Surveying Attitudes: Questionnaires vs. Opinionnaires

Jay Alden
February 12, 2007

Performance analysts often seem disdainful of measuring attitudes – putting numerical scores on how people *feel* about things. Surveys are often called *smiles tests*. The idea of *subjectivity* in measurement is treated more as a condemnation than a description. Yes, performance analysts greatly prefer measures that are objective in nature such as percent *of projects that meet cost projections* and *number of engineering changes following design review*. These types of data, while not always easy to capture, typically require straight-forward measurement actions like recording, counting, and calculating. Often, these objective kinds of measures are byproducts of internal data processing actions that make them easily accessible and highly credible. On the other hand, the measurement of people's subjective feelings is neither easy to do nor necessarily compelling in the presentation of measurement results.

Why Measure People's Attitudes

In some circumstances, knowing exactly how a particular group of people *feel* about certain things is vital for effective organizational decision making. For example, most large organizations have elaborate processes for tracking *customer satisfaction* – how customers or clients of an organization feel about products or services they have acquired from that organization. If the measurement drops below some targeted figure, they take great pains to diagnose the root causes of the decrease in customer satisfaction so that they can take appropriate corrective actions. They might also use the trend data as a baseline to see if customer satisfaction changes as a result of modifications to their products or services (e.g., did a part substitution or the elimination of a service feature have a significant impact on customer satisfaction?). Similarly, many organizations systematically track the attitudes of their own employees using measures of *employee satisfaction* or *organizational climate* in order to make critical decisions about employment policies and working conditions. The measurement of attitudes also plays a major role in market research when organizations attempt to determine the types of product features that offer a high probability of ultimate sales success.

If so many organizations depend heavily on measuring people's attitudes to make critical business decisions, why is there so much bias against their use among many performance analysts? It is possible that these performance analysts are not making the critical distinction between the two common survey techniques for measuring attitudes. One type – a *questionnaire* – takes a quantitative approach to measuring attitudes and is typically used for determining trends and making critical organizational decisions (i.e., *summative evaluation*). The other type – an *opinionnaire* - takes a qualitative approach to measuring attitudes and is used to explore ways to improve performance (i.e., *formative evaluation*). Each type of survey offers extensive benefits when used appropriately for its intended purpose. However, an *opinionnaire* is sometimes used inappropriately to make major decisions and the poor results have contributed to the disdain and contempt in which the measurement of attitudes is often held.

Part of the problem is that questionnaires and opinionnaires often look identical. They are both composed of multiple questions to which respondents select one of several options, often on a common scale. See Figure 1 for a portion of a sample user satisfaction survey.

2. When I go to personnel staff for service, they generally:

a. inform me about important changes in personnel rules or benefits.

[ ] Always [ ] Usually [ ] Sometimes [ ] Rarely [ ] Never

b. give me accurate information.

[ ] Always [ ] Usually [ ] Sometimes [ ] Rarely [ ] Never

c. treat me courteously..

[ ] Always [ ] Usually [ ] Sometimes [ ] Rarely [ ] Never

d. provide timely service..

[ ] Always [ ] Usually [ ] Sometimes [ ] Rarely [ ] Never

e. provide the service I expect.

[ ] Always [ ] Usually [ ] Sometimes [ ] Rarely [ ] Never

f. offer innovative solutions.

[ ] Always [ ] Usually [ ] Sometimes [ ] Rarely [ ] Never

Figure 1: Portion of Sample User Survey

But, beneath the surface, huge differences exist between a questionnaire and opinionnaire in the ways they are usually developed, scored, and tested.

Development, Scoring, and Test of Questionnaires

A questionnaire undergoes a rigorous systematic process before it is administered to collect information. Here's a typical process:

1. Create an Instrument Outline

The first step in the process is to lay out the organization of the survey. This usually includes a descriptive name for the factor or factors to be measured, the type of item scales that will be used for each factor, and the approximate number of items to be used in the survey to measure each factor. Item scales typically involve qualities such as *frequency*, *excellence*, *importance*, *comparison,* and *agreement*. Generally, the number of options provided on the scale ranges from

three to seven. The number of items chosen for a section is the minimum number necessary to fully cover the factor being measured and produce a reliable score, yet not place a burden on respondent time. For example, the set of six items shown in Figure 1 is part of a larger questionnaire whose outline is depicted in Table 1. This particular survey instrument will yield four different quantitative scores, one for each of the first four sections. The last section on respondent demographics will be used to collect information on such factors as age, gender, work location, and tenure. This information can be used to check the similarity of the survey respondents to the full population of users and to allow a breakdown of the data by the various demographic factors (e.g., does satisfaction with employee development services differ by work location?).

Table 1. Instrument Outline for a Personnel Services User Satisfaction Questionnaire

FACTORS TO BE MEASURED	PRIMARY ITEM SCALE	NO.OF ITEMS
1. Satisfaction with employee development services	5-option Frequency Scale	4
2. Satisfaction with professionalism of personnel staff	5-option Frequency Scale	6
3. Satisfaction with hiring & orientation practices	5-option Frequency Scale	9
4. Satisfaction with benefits administration	5-option Agreement Scale	7
5. Respondent demographics	Various	4

2. Decide on Scoring Rules

As shown in the instrument outline, each section of a questionnaire contains multiple questions or items. Although these individual items might deal with different dimensions of the factor being measured, their collective score is reported as the "measure." The scores from the individual items may be added, averaged, or combined in some other way to yield a single score that represents the factor being measured. For example, the measure of *Satisfaction with professionalism of personnel staff* is based on the six items shown in Figure 1. If a response of *Always* is assigned a score of 5 and a response of *Never* assigned a score of 1, the average score across all six items will fall within a range of 1.0 to 5.0 with the higher scores representing the higher level of satisfaction. With abstract data such as people's *attitudes*, it is virtually impossible to produce a reliable measure with just a single item. Typically, the more items in an instrument, the greater the reliability of the instrument. If desired, an overall *User Satisfaction* score may also be derived by calculating a straight average or weighted average of the scores from the first four sections of the questionnaire.

3. Prepare the Item Pool

The development of a questionnaire requires the initial construction of a large number of items (typically three or four times the number of items specified in the Instrument Outline) from which the final items will be chosen for the questionnaire. The collection of these items is called an *item* pool. For example, the item pool for the section on *Satisfaction with professionalism of personnel staff* should contain anywhere from 18 to 24 items since the questionnaire is to contain six items for this section. The developer would then use item analysis as described below in Step 4 to select the best six items from the pool for use in the questionnaire. To prepare this large pool of items, the developer needs to first search out the kinds of information associated with the factor being measured. Relevant sources usually include such documents as questionnaires used by other organizations, reference texts on the topic, published reports, and even complimentary or complaining letters received from customers or users. It is also helpful to interview people knowledgeable about the factor being measured such as managers, practitioners, customers or users, technical experts, and the decision makers to identify important qualities to include in the items. Table 2 lists some general guidelines for writing the actual items for the pool

Table 2. Guidelines for Writing Questionnaire Items

DO	AVOID
State the item as briefly as possible	Ambiguous, technical, bureaucratic, and colloquial language
Emphasize *Crucial* or **NEGATIVE** words	Negatively worded items
Make consistent and careful use of 1st, 2nd, and 3rd person styles	Doubled barreled items (asking about two different issues in the same item)
Match the sense of the question or statement with the chosen scale (e.g., *frequency, agreement*)	Leading, loaded, and embarrassing questions
	Extreme words (*always, never*)

4. Conduct an Item Analysis

The next step in the systematic construction of an attitude questionnaire is the selection of the best items from the item pool. That is, all the items in the large pool are examined in one way or another to reveal the particular items destined for the final instrument that will do the best job for the intended measurement. This process, called *item analysis*, may involve either judges or subjects

Judges: A group of 10 to 30 people are asked to make certain qualitative judgments about all the item statements in the pool (e.g., relevance, importance, clarity) and those statements with the best average ratings are retained for the final instrument.

Subjects: The entire pool of items is administered to a test group of 20 to 30 subjects. The test subjects are then placed in three categories based on their score across all the items – *highly favorable, relatively neutral*, and *highly unfavorable*. The results of each individual item are then analyzed to see how well the item discriminated between the groups of test subjects that were *highly favorable* and those that were *highly unfavorable*. The items with the best ability to discriminate between these two extreme groups are selected for the questionnaire.

5. Test Final Instrument

Even after the seemingly best items have been selected for the attitude questionnaire using item analysis, the instrument is still not ready for actual use. The instrument should first be tested for *reliability* and *validity*.

Reliability: This quality represents the consistency of scores produced by administration of the questionnaire in spite of variations in irrelevant measurement conditions. That is, the score obtained from an instrument should not be affected by the environmental conditions that exist at the time of measurement or the way it is administered or the precise wording of the items. The score on a questionnaire should be the same whether it is raining or sunny outside, whether you make a selection with a number 2 pencil or type in your selection, or whether the item says "larger" or "greater." Some typical ways of testing for reliability include (1) administering the survey twice to the same group of people under different measurement conditions (*test-retest reliability*), (2) administering two versions of the same survey with slightly different wording to a group of people (*equivalent forms reliability*), and (3) administering the survey to a group of people and mathematically comparing the results on each item to all other items (*internal consistency reliability*). In each case a numerical score in the range of 0.00 to 1.00 is derived to represent the reliability of the questionnaire. If a figure of less than 0.80 is obtained, the instrument should be modified before it is used (e.g., adding more items, standardizing survey administration, revising the wording of the items).

Validity: This quality of an instrument refers to the degree to which the resulting score truly represents the factor we intended to measure. For example, in some situations, the number of customer complaints is a *valid* indicator of customer satisfaction – the fewer the complaints, the higher the satisfaction. But in other situations where customers have no choice in the matter or do not have a convenient way to submit complaints, the number of customer complaints has little relationship to customer satisfaction. In this situation, the number of complaints would be an invalid measure of customer satisfaction. The most common ways to determine validity are to (1) have a group of knowledgeable people review the questionnaire and come to a consensus that the set of items seems to measure the factor intended to be measured directly and completely (*face validity*), (2) show that the information covered in the items matches directly and completely with theoretical models and descriptions in accepted reference materials dealing with the factor intended to be measured (*content validity*), and (3) demonstrate that the scores from the new instrument correlate highly with scores from an existing instrument that measures the same factor intended to be measured but is impractical to use for the new application (*concurrent*

validity). If the *validity* of the questionnaire is found wanting, the number, type, or wording of the items should be revised, and the instrument retested again.

Both the techniques and the results of reliability and validity testing should be reported to management before the final instrument is administered

Development, Scoring, and Test of Opinionnaires

Since an opinionnaire is designed more for exploration of an issue than as a basis for critical decision making, the development process need not be so rigorous. The development process for an opinionnaire differs in three primary ways from the development of a questionnaire:

1. Scoring of an Opinionnaire

The results from each item of an opinionnaire are reported individually, usually in the form of an average score or as a frequency distribution as shown in Table 3.

Table 3. Reporting of Opinionnaire Results

Section 2	**Always** [5]	**Usually** [4]	**Sometimes** [3]	**Rarely** [2]	**Never** [1]	Mean
Item a	12	18	7	1	2	3.93
Item b	9	15	10	4	2	3.63
Item c	26	12	2	0	0	4.60
Item d	15	11	8	5	1	3.85
Item e	7	8	14	5	6	3.13
Item f	2	6	10	15	7	2.53

The results in the table show the number of respondents who selected each of the five possible options (i.e., *Always* through *Never*) for the six items shown in Figure 1. The average rating for each item is shown in the right-hand column. These results suggest that users of the personnel services are most satisfied with the *courtesy* of the personnel staff (Item c) and least satisfied with their ability to offer innovative solutions (Item f). Based on these results, the personnel organization might very well reward the staff for their courteousness and further explore ways in which they might become more imaginative in offering users possible solutions to their problems. But, it would be unreasonable to attempt a costly reorganization or to fire staff members based solely on these scores. After all, unlike a questionnaire that reports data as the average score across all six items, each of the scores from an opinionnaire are based on a single item which is unlikely to be reliable. An organization would be foolish to make critical decisions on the basis of these results. It can use the opinionnaire results to consider alternative ways of doing business, but shouldn't make firm decisions that would place lives or a lot of money in jeopardy.

2. Construction of an Opinionnaire

The item pool developed for an opinionnaire typically contains the same number of items that are destined for the final survey. If an employee satisfaction survey designed to explore opportunities to improve working conditions, management controls, and employment policies is to have 12 items covering these various issues, only 12 items are constructed. Item analysis is generally not part of the process for constructing opinionnaires. The developer will engage in the same kind of research used in the construction of questionnaires and follow similar writing guidelines. He or she may also seek the advice of other people concerning the content and wording of the items, but a systematic item analysis process is believed unnecessary considering the way the survey results will be applied. Some opinionnaires also include open-ended questions to elicit clarification for a response or suggestions for improvement.

3. Testing of an Opinionnaire

As with a questionnaire, a draft opinionnaire is subject to testing before it is used for its intended purpose. The technique used for testing each type of survey is consistent with the nature of the data it provides. Questionnaires undergo *quantitative* testing (i.e. reliability and validity) while opinionnaires undergo *qualitative* testing. The standard technique for testing opinionnaires involves *pretesting*. The draft survey is administered to a group of typical respondents. The respondents are observed while taking the survey and they are debriefed afterwards. The developer looks for the following kinds of information during the pretest:

- Are the directions clear?
- Are there any errors or confusion in the wording of items or scales?
- Do the respondents have the background to understand and answer each question?
- Can the survey be completed in a reasonable time period?
- Will all questions be answered?

Any problems discovered during the pretest are, of course, corrected before the survey is formally administered to the target audience. In some cases, opinionnaires are subjected to a low level of validity testing (e.g., *face validity*) before formal administration.

Differences in Non-Response Bias

A further distinction between questionnaires and opinionnaires lies with their approaches to controlling non-*response bias*. This type of bias occurs when the people who actually respond to a survey are not truly representative of the intended audience. For example, let's say that many of the 175 people who completed and returned a user satisfaction survey for a particular software product had very different feelings than the vast majority of the 20,000 or so total users of the software. This situation would be disastrous for a quantitative study involving a questionnaire designed to determine whether the company should merely release a slightly improved version of the software or introduce an entirely new product family. On the other hand, the study findings

might be useful in spite of the non-response bias if the design used an opinionnaire to elicit ideas on how the software might be improved.

Non-response bias can occur in either of two ways:

Non-statistical sampling: Surveys are usually administered to just a sample of the people whose attitude is of interest. Typically, the target audience may involve thousands of people so it is too expensive to send the survey to everyone. Instead, the survey is sent to just a sample of those people. If the study involves a questionnaire, the sampling technique must be statistical in nature. That is, the sampling process must assure that the selection of respondents is randomized (i.e., everyone has an equal chance of being chosen) and that the size of the sample is sufficiently large to assure its representativeness. In this case, non-response bias is intolerable. If the study involves an opinionnaire, the selection of people to return the survey may involve non-statistical sampling techniques like choosing people based solely on the judgment of the performance technologist, or choosing an available group of people because it is convenient to do so. Also, the size of the sample is of much lesser concern than in a quantitative study. A moderate degree of non-response bias is acceptable with opinionnaires.

Response Rate: In almost all circumstances, not everyone who receives a survey completes and returns it. There are many reasons to account for people not returning survey. The intended respondents may be too busy, the topic may be of little importance to them, there may be no easy way to return the completed survey, or the survey may be just too long. The concern is that the people who choose to return the survey may feel very different than the people who discard it. With a low response rate, survey scores are often not generalizable; the results are laden with non-response bias. Response rate is calculated by dividing the number of people who complete and return a survey by the total number of people who receive the survey. For a questionnaire, response rates above 70 percent are necessary because non-response bias is unacceptable. This level of response rate is not easy to attain. Therefore, the performance technologist needs to take special action beyond just distributing the survey (e.g., making it very easy to respond, offering incentives for returning the completed survey, showing how the use of the results is important to the respondent). For an opinionnaire, response rates greater than 25 percent are completely acceptable. The survey responses do not have to generalize to everyone; they just need to offer some insights as to how things might be better.

Conclusion

Questionnaires and opinionnaires are both valuable devices for assessing how people feel about certain things. However, these two types of surveys are developed in entirely different ways to serve very different purposes. Questionnaires provide highly accurate numerical scores that validly and reliably represent people's attitudes. As such, it makes sense to track scores from questionnaires to establish trends and, when appropriate, to make critical organizational decision on the basis of the results (i.e. *summative evaluation*). Questionnaires are quantitative tools for problem detection. Opinionnaires provide indicators of potential problems and point to possible

ways by which performance might be improved (i.e., *formative evaluation*). They are qualitative tools for problem diagnosis.

Since questionnaires and opinionnaires have a similar appearance, the danger arises that one type of survey will be used for the other's intended purpose. It would be *inefficient* to use a questionnaire for the purpose of formative evaluation. The extra time and effort devoted to item analysis, quantitative testing, and avoiding non-response bias would add little value. Conversely, it would be *ineffective* to use an opinionnaire for the purpose of summative evaluation. The quantitative results from an opinionnaire have questionable accuracy. Poor organizational decisions may result.

Performance analysts need to be aware of the distinction between questionnaires and opinionnaires. If their purpose for assessing people's attitudes is formative evaluation, they can feel comfortable going through the less rigorous process in the development and testing of an opinionnaire. But, if their purpose is to establish a system of performance management involving summative evaluation (i.e., long-term tracking of people's attitudes), they need to invest the necessary time and effort in the development and test of a questionnaire; one that will yield valid and reliable numerical scores. Or, if an existing survey is proposed for this purpose, they should demand evidence that the questionnaire had been produced by a systematic process that includes the reporting of the instrument's validity and reliability. By making this distinction, perhaps the subjective measurement of people's feelings will have greater credibility in the minds of performance analysts.

Jay Alden is a Professor at the Information Resources Management (IRM) College of National Defense University in Washington DC. He conducts courses on performance measurement. He is also the Chief Editor for the College's Community of Practice supporting federal Chief Information Officers. Dr. Alden previously was the Director of Executive Programs at the University of Maryland University College and the Director of Evaluation and Research at Xerox Corporation.