enjoyment when completing the Q sort. Conversely, 32 comments specifically described the Likert assessment as "boring." Representative comments:

- It was different. It kept my attention because it made me think and I wasn't mindlessly filling in bubbles.
- Loved it! Nice idea.
- I think it's a fun thing to do; it reminds me of Facebook quizzes.
- It was a lot more interactive and fun.
- I thought it was good because it was interesting and set up differently compared to others. Different is good!
- Drop/drag format was more engaging.

Theme 2: Q sort was more difficult and/or restrictive. Some participants (38) focused on the difficulty or restrictiveness introduced by the Q sort. Participants who found the Q sort to be difficult also were more likely to comment about the restrictiveness of the Q-sort procedure. In short, the general preference for Q sort found in this study does not mean that this preference was shared by all participants, as some sample comments show:

- [It] was a bit confusing with all the shuffling of boxes
- I found that if you have too many items that are "not like me" you cannot fit them on that side of the pyramid.
- This assessment did not allow enough flexibility in answers
- I feel put into a box with this assessment.
- This one was more difficult.
- I wish there were an unlimited amount of rows.

Theme 3: Skepticism about applied personality measurement. Although this theme does not directly address differences between the two assessments, it is worth mentioning that approximately half of participants who provided feedback indicated some degree of skepticism about the use of personality measures in the job applicant context. In fact, some participants were quite vehement in their objections to use of such measures. This should not be taken as an indictment of either method, but rather a caution to remember that positive reactions to the measure itself do not necessarily indicate positive reactions to use of the measure to make decisions about individuals. Comments on this theme included:

- Does not fully assess a person's skills, abilities or personality very well. Would give an employer only a vague idea of the type of person they will be hiring or interviewing.
- Assessments like this are meaningless.
- A person can answer most of the questions just to look good for the job. I don't believe job performance in the future could be measured with just this test.
- People can fake their personalities and get the job even though they aren't fit for the job.

Conclusion

This is the first study (at least to our knowledge) to have Big 5 personality items placed into a Q sort and to investigate participant reactions to a Q-sort measure in comparison with the usual Likert administration of the personality test. The results provide a number of interesting themes. Perhaps most importantly, the finding that a Q-sort measure containing identical personality items as a Likert measure was

generally viewed more positively by participants is useful to researchers and practitioners who are concerned about user favorability of assessments. These results indicate that variables such as assessment format may in part be accountable for low favorability of personality measures (Hausknecht, Day & Thomas, 2004).

This study makes important contributions to both the Q-methodology and personality assessment-reactions literatures. First, evidence that a Q sort generates more favorable reactions than a Likert measure is useful to scientists and practitioners who are concerned about negative reactions to personality and other non-cognitive assessments. Second, this study indicates that specific reactions (such as to visual appeal and ability to express oneself) are particularly important in driving the general participant preference for Q sort over Likert. Taken together, the results support the use of Q sorts in situations where participant reactions to the assessment are deemed important.

References

Hackert, C., & Braehle, G. (2006). FlashQ (computer software). Available from http://www.hackert.biz/flashq/home/.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57,* 639–683.

Lim, B., & Ployhart, R. E. (2006). Assessing the convergent and discriminant validity of Goldberg's International Personality Item Pool. *Organizational Research Methods, 9,* 29–54.

McKeown, B. (2001). Loss of meaning in Likert scaling: A note on the Q methodological alternative. *Operant Subjectivity, 24,* 201–206.

Stenner, P. (2011). Q methodology as qualiquantology. *Operant Subjectivity: The International Journal of Q methodology, 34,* 192–203.

Szeto, G. P. Y., Straker, L. M., & O'Sullivan, P. B. (2005). EMG median frequency changes in the neck-shoulder stabilizers of symptomatic office workers when challenged by different physical stressors. *Journal of Electromyography and Kinesiology, 15,* 544–555.