Chapter 22

Measurement Development and Validation in Research:
Statistical Techniques and Illustrations

Lihua Xu
University of Central Florida, USA

ABSTRACT

This chapter describes the importance of measurement in social research and education research. In order to validly compare across groups, whether it is age, gender, ethnicities, or cultures, measurement invariance needs to be established. This is accomplished through single-group and multi-group confirmatory factor analysis. The procedural approach is presented with a detailed illustration from real research in servant leadership in K-12 principals in the United States. Second-order confirmatory factor analysis is described due to its popularity. Procedural steps are cited, and an example is given for illustration. As a major statistical technique in instrument development, exploratory factor analysis is discussed and illustrated at the beginning of this chapter.

INTRODUCTION

Data collection is an important aspect of any type of research study. The quality of data collection tools, measuring instruments, is extremely important in both quantitative and qualitative research. This chapter will primarily focus on the measurement in quantitative research, and particularly, its validations. Validity of a measurement focuses on whether the measurement measures the right thing. This is the core concern of a measurement. A measurement need to be validated across populations and cultures etc. first before we could trust our data generated by the measurement and conclusions we make based on them.

For example, culture can influence construct comparability/measurement equivalence in at least two ways. First, the psychological measurement developed under one culture may not be able to measure the same construct in another culture. Referring to psychological assessments conducted in Asia, Sue and Chang (2003) pointed out that research on cultural differences and similarities assumes that we have valid and equivalent tools with which to evaluate these findings. Cultural values and beliefs can greatly affect item responses

DOI: 10.4018/978-1-4666-7409-7.ch022

to measures assessing social and psychological constructs that are developed and administered. Second, cultural tendency to respond in a particular way (e.g. frequent use of the low and high end of the response scale) might cause nonequivalence. Like cultures, gender and age are two other common factors that can influence measurement equivalence.

Measurement equivalence allows researchers to have confidence in: first, measure constructs are applicable across groups (e.g., gender, age, ethnicity) or cultures; second, scale items are interpreted consistently by respondents across groups; third, rating scales are calibrated similarly across groups; finally, observed mean differences reflect the mean differences of underlying latent traits (Drasgow, 1984). In order to compare scale scores of a measurement between groups, researchers need to make sure that the traits, or constructs that the measurement is trying to measure are applicable in different groups. If it is not, cross-group mean difference comparisons are not feasible. Traits and constructs are latent variables and they cannot be directly observable (e.g., love, motivation etc.). They are measured through behavior manifestations. Usually a Love scale include a few items, or a few subscales. Those items are behavioral manifestations, or operational definitions of love. What is love? Does love mean the same thing to both girls and boys? Do girls think getting presents from their boyfriends or husbands mean they are loved? Do boys think if their girlfriends or wives have sex with them, they are being loved? Only when loves means the same thing to both girls and boys, you can compare how different or similar they are.

To test measurement equivalence, this chapter will introduce the powerful measurement validation techniques, such as exploratory factor analysis (EFA), single-group confirmatory factor analysis (CFA), multi-group confirmatory factor analysis, and higher-order confirmatory factor analysis. EFA is mainly used during the preliminary stage in developing a measurement while CFA is theory-driven and it provides a means for construct validation not only within a culture but also across cultures. For within-culture validation, it tests how well the hypothesized model in terms of relationships of constructs with measuring items can explain the observed data. For cross-cultural validation, it tests whether constructs are measured the same way across groups.

In this chapter, EFA will be briefly introduced for its role in instrument development and construct validation and the emphasis will be on the use of confirmatory factor analysis. Their functions and major related concepts will be described and analytical steps will be introduced and illustrated. Examples based on real research data will be provided to help illustrate the conceptual descriptions. Statistical software such as SPSS and LISREL will be demonstrated.

STRUCTURAL EQUATION MODELING AND CONFIRMATORY FACTOR ANALYSIS

Structural equation modeling (SEM) refers to a general approach of multivariate data analysis that models the relations between observed and latent variables. SEM is the multivariate data analysis method that has undergone the most refinement and extension over the years and has continued to be developed (Hershberger, 2003). In comparison to the traditional regression analysis, ANOVA or MANOVA, SEM has the advantage of taking the measurement error into consideration while comparing group differences.

Structural equation modeling is a comprehensive statistical approach to testing hypotheses about relations among observed and latent variables (Hoyle, 1995). It is a methodology for representing and testing a theory-driven network of linear relations between variables (Rigdon, 1998). It tests hypothesized directional and nondirectional relationships among a set of observed and latent variables (MacCallum & Austin, 2000).