Random Qualitative Validation: a Mixed-Methods Approach to Survey Validation

Eric Van Duzer

2/1/2012

Random Qualitative Validation: a Cross-sectional Approach to Survey Validation

The purpose of this paper is to introduce the process and value of Random Qualitative Validation (RQV) in the development and interpretation of survey data. RQV is a method of gathering clarifying qualitative data that improves the validity of the quantitative analysis. This paper is concerned with validity in relation to the participants' understanding of the survey author's intent, and the ability to accurately interpret the quantitative response on forced-choice scales. A campus study of rigor in general education courses conducted in 2006/07 provides examples and will be used to demonstrate the benefits and challenges of the RQV approach.

Currently, a variety of methods are used to improve the interpretive validity of survey data including quantitative techniques to establish convergent and discriminate qualities of survey scales. Assessments of internal consistency such as the use of Chronbach's Alpha and methods for examining psychometric properties such as Factor Analysis are valuable tools, but fall short of addressing the question of participant interpretation of survey items. In discussing his work on the National Center for Higher Education Management Systems surveys Dillman (2000) writes

"Psychometric analyses showed that most items are reliable. However, these kinds of analysis could not inform us whether students from different institutions interpreted the items consistently or if their responses accurately represented the behaviors or perceptions the survey designers intended. That is, even when psychometric indicators

are acceptable, students may be interpreting some items in disparate or unintended ways."

Unless a method is used to ensure the adequacy of the communication, construct irrelevant data can undermine the validity of the data interpretation. In order to better interpret the survey data, some form of qualitative data collection is usually a component of survey development procedures (Pressor, S., et al., 2004). Checks for face validity using expert review, focus groups, one-on-one interviews, cognitive interviewing and pilot studies with immediate debriefing are used to investigate how the respondents interpret the meaning of the survey items and what meaning they are trying to communicate in their responses (e.g., Huichcock, J.H., et al., 2005). However, the typically small scale of these techniques generally makes it impossible to ensure the representativeness of the sample leading to a greater probability of error in generalizing the results of these small scale pilot tests to interpret the full survey data, particularly with diverse populations. Therefore, an alternative method to gather qualitative data from a broad cross-section of respondents was needed to augment current practices and improve interpretive validity.

In attempting to address the need for representative data regarding the interpretation of questions and responses, the Random Qualitative Validation (RQV) process was develop and piloted in the development of a common instrument to assess student's perceptions of rigor in university General Education courses. The survey was used across disciplines and class levels requiring a method for understanding the students' interpretation of the survey in very different settings.

Random Qualitative Validation

The random qualitative validation (RQV) process described below provides cross-sectional qualitative data targeted to help interpret the quantitative data. The RQV process involves using multiple versions of the same survey with each version asking respondents to explain their ratings on a small number of items (see Appendix A). For example, if a survey instrument consists of 30 items, ten versions of the survey could be created, each asking for an explanation of three different ratings. Version one may ask the respondent to "please explain why you gave this item the rating that you did" for items #1, #11, and #21, while the second version would ask for an explanation of ratings #2, #12, and #22. In this fashion, respondent fatigue, often identified with open-ended surveys requiring many written answers, does not occur and does not result in significant loss of cooperation or missing data. Prior to distribution, the ten versions are collated resulting in data collection method sufficiently similar to a systematic random sample to accept as representative (any class/group with ten or more students will have at least one respondent explanation for each item on the survey).

In the GE Rigor survey pilot study introduced above, 591 respondents distributed across upper and lower division courses in science, humanities, art and English had a response rate between 85% and 92% for the RQV writing prompts. An average of 52 comments for each survey item was collected. There were very few random or non-responsive comments such as "I hate surveys."   These non-responsive comments were not included in the response rate reported above.

The comments provide a relatively quick method of identifying unintended interpretations of what is being asked as well as improving the researchers' ability to interpret what the respondents' numerical rating is intended to convey. The RQV written comments tend to be short and directly targeted to the interpretation of the quantitative scores. This provides an efficient source of information for quickly evaluating the respondents' interpretation and intent, and where used as part of the process of instrument development, RQV provides an effective mechanism for improving the survey items to increase the clarity of communication with respondents.

The following comments are illustrative of the type of data collected with the RQV method:

1. This class requires a lot of effort on my part because it is a whole different language for me. Arts is something I appreciate but not something that I have used in order to express myself.

2. This class involves more work on one particular thing rather than a broad overview, so we go more in-depth in our writing.

3. This is not a science – there is no memorization – the effort is another kind more applicable to your life

In the RQV process, the comments are collected for each survey item and then grouped into content categories. This can be a rather informal process that does not require formal coding procedures given the specificity and brevity of the individual comments. In the GE Rigor survey several items were organized into categories independently by multiple researchers and

then compared demonstrating a high level of agreement between on the major categories and intent of the comments. This process of evaluating the comments can be very efficient and informative. In the case of the GE Rigor survey, three survey items were identified as having little commonality across disciplines making the analysis difficult if not impossible. As a result those items were removed or completely rewritten. Eleven other items were revised to some degree.

Many of the needed changes to wording would have been difficult to identify from a quantitative analysis. For example the item that asked students to rate the extent to which the course required them to "critically examine personal, disciplinary, or social values" resulted in two groupings of comments. About three quarters of the comments reflected the intended interpretation involving an exploration of personal/professional values and their application. However, approximately one quarter of the comments fixated on the word "critically" and reflected the degree to which the course required them to think critically, an issue addressed in a separate item on the survey. The result of this analysis was to remove the word "critically" and simply ask respondents to what extent the course required them to "examine personal, disciplinary, or social values." Similarly, one item asked students to rate how many oral presentations were required in the course. The qualitative comments revealed that students included participation in class discussions and other informal communications in their ratings. The item was revised to ask about the number of "formal" presentations required.

A Not applicable (N/A) option is often employed as one position on the response scales. The N/A response is intended to provide a legitimate "out" for respondents who feel that the item does not apply to them. However, using the RQV process with the GE Rigor survey

provided a unique insight that revealed instances where the respondent had tried to "force fit" their response to an item thus making interpretation of the ratings difficult. To address this problem, a variety of modifications were made including providing examples to narrow the range of likely interpretations; For example, "To what extent did the course require you to develop skills/specialized techniques (e.g., drafting, painting)." In some cases, providing a reminder to use the N/A option where appropriate as in the following example, In a typical week, how many hours do you spend completing quantitative problem-sets? (*If not that type of class mark **N/A***).

The RQV results can identify when researchers and respondents do not share a common understanding of language. For instance, an item asking how frequently the course required students to "use scholarly material for class activities/projects" was answered by some students as being very frequent since they used their *textbooks* every week. In writing the item, the author's intention was to determine if students were reading and analyzing research papers or other scholarly documents/ materials. In a quantitative analysis it might have been difficult to uncover this misinterpretation (particularly if the instrument lacked other similar items or an overall scale on the use of research resources to compare it with), whereas with the RQV process it was relatively easy to identify this as a problem requiring a wording change on the survey.

In addition to revealing problems with wording and directions, the RQV process can provide explanatory data. An example from the GE Rigor survey relates to an investigation of the concern by some faculty that students majoring in science felt that GE was useless in their education. Two items were relevant to investigating this concern. The items were part of a

semantic differential scale designed to assess the students' attitudes about the course. They were:

INSERT TABLE 1

The qualitative data provided important insight into the meaning of the students' ratings. On the item measuring how meaningful the content of the course was, respondents routinely addressed issues of being prepared as a well educated citizen in the world with comments such as "The subject of the class really relates to everyone. It is nice to see how it relates to you as a person in society as a whole. It is important to be informed on race as it Impacts everyone and it is so interesting to see how."

However, in responding to the second item asking how valuable they felt the content was as a part of their education, students responded with comments such as "While I find this course incredibly interesting it is of little relevance to my education/professional goals." There was such a strong correlation between the rating on this item and an item asking if the GE course satisfied requirements for their major, that the former was removed from the final survey altogether. If a course did not satisfy a major requirement, regardless of how interesting or meaningful it was, the item received a low rating on the issue of value to one's education. This explanatory data provided an understanding of the students' ratings that would have been simply confusing without it.

In this paper I have described the concept of RQV, the use of several versions of a survey instrument to implement RQV, and provided examples of the types of information gained and

utility of the process. By providing a mechanism for cross-sectional, random collection of

qualitative data explaining survey ratings while limiting the demands on individual respondents'

time and effort, Random Qualitative Validation adds a useful tool for evaluating the degree to

which respondents understand survey items as intended by the author, and sheds light on the

meaning of the respondents' ratings that improve the ability to make valid interpretations of

the data.

References

Dillman, D. (2000). Mail and Internet Surveys: The Tailored Design Method. J. Wiley. New York.

Huichcock, J. H., Nastasi, B.K., Dai, D.Y., Newman, J., Jayasena, A., Bernstein-Moore, R., Sarkar,

S., & Varjas, K.. (2005). Illustrating a mixed method approach for validating culturally

specific constructs. *Journal of School Psychology, v43, n3*. Pgs 259-278

Presser, S., Couper, M.P., Lesser, J.T., Martin, E., Rothgeb, & J.M., Singer, E. (2004). Methods

for Testing and Evaluating Survey Questions. *Public opinion Quarterly, v68, n1*.

Appendix A. Sample Survey Page

I. This course requires me to:	Every week	Most weeks	Some weeks	Never	Not Applicable
1. Think critically/analytically about the content of the class.	1	2	3	4	5
2. Memorize and repeat basic facts from the readings or lecture	1	2	3	4	5

Please briefly explain why you rated question # 2 the way you did:

	Every week	Most weeks	Some weeks	Never	Not Applicable
3. Apply the ideas/content in the course to solve problems or make arguments	1	2	3	4	5
4. Critically examine personal, disciplinary or social values	1	2	3	4	5
5. Understand the underlying theories or connections between ideas in the class	1	2	3	4	5
6. Synthesize ideas, information or experiences into new, more complex interpretations	1	2	3	4	5

I

II. Compared to other <u>General Education</u> courses that you have taken at the same level (100 level, etc.) this class:	Agree	Somewhat agree	Somewhat disagree	Disagree	Not Applicable
7. Is more intellectually challenging	1	2	3	4	5
8. Requires more effort to master the material	1	2	3	4	5
9. Requires more critical thinking	1	2	3	4	5
10. Requires more careful attention to details	1	2	3	4	5
11. Requires a deeper exploration of my creative abilities	1	2	3	4	5
12. Requires me to more critically examine my assumptions, beliefs, or biases	1	2	3	4	5

III. What are the preparation requirements for this class?	Agree	Somewhat agree	Somewhat disagree	Disagree	Not Applicable
13. I can get A's or B's on quizzes, exams, or other forms of evaluation even without completing all of the coursework	1	2	3	4	5

Please briefly explain why you rated question # 13 the way you did:

Table 1

Overall the <u>content</u> in this class is:

	Meaningful	1	2	3	4	5	Meaningless
Valuable addition to my education		1	2	3	4	5	Useless addition to my education