



Is Your Organizational Attitude Survey Reliable?

Avi Benita, *Metrics Institute*

Background

Any student doing a quantitative research learns that the quality of a survey, or any other measurement tool, is determined by its **validity** (the ability to reflect the subject being measured) and **reliability** (the ability to be consistent: for example two employees with a similar attitude regarding a specific question, would score it the same way every time).

While validity is difficult to evaluate, since it requires additional, “objective” measures, the subject of reliability is relatively more convenient to handle since it mainly depends on the approach used to build the questionnaire and analyze the data.

An unreliable survey might be not only worthless, but also have negative consequences since it creates a distorted view of the subject being measured and may lead to erroneous and harmful actions. It is possible that of all the factors which determine the quality of a survey, the subject of reliability is the most critical.

Despite its academic aura, the subject of reliability is completely practical and is an essential part of the survey and not an esoteric aspect thereof.

Therefore, it is astounding that the subject does not receive the appropriate attention which it deserves from those engaging in this field, to say the least. Generally, it can be said that the awareness of methodological issues in the field of organization attitude surveys is very sparse and not in line with the popularity of surveys over the past few years.

There are many reasons for this being the case, but we believe the most significant one lies in the difficulty to distinguish the difference between a reliable survey and one which is not. The emergence of “automatic” internet systems has worsened the situation since emphasis has shifted towards processing and presentation rather than content and analysis, and the survey’s “soul” has been somehow lost along the way.

Purpose

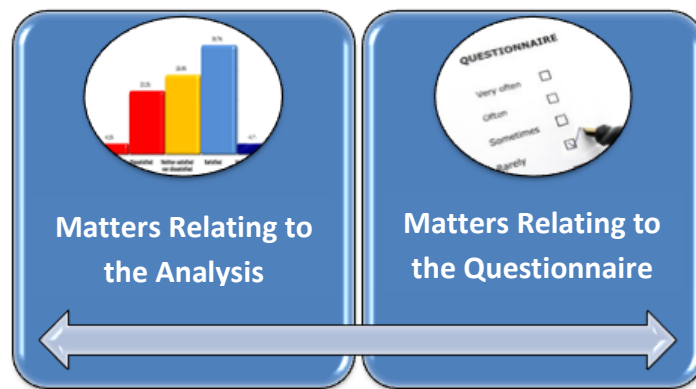
This article reviews some of the main issues that affect a survey’s reliability and the possible ways of dealing with them. It was written from a very practical approach and one which is less academic. It is based on experience with hundreds of organizational surveys and expresses the

opinion of the article's writer. Some of the issues presented in this article have been reviewed more concisely in a previous article regarding paradigms in organizational attitude surveys.

Even though the subject of reliability can be applied to any survey or measurement tool, this article refers to four types of surveys which any common organization conducts periodically at varying intervals:



The article examines the subject of reliability according to the two main domains in which it manifests:



Matters Relating to the Questionnaire

Needless to mention how much a questionnaire's quality is critical. A well-built questionnaire will provide valuable results even if the analysis is basic and simple. However, the most advanced statistics could not rescue an ill designed questionnaire.

Questions wording

This constitutes the Achilles heel of many surveys. It stems from the person composing the questionnaire being more focused on the output (what to ask?) and less on the person filling out the questionnaire (how to ask?)

In many instances, this results in questions that are unclear or not precise enough, which can be subject to plenty of interpretations. As a result, the score given to a specific question is overly dependent on the way in which it was understood, which harms the reliability of the result since two employees with a similar stance towards a subject might score the same question differently.

The following are concrete (real) examples which might illustrate the most common errors:

Question	What might impact reliability?
<p>“I receive recognition and praise for my work/ achievements”</p>	<p>“My work” and “my achievements” are two different things. So what does this mean?</p> <p>It is unclear who is supposed to provide recognition and praise (The direct supervisor? The team? Human resources?)</p> <p>“Recognition and praise”: why both?</p> <p>Maybe the person filling out the questionnaire does not have any extraordinary achievements and thus is not worthy of praise (and not because the company he works for does not praise achievements)</p>
<p>“The procurement unit keeps up with obligations and schedules, and works to fulfill its assignments”</p>	<p>Too many attributes in on question...</p>
<p>“In my team, there is a clear sense of identity and belonging”</p>	<p>Apart from the issue of double concepts (identity and belonging), the way in which an average employee understands the meaning is more troubling. The subject is definitely worthy, but might be understood differently by different employees... if at all.</p>

<p>“I am part of the decision-making process in areas pertaining to my work”</p>	<p>The question assumes that every employee must be included in every decision made which pertains to his or her work. Is this the case? Maybe the average employee does not expect and does not even aspire to be part of the decision-making process and assumes this is the duty of managers?</p>
<p>“I feel that I am motivated and committed to my workplace”</p>	<p>Does anyone expect an employee to state that he or she is not motivated and committed? It is then not surprising that questions phrased in this manner always receive very high scores and very low variance.</p> <p>Subject such as this are better approached from a less individual centered perspective, as in <i>“The employees in my team are motivated and committed to their workplace”</i></p> <p>Experience shows a stark difference in scores between the two wordings.</p>

Response Scale Span

Another dilemma, which is related to the response scale, concerns the size of the scale. Many of those who engage in the field of surveys tend to use a 10-level scale out of the belief that it is somehow more sensitive than the much more common 5-level scale and, as such, is expected to yield a greater variance between responses. Another explanation provided by those devoted to the expanded scale is based on its intuitiveness since it is reminiscent of the grade scale in school or reality shows:

2. To what extent do these attributes fit your experience with the Service Provider on a scale from 1 (Very negative) to 10 (Very positive)?

Demonstrates expertise and proficiency

1 2 4 5 6 7 8 9 10

The biggest disadvantage of the 10-level scale lies in the inability to provide an explicitly worded degree for each level on the scale, thus it is not possible to be sure that a person who scored a 7 is indeed less satisfied than one who scored an 8, which might greatly harm the reliability of results.

In opposite, a worded expression can be assigned to each level on a 5-level scale, and in practice the person filling out the questionnaire does not provide a numeric grade, but rather selects an expression out of the five being offered.

3. To what extent do these attributes fit your experience with the Service Provider?

To a very high extent To a high extent To a moderate extent To a low extent To a very low extent

Demonstrates expertise and proficiency

Another disadvantage of an expanded scale lies in the limited ability to present a distribution of grades, and it is why most of whom using it bypass this limitation by artificially reducing it to 5 levels. If so, why is an expanded scale even needed to begin with?

The argument regarding the scale's sensitivity due to its broader size is pretty limited since lower scores are seldom used. A greater variance might be achieved if compared to the 5-level scale, but we are of the opinion that this is an artificial inflation which stems from the choice between adjacent levels (such as 7 or 8) being quite random.

Moreover, most of the benchmark publications around the world are on a 5-level scale. These and other reasons (such as the difficulty in presenting a distribution of 10 categories) lead us to recommend a 5-level scale in place of the 10-level scale.

A 10-level scale might have some advantages when used in long telephone surveys where the verbal scale is difficult to convey to the person providing responses, thus allowing them a choice between 1 and 10, while assuming this does not have decisive consequences since these questionnaires are usually used for marketing and not for organizational attitude surveys which can a strategically impact.

Matters Relating to the Statistical Analysis

Averaging of Questions

Reliability can also be harmed by the manner in which results are analyzed. The main problem here is the way in which different questions are grouped together.

Many questionnaires are built from various clusters, where each one contains several questions. For example, in an organizational climate survey, such a subject might be "work environment", which would include questions regarding the computers, maintenance, comfort of seating arrangement and maybe even the quality of food and relationships with coworkers.

Therefore, it seems only natural that when analyzing the survey, an averaging of all these questions is performed and the average grade is assigned to the subject of "work environment". Some go even farther and calculate the average grade for all the questions in the survey and compare it with the grades given to the various questions or subjects. This grade is then used as a



sort of leading index by which a benchmark and comparison is performed between units and is used to calculate trends over the years.

This practice is **completely invalid and erroneous from a statistical perspective** since it actually “mixes tomatoes and oranges” and creates a very unreliable result. What for example does the average score for satisfaction with *room cleanliness* and *relationships with coworkers*, represents?

The problem with the aforementioned approach cannot be fully explained, but it is generally connected to the issue on averaging questions that have a low correlation, meaning they do not probably represent the same content.

As a rule, preliminary tests should be conducted prior to combining questions under one category and averaging them. The most recommended one is the Factor Analysis technique, which helps to identify coherent dimensions. Having sat that, it is important to remember that Factor Analysis or any statistical technique is a support tool that must always be supplemented by inner content and experience.

Sample Representativeness

In organizational surveys, the entire organization is usually asked to fill out the survey, but the response rate will be less than 100% more often than not, and instances of a 50% response rate are not uncommon. Therefore, analysis of the survey actually refers to a sample and not the entire population.

As is well known, a representative sample depends on two different factors:

- **Size (n)** – relevant to the issue of significance, which receives too much attention.
- **Randomness** – meaning the extent to which the response rate is similar among all employee groups.

While the size of the sample is not usually a problem when looking at the entire company or its major groups, randomness might be much more significant and influential on the reliability of results. A simple example demonstrates the issue:

	In the population	In the sample
Managers	N = 100 (16.6%)	N = 50 (25%)
Employees	N = 500 (83.3%)	N = 150 (75%)



In the example above, it is obvious that the Managers group is overrepresented in the sample, thus any company wide measure might be biased. The severity of the bias grows insofar as there is a difference between the various groups' attitudes.

One of the ways this can be handled is by weighting the results on a companywide level in a manner which gives each group its true weight. It is recommended that such weighting be performed only in cases where there is a significant difference in response rates between the groups. When analyzing each group individually, the issue of randomness is less relevant.

Conclusion

In this short article, we attempted to highlight a critical subject which we believe does not receive the attention it deserves, even by "experts".

As mentioned, the main reason lies in that it is difficult for a nonprofessional, to differentiate between an unreliable and a reliable survey. Precisely because of this difficulty, it is important to maintain a professional and consistent approach, and not to compromise just because "nobody feels the difference anyhow".