Using Latent Variable Models in Survey Research

Roger E. Millsap

Arizona State University

Contact: millsap@asu.edu

(480) 965-2584

What is a "latent variable" in the context of survey research?

A latent variable is an unmeasured variable that is believed to influence responses to one or more survey questions.

"Unmeasured" means that we do not have a perfect measure of the latent variable, although we may have one or more items that attempt to measure it.

Example: "satisfaction with dental services" as measured by several survey items on a dental consumer survey

What is a "latent variable" in the context of survey research?

Some latent variables are **intended targets** of the measurements: if we want to measure "satisfaction with dental services", this is a target latent variable and we include items to measure it.

Other latent variables are **unanticipated** and contribute to error of measurement:

--sources of random error or noise in the measures

--sources of bias

Example of bias: under-reporting of dietary intake
selective reporting of behavior

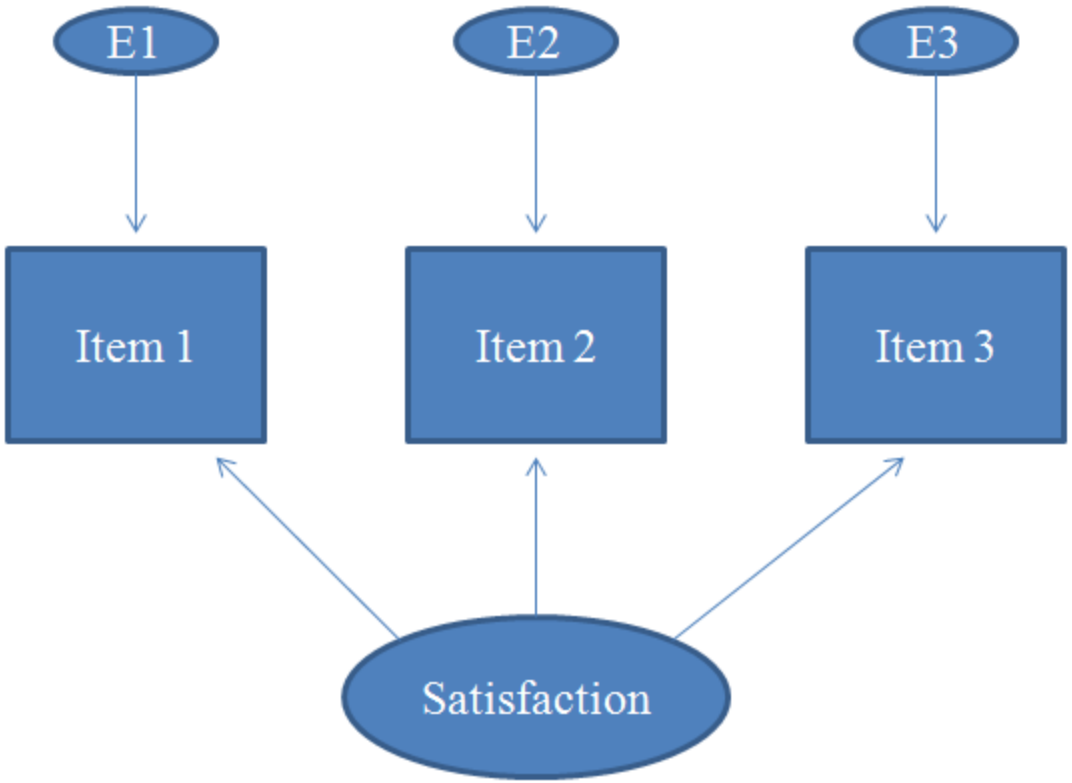
What is a "latent variable" in the context of survey research?

Targeted latent variables will ordinarily need to have multiple **indicators**, or observed items/scales that are intended to measure those latent variables.

Example: Three satisfaction items (squares) that are intended to measure one satisfaction latent variable (the circle)

What is a "latent variable" in the context of survey research?

We might further consider measurement errors (smaller circles) that influence each item separately, producing some unreliability in measured responses to these items:



What is a "latent variable" in the context of survey research?

A **latent variable model** specifies the relationship between the measured variables (items, scales) and the latent variables.

The model will include consideration of two important issues:

1) How many latent variables are there?

2) Which latent variables are related to each measured variable?

The first issue is denoted the **dimensionality** problem. For example, with the satisfaction survey, we might ask "how many different facets of satisfaction are there?"

The second issue arises with multiple targeted latent variables to be measured by different sets of items, for example.

How can a latent variable model be helpful in survey research?

1) The model will help separate variance due to measurement error from variance due to the targeted latent variable.

   -- we can estimate scale **reliability**

   -- we can estimate scale relations with other variables without the influence of measurement error

   --we can design shortened versions of longer scales that have the same reliability as the long versions

How can a latent variable model be helpful in survey research?

2) We can use the latent variable model to help understand or verify the number of targeted latent variables.

--In new surveys, we can evaluate whether items function together as a measure of a latent variable.

--We can use the model to help eliminate items that do not function as desired, or that introduce biases into the measure.

--We can use the model to help understand changes in item functioning in longitudinal surveys.

--We can use the model to evaluate the equivalence of surveys translated into different languages.

Types of latent variable models

1) Models for **continuous** latent variables: these models assume that a latent variable is continuously scaled, with respondents arrayed along a quantitative dimension.

--**factor analysis**: exploratory and confirmatory

--models based on **item response theory** (IRT)

Both of these models can now incorporate multiple latent variables. IRT represents a broad array of models for item-level data. It has been used in cognitive testing but is now expanding to include other forms of measurement.

Types of latent variable models

2) Models for **categorical** or **discrete** latent variables. These models assume that respondents are in groups or types, with the groups being unknown. A key issue here is the number of groups.

-- **latent class analysis** (LCA)

-- **latent mixture modeling**

These two modeling approaches are closely related. LCA is usually applied to item-level data with binary or polytomous items. Latent mixture modeling can apply to continuous measured variables.

Types of latent variable models

The choice between the continuous vs. categorical models depends on how you define the latent variable.

If you believe that the survey items measure a quantity that is best conceptualized as continuous, use factor analysis or IRT.

Example: satisfaction as a continuous latent variable, or with multiple satisfaction latent variables

If you believe that the survey responses are determined by the type of respondent, you may prefer one of the discrete models.

Example: consumers defined within a typology or market segment, with different segments showing different response patterns

Types of latent variable models

You might also decide to combine several models in order to appropriately analyze your survey.

For example, you may feel that there are distinct groups of respondents (e.g., latent classes), and within each group, there is a different factor model.

An illustration would be that there are distinct market segments, each having one or more continuous factors that underlie the survey items, with different factor models in each segment.

This situation could be represented by a **factor mixture** model.

What steps do I take to actually use a latent variable model?

1) **Decide on the type of model**: Here you consider the research questions that you want to answer, and the most appropriate form for the latent variables of interest.

2) **Locate appropriate software**: Latent variable modeling requires software. Available software is described below.

3) **Specify the model and evaluate fit**: Decide on the number of latent variables, and their relations to the survey items. Evaluate whether the model fits the data.

4) **Revise the model or change the items as needed**: You may alter the model, or decide to eliminate some items.

Software for latent variable modeling

Exploratory factor analysis (EFA): Programs for EFA are widely available. SPSS and SAS both offer EFA programs as part of their general menu of statistical procedures.

Confirmatory factor analysis (CFA): Software for CFA is more specialized, but there are many options.

Mplus: http://www.statmodel.com/
LISREL: http://www.ssicentral.com/lisrel/new.html
EQS: http://www.mvsoft.com /
AMOS: http://www-01.ibm.com/software/analytics/spss/

Software for latent variable modeling

Item Response Theory (IRT): IRT software is available, but with the exception of the first program listed below, it is often challenging to use without considerable knowledge of IRT.

IRTPRO: http://www.ssicentral.com/irt/index.html
BILOG-MG: http://www.ssicentral.com/irt/index.html
PARSCALE: http://www.ssicentral.com/irt/index.html
WINSTEPS: http://www.winsteps.com/index.htm

Latent Class Analysis (LCA): Several options exist here as well. Mplus (mentioned above) can be used for LCA.

LATENT GOLD:
http://statisticalinnovations.com/products/aboutlc.html

Software for latent variable modeling

Latent Mixture Modeling: Software here is more specialized generally. A good choice is the Mplus program mentioned above, as it provides a very general model specification that could be applied to many specific problems. It also allows hybrid combinations of mixture models and other types of models.

Free software: Many of the latent variable models can be implemented with free software that can be downloaded from the web. For example, the **R statistical package** contains programs for factor analysis, IRT, and latent class analysis. Most of these programs require some user sophistication however.