CREDIBILITY OF THE 2003 MIT CUSTOMER SURVEY

To determine if the results of a survey are meaningful there can be substantial confusion because many of the key concepts have fairly precise definitions in a research context, and this is often at odds with how these terms are used in everyday language. Within this short discussion, the term credible is used to describe the extent to which a survey is meaningful. Can the results be believed? There at least three key criteria that should be met, and these are often confused with overall credibility: is the survey representative, reliable, and valid?

For the 2003 MIT survey:

- Random sampling and responder vs. nonresponder analysis ware used to establish that the survey was *representative*.
- Cronbach's Alpha, an estimate of internal consistency, was calculated to quantify and establish *reliability*.
- An extensive peer review was used to address *validity*.

Some basic details are provided below.

Is the data representative?

If the data isn't representative of the target population, the other issues are irrelevant. For surveys where the response rate is substantially less than 100% (virtually all satisfaction surveys), nonresponse bias may be an issue. Drawing a random sample and contacting them to fill out a survey is a fairly straightforward exercise in the context of an internal satisfaction survey. The key issue is not whether the survey was sent to a representative (random) group of customers, but rather, are the persons that responded still representative of the population?

Two basic techniques to estimate response bias are:

- compare the responses of early to late responders,
- compare the demographics of responders and nonresponders.

Both techniques were used for the MIT survey and the results indicated that nonresponse bias was not a problem. The results can be treated as representative of the target population.

Is the survey reliable?

Reliability refers to the extent to which an instrument yields consistent (repeatable) results when measuring the same thing multiple times. Reliability is a necessary, but not sufficient, condition for validity; an instrument can be extremely reliable and have virtually no validity.

There are a number of methods to estimate reliability and 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 MIT's survey,

Cronbach's Alpha ranged between .75 and .94 for various survey subsections such as help desk, software, hardware, etc.

Is the survey valid?

Validity addresses whether we are measuring what we claim we are measuring. Face validity is a common way of dealing with validity in developing surveys with straightforward constructs, and was the approach used for the MIT survey. Multiple reviewers, internal and external to MIT, 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.

A note about confidence levels and confidence intervals

By themselves, the confidence level and confidence interval don't reveal much about the credibility of a survey. If the survey isn't representative, reliable, and valid, the confidence interval and confidence level are simply garbage-in garbage-out calculations. Given that the other criteria have been established, we can also treat the published confidence levels and confidence intervals as credible.