Who Is Being Heard? Response Bias in Open-ended Responses in a Large Government Employee Survey

Mark Andrews

ORC Macro International

mark.w.andrews@orcmacro.com

Abstract

A great deal of research has been conducted in order to describe and measure both unit and item non-response bias on quantitative surveys. Relatively less work, however, has been done to adapt this work to open ended responses. This lack of attention is unfortunate because the topic is relevant for a number of reasons. First, while some purists treat verbatim comments as purely qualitative data, most researchers go through great lengths to code and tabulate responses for analysis. If a large bias is detected, this method of quantifying qualitative data will be brought to question. Second, this topic is important because there are reasons to believe that item non-response bias is greater for open ended comments. Open ended comments have traditionally low response rates. Furthermore, there are also obvious education, language, or other demographic issues that may have an effect on the likelihood and quality of response. Finally, open ended responses have an added dimension—response quality. Item non-response, therefore, is not the only source of bias in verbatim comments.

This paper will examine the issue of item non-response bias and response quality bias in a large (13000+ respondents) government employee survey. This survey provides a substantial quantity of demographic data including income, race/ethnicity, gender, age, education, and whether the employee is in a supervisory position. To measure item non-response bias, we develop a logistic regression model to predict response likelihood to the one open ended question in the survey. For the second stage, we will use length of response as a proxy for response quality. Again we will compare means and develop a regression model to predict response length using available demographics.

1. Introduction

Open ended responses are an invaluable source of information for the researcher. These comments often paint pictures that are more vivid than can be achieved through more traditional survey questions and have the potential to uncover areas that were overlooked in the survey design or can provide details that help unravel the complex relationships between more quantifiable variables.

Researchers use open ended response in a number of ways. Some take a very subjective approach; evaluating each comment, assessing its importance to the organization or the subject matter in hand, and searching for quotes that demonstrates an important theme, supports a hypothesis, or provides valuable insights about the data. Others are more methodical implementing complex coding systems in an attempt to integrate qualitative responses with quantitative data.

Much attention has been given to item non-response of quantitative survey questions. This is a particularly important issue when sensitive questions are asked or where the question is a necessary variable for key analysis or weighting. To solve this problem, researchers have developed logarithms to replace missing values or weight the data to adjust for item non-response.

For a typical question, however, the potential for item non-response bias is usually quite low. For the survey used in this paper the item non-response rate was just under 2%—a figure that is probably higher than one would expect from a shorter survey. In comparison, non-response for open ended questions is comparatively high. For this survey non-response rates for the two open ended questions were 41% and 76% percent. With such high levels of non-response the chance for bias is high.

It is the objective of this paper to explore these potential biases not only by examining item non-response, but also by exploring biases relating to response quality. Response quality for the sake of this paper will be measured by the length of the responses provided. Response length is certainly not a perfect measure for response

quality. It is easy to think of a hypothetical example where a concise response would be more meaningful. But, short of subjectively rating responses for quality, we have to assume that response length is an accurate proxy for quality.

The potential sources of item no-response bias are many. Anyone who has analyzed employee or customer satisfaction surveys can attest to the fact that dissatisfied customers are more likely to respond to open ended questions. Open ended responses are often used by respondents as a forum to vent their frustrations. It should be no surprise that this paper confirms this fact, indicating that dissatisfaction is one of the strongest predictor of whether a subject responds or not to on open ended question.

Another obvious source of bias is education level. One could expect that those with lower levels of education or limited English skills would be less likely to respond to open ended questions. Other biases might be less obvious. How about the respondent's gender? Age? Race? Level in the organization?

The potential for item non-response and response length bias is particularly problematic for analysis that attempts to quantify the qualitative results through coding. The problem with coding strategies is that they give the researcher a false sense of accuracy. The result of the coding produces an item that looks like and can be analyzed like any other survey question, but in reality it isn't. In addition to a sharp increase in the potential for item non-response bias, there is also bias introduced through differences in response quality as well as bias introduced by the coding process itself. It is very important to be aware of these biases when interpreting the results.

This is not to say that these issues are not of concern for more subjective qualitative analysis. While a researcher conducting analysis has the ability to judge the relative importance of individual comments, that importance is often influenced by a mental or physical tally of the frequency of similar comments. If important comments are under-represented or never written due to item non-response or the inability for the respondent to articulate it, they may be overlooked. Researchers should be aware of the potential for bias and avoid the misconception that qualitative research is somehow removed from the stringent constrains of quantitative research.

It is the objective of this paper to explore these potential biases not only by examining item non-response, but to explore biases relating to response quality. After a brief analysis of response rates and average response length, this paper will use regression analysis (both logistic and least squared) to adjust for the dominant demographic elements.

2. The Study

This study analyzes the open ended responses of a large federal government organization. The employee climate survey was administered by ORC Macro in the fall of 2004 to the entire population of 22,452 employees as both a paper and web survey. The effort yielded 13,793 responses for a response rate of 61.4%.

The survey was a long survey containing 125 questions, including 110 employee climate questions, 13 demographic questions, and 2 open ended questions. The survey was not forced response. The two open ended questions were asked toward the end of the survey prior to the demographics section. The first question (What is the single most important thing [the respondents division] could do to improve?) was followed by a space for more general comments (General: Please use this space to provide any additional comments you may have).

Roughly half (59.4 %) of the respondents responded to one or both of the open ended questions. A small portion of this non-response is likely due to incomplete surveys. Of those that answered the question prior to the open ended question, 60.1% answered one or more open ended question. Since the demographic questions were asked after the open ended questions, this small discrepancy is not likely to affect the results here. Only when we analyze item non-response solely by level of satisfaction (the first question on the survey) is it likely that survey fatigue becomes an issue.

It is the purpose of this paper to not only examine item non-response but to examine response quality. For the purpose of this paper we will only use a proxy for quality—number of characters entered .   The number of characters was counted for each question and aggregated in one score. The mean length was 261.2 characters

for respondents who replied to one or both open ended questions. Response lengths ranged from a handful to nearly 4000 characters (roughly 800 words). Response rates and mean response lengths were calculated for a number of demographic segments of the population. The results begin to reveal how certain groups are under-represented in the qualitative data collection.

2.1 Overall Satisfaction

Table I reveals a clear inverse relationship between the respondent's level of satisfaction and both whether they responded to the open ended questions and the length of their response. According to these figures, very dissatisfied respondents are 1.28 times more likely to respond to either or both open ended question and provide responses that are on average 66.7 characters longer than very satisfied respondents..

2.2 Education

These results also show a similarly clear relationship between a respondent's educational level and both item response and response length. Those with bachelor's degree or higher are 1.46 times as likely to respond to open ended questions than those with a high school education or less and provide responses that are on average 181.5 characters shorter.

2.3 Age and Tenure

These preliminary results also indicate a number of other potential biases. Item response rates fall as the number of years with the organization increases, as does the length of the responses given. This finding is consistent with the finding that older employees, in particular those over 60 years of age, are less likely to respond and on average write less when they do. Those over 60 are 1.2 times likely to respond and typically leave responses 37.2 characters shorter than other respondents.

2.4 Gender, Race, & Ethnicity

White respondents are only marginally more likely to respond to the open ended questions than non-whites but on average write 52.2 more characters in there responses. African Americans are likely to respond as other respondents, but on average write 80.4 less characters when they do.

With a 49.8% response rate, Asian & Pacific Islanders are the least likely to respond and on average provide 64.0 less characters in their responses. Interestingly, we would expect to see relatively lower response rates for Hispanic respondent due to language proficiency issues, but we do not. Hispanic respondents are just as likely to respond as the general population. They do, however, leave responses that are on average 52.6 characters shorter.

Differences in item non-response by gender are small. Women are as likely to respond as men, but write more when they do reply—providing on average an additional 52.1 characters. While women are only 43.2% of respondents, over half of the responses over 400 characters were provided by woman.

2.5 Level in Organization

Lower level employees (GS 1-8) are less likely than other employees to respond to open ended questions and write substantially less when they do. In contrast, senior level professional employees (GS 13+) are more likely to respond than other GS grades and on average write significantly more. Interestingly, supervisors while more likely to respond to the open ended questions but do not write significantly more than non-supervisors. It is therefore likely that supervisors outside or the GS pay scale are driving these lower than expected numbers.

2.6 Survey Mode

Not surprisingly a mode effect was also detected. As would be expected, subjects that responded to paper surveys wrote less—on average 159 characters, or 52.3% less. Surprisingly, this same group was only slightly less likely to respond to one or more open ended question. It is important however to take these results lightly because the mode of the cases was not determined randomly, a selection bias may be responsible for some of this variation. Those who replied by paper were more likely to be older, male, and from a lower pay scale and educational background.

3. Regression Analysis

Least squared and logistic regression analysis was conducted to test for bias in the rate and length of the response. The regression analysis allows us to measure the strength of the relationship while controlling for the dominant

predictors of item non-response and response length (satisfaction, education, and survey mode).

It is important to note that the regression analysis in no way negates the preliminary analysis of item-response rates and means. If our concern is achieving an appropriate racial mix of qualitative data, for example, there is little need to go beyond the preliminary analysis. The regression analysis, however, allows us to dig deeper into the data to help us understand how different demographics interact and affect item non-response and response length.

The categorical nature of the data necessitated the creation of dummy variables. With the exception of the satisfaction variable, the largest group was selected as the base variable. For satisfaction the "neutral" response was chosen as the reference point. In order to preserve a progressively increasing nature of the GS-scale variables, employees not on the GS scale were removed from the model. Non-GS-scale employees represent 29.2% of the respondents and are much less likely to respond and leave shorter responses.

4. Likelihood To Respond

Binomial logistic regression analysis was conducted to determine what demographic variables best predict whether a respondent answered one or both of the open ended questions. We ran a number of models using response to at least one open ended question as the dependent variable. Due to their consistent strength as predictors of item non-response in open ended questions, overall satisfaction and education were included in all models.

4.1 Overall Satisfaction

Overall satisfaction is the strongest predictor of item non-response to open ended questions with respondents who are least satisfied being more likely to respond. This relationship proves significant in all the models and across all levels of satisfaction. In all models, very dissatisfied employees are nearly twice as likely to respond to one or more open ended question as those who are neither dissatisfied nor satisfied and are roughly 2.4 times as likely to respond as very satisfied respondents. Likewise, those who are dissatisfied are roughly 1.5 times as likely to respond as neutral respondents and 1.9 times

more likely as compared to "very satisfied" respondents. These figures show a much stronger relationship between education and non-response than the preliminary analysis revealed thus indicating that overall satisfaction plays an even more important role when education is controlled for.

4.2 Education

The level of education is also a consistently strong predictor of response to open ended questions. In particular, those with lower education (some college or technical training, high school graduate or equivalent, & less than high school) are less likely to respond. This relationship remained robust in every model. Those with a four year college degree are roughly 2.2 times as likely to respond as those with only a high school education. This is considerably higher than preliminary analysis of item response rates indicated.

Despite early indications that the relationship is strong across all education categories, associates degree respondents are not significantly more or less likely to respond than bachelor's degree respondents. Advanced degree respondents, however, are significantly more likely to respond in 7 of the 10 models presented here—1.1 to 1.2 times as likely depending on the model.

4.3 Gender, Race, & Ethnicity

Women, African American, and Hispanic respondents are not significantly more or less likely to respond to open ended comments when adjusted for satisfaction and level of education attained. Non-Asian/Pacific Island respondents however are 1.4 times as likely as those of Asian/ Pacific Island decent to respond to open ended questions—a significant difference. Interestingly, those choosing "American Indian or Alaskan Native" were the most likely to respond—1.38 times as likely as whites.

4.4 Age and Tenure

Older employees are less likely to respond to open ended questions. This is particularly the case for those over 60, who, with a likelihood ration of 0.65, are less likely to respond than those between 40 and 50 years of age—a difference lower than in preliminary analysis. Those between 40 and 50 years of age are the

most likely to respond. Interestingly, tenure is not a good predictor of item non-response.

4.5 Level in Organization

Lower level employees (GS 1-8) are significantly less likely to respond than junior professional employees (GS 9-12)—a 0.87 likelihood ratio. More senior employees are (GS 13-14) are 1.2 times more likely to respond. Surprisingly, those who indicate that they supervise other employees are less likely to respond to open ended questions—a likelihood ration of 0.83. This result is contrary to preliminary analysis that shows that supervisors were 1.1 times more likely to respond than non-supervisors—a finding that is understandable when we take into account that in this model we are adjusting for education. It is also worth noting that this relationship may be driven by the more than 4000 respondents who are not included in the GS-Level analysis.

5. Response Length

Least squared linear regression analysis was conducted to determine what characteristics of the respondents best predicts the length of the open ended responses that they provided. As in the previous analysis, level of education attained and overall satisfaction were included in all models. In addition, we also included a variable for the mode of the survey.

The regression models presented below only explain a small portion of the variation in response length—roughly 6%-7% of total variance with one exception (GS scale) which was lower due in part due to a smaller sample. The limited predictive power of the models is understandable considering the great variation in response length that is skewed towards the high end of 4000 characters. To adjust for this distribution we tested a number of curve estimates. The resulting models explained a slightly greater portion of the variance (6.8% to 8.4%) but resulted in no large changes in the significance of coefficients. Therefore we determined that a linear model was preferable due to the ease of interpretation of the coefficients.

5.1 Survey Mode

Unlike in the logistic regression models, a mode variable was also included in all response length models. This variable was excluded from the previous analysis because it didn't prove to be a significant predictor of item non-response.

For obvious reasons, the mode of the survey has a substantial effect on the length of response. Those responding to paper surveys wrote on average between 114 and 129 characters less than those responding on the internet (depending on the model). These figures are lower than preliminary comparison of mean response length which indicated that those who responded to a paper survey provided on average 160 characters less than web respondents.

5.2 Overall Satisfaction

Overall employee satisfaction proved to be a strong predictor of length of response, as it was for item non-response. A very dissatisfied respondent can be expected to provide roughly 100 characters more than a very satisfied respondent. This difference is higher than the 67 character difference suggested by the preliminary mean analysis. This adjustment appears to be due to a fairly large gap in satisfaction between paper and web respondents (64% satisfied or very satisfied vs. 76%).

5.3 Education

Education is also a strong predictor of length of response. Those with a high school education or less provided nearly 100 characters less on their open ended responses compared to college graduates. This is less than the 153 character difference found in the preliminary mean comparison analysis. This is due to the lower level of educational attainment of paper survey respondents.

Both those with a high school education and some college or technical training wrote significantly less across all models, while advanced degree respondents provide significantly longer comments.

5.4 Gender, Race, & Ethnicity

When adjusted for education, satisfaction, and survey mode; women provide significantly longer comments. The difference (31 characters) is lower than the difference observed in the preliminary comparison of means. This is because women are less than half as likely to have completed a paper survey.

While African Americans are not significantly less likely to respond to the open ended questions, they provide significantly shorter responses (80.7 less characters than white respondents)—a difference that is well in line with the 84.9 character difference observed in the preliminary analysis. A 10 variable stepwise regression indicates confirms the strength of the inverse relationship between the African American demographic and response length is strong indeed. This model indicates that the African American variable is the second strongest predictor of response length, second only to survey mode. This result is disturbing because it indicates that African Americans are consistently under-represented in the volume of opened ended data.

For Asian Americans, the under-representation is two fold. Not only are Asian Americans in this sample significantly less likely to respond, but they also provide significantly shorter responses. The model indicates that Asian Americans provide 55 less characters than white respondents—a difference that is less than preliminary data indicated (81 characters).

While preliminary analysis indicated shorter responses for Hispanic employees, there does not appear to be a significant difference when adjusted for education, satisfaction, and survey mode.

5.5 Age and Tenure

Respondents over 50 are less likely to respond and provide significantly shorter responses when they do. Those between 50 and 60 on average provide 27.1 less characters and those over 60 provide 47.8 less characters than those 40 to 50 years of age—a difference similar to that observed in the initial analysis of means. Despite age differences, there is no significant difference in response length related to the number of years with the organization.

5.6 Level in Organization

Lower level employees (GS 1-8) not only are less likely to respond to open ended questions, but also provide significantly shorter responses—providing on average 39.4 characters less than low level professional respondents (GS 9-12). There is no significant difference in length of response of supervisor.

6. Discussion

This paper explores how item non-response and response quality vary among different populations in a large government employee survey. In doing so, this paper confirms that a number of groups are under-represented in the volume of qualitative data collected in this survey.

As expected the level of satisfaction of the employee is a strong predictor of the item non-response and response length. When adjusted for education, very dissatisfied participants are 2.4 times as likely to respond to either or both of the open-end questions as very satisfied respondents and provide 100 more characters when they do respond. It is debatable how much these large numbers should concern the researcher. Those who have analyzed qualitative data of this type will contest that there is much more to be learned from comments that are critical of the organization. Praise simply doesn't provide much information unless it is qualified. Unfortunately, satisfaction is not evenly distributed in most organizations. Therefore data collected may provide good information about a small subset of the organization—information that may or may not be useful for the rest of the organization.

Although the results were also expected, the under-representation of less educated respondents is more alarming. Those with a four year college degree or greater education are 1.25 times more likely to respond to one or both open ended questions and provide much longer responses—on average 111 characters longer. The opinions and observations of lesser educated employees are greatly under-represented in the qualitative data.

Not surprisingly, lower level employees are similarly under-represented in the data. The lower level (GS 1-8) employees are less likely to respond and provide on average 91 fewer characters in their response. Regression analysis reveals that only half of this difference is explained by education.

When analyzing open ended comments in employee surveys, special attention should be given to assure lower level employees voices are heard. In some management perspectives, where a major concern is the retention of its skilled

workforce, the over-representation of higher educated employees is less of a concern. This however is a very narrow perspective of organizational assessment. The private sector is growing to realize the important contribution that employees on the ground can make in suggesting organizational and process improvements. A climate survey is one of the few times that lower level employees have an opportunity to contribute to this dialogue. Effort should be made to assure that their perspectives are heard.

Race also is a factor in item non-response and response. Asians & Pacific Islanders in this population are roughly two thirds as likely to respond as white respondents (adjusted for education and level of satisfaction), and provide responses on average 64.0 characters shorter (55.5 characters when adjusted). African Americans in this population are as likely to respond as white respondents but, when they respond, provide responses on average 80.4 characters shorter (80.7 characters when adjusted for education and satisfaction).

The potential for bias due to the under-representation of racial groups in the qualitative data is great. Therefore it would be prudent to do a preliminary analysis of the open ended comments to determine if there are significant differences in how people in under-represented groups respond. We would hope to find little difference in the responses, but might instead uncover important issues that might otherwise be overlooked.

Gender is also, to a lesser extent, a potential source of bias. In this study, women are over-represented in the qualitative data simply due to the length of response—on average 52.1 characters longer then men (31.4 when adjusted for education and satisfaction). Since the difference is relatively small and in the opposite direction of societal bias, it is probably unnecessary to probe deeper.

In this study, age is also a significant indicator of item non-response and response length. The difference, however, is only significant for the oldest two age groups (51-60 & 60 or mores years of age). These groups are significantly less likely to respond to open ended questions and provide shorter responses than other age groups. Interestingly, those in the age group of 41-50 years are the most likely to respond and have the longest response (adjusted and non-adjusted).

This would seem to indicate that this demographic has a greater vested interest in the improvement of the organization especially in comparison to older employees.

7. Implications

It is important to note that simply because a particular group is over or under-represented in the qualitative data, it doesn't necessarily indicate the presence of bias. Bias is the under-representation of ideas, not demographics and in so far as people of different backgrounds share common beliefs, bias will not materialize. It is, of course, naive to think that this is the case. In large organizations, people of different backgrounds can have vastly different experiences and opinions. It is the researcher's job to capture these opinions.

In light of the findings of this paper, it would be prudent to suggest that researchers should test their own qualitative data. A simple comparison of item response rates and mean response length by available demographics would give a researcher a good indication of potential sources of bias. The next step would be to compare responses by the demographics identified in the initial analysis. If a coding system is utilized, this process can be as easy as running crosstabs. For a subjective qualitative analysis this process is more difficult. A group to group comparison can be a lengthy and unnecessary endeavor. In most case it would suffice simply to code open ended responses with key demographics. Researchers should know the source of the comment they are analyzing so they can detect patterns in how different groups respond. In all cases awareness of the potential for bias is key.

For a complete version of this paper with charts, please contact the author via email at Mark.W.Andrews@orcmacro.com.