

RELIABILITY OF THE STANFORD CUSTOMER SURVEY

The Stanford Customer Survey was found to be very reliable in measuring customer satisfaction. This conclusion is based on a Cronbach's Alpha (described further below) of .89 for the Likert satisfaction scales on the general section of the survey. A Cronbach's Alpha of .70, or higher, is generally considered to an adequate indicator of reliability.

Defining Reliability and Validity

The reliability and validity of social research instruments are often confused because they have fairly precise definitions in a research context that can seem somewhat at odds with how these terms are used in everyday language.

Reliability refers to the extent to which an instrument yields consistent (repeatable) results when measuring the same thing multiple times. *Validity* addresses whether we are measuring what we claim we are measuring. Reliability is a necessary, but not sufficient, condition for validity; an instrument can be extremely reliable and have virtually no validity. Reliability and validity are both required for a measure to be meaningful.

The "yardstick" analogy is often used to illustrate the interaction between reliability and validity. A rigid yardstick that is accurately calibrated in inches would be highly reliable and valid for reporting the length of an object in inches. An incorrectly manufactured yardstick, where each indicated inch is actually an inch and a half, would still be just as reliable, but have very low validity. A yardstick made of a stretchable material would be highly unreliable and therefore have little validity.

Determining Reliability and Validity

There are several subcategories of reliability and validity, and numerous, often complex, methods for quantifying them. There are limited, but reasonable, choices for addressing these issues in general surveys.

Internal consistency reliability can be used in instrument development, or as a post hoc method of determining reliability. Internal consistency reliability is determined by comparing scores of questions that should yield similar or opposite results, and how well they correlate with overall ratings.

Cronbach's Alpha is a commonly used statistical test to estimate the internal consistency of a survey along some dimension – in our case customer satisfaction. Cronbach's Alpha can vary between -1 and +1, and, in general, the higher the number, the better. A Cronbach's Alpha's of .70 or greater is typically considered adequate to demonstrate instrument reliability. For the general section of the Stanford survey, the overall Cronbach's Alpha was .89, and ranged between .81 and .94 for subsections such as help desk, software, hardware, etc.

Face validity is a common way of dealing with validity in developing surveys with straightforward constructs, and was the approach used for the Stanford survey. Multiple reviewers, internal and external to Stanford, reviewed questions to basically ensure they made sense. Does the item appear to measure what's intended? Simple, unambiguous, questions that multiple reviewers interpreted the same way were regarded as having high face validity.