Chapter 8
Scale Development and Factor Analysis

Lihua Xu
Western Carolina University, USA

ABSTRACT

Scale development is an important step in empirical research. This chapter describes the common procedures to follow in scale development with essential factor analytical methods. The concept of measurement invariance, the importance of its testing prior to group comparisons, and testing procedures are discussed. Single-group, multi-group, and hierarchical confirmatory factor analytical methods and associated decision making are described. Procedural steps in scale development and measurement invariance testing are illustrated at length using a real dataset in stereotype threat and principals' leadership style in the United States.

INTRODUCTION

Data collection is an important aspect of empirical research. The quality of data collection tools, scales or measuring instruments, is paramount in this process. Psychometric properties of these instruments, reliability and validity, are essential quality indicators. It is a common practice for educational and social researchers to develop their own instruments. To maximize the likelihood to create an instrument with sound psychometric properties, this chapter attempts to offer some procedural guidelines from the instrument development and validation literature. Illustrative examples are incorporated to shed light on the explanations. Exploratory factor analysis will be introduced first for its role in instrument development and the emphasis will be on the use of confirmatory factor analysis. Their functions, related concepts and procedures will be described and analytical steps will be introduced and illustrated.

Instrument Development Using Exploratory Factor Analysis

Instrument development is a fairly complicated process. The recommended procedures are: first, determining clearly what it is that you want to measure; second, generating an item pool; third, determining

DOI: 10.4018/978-1-5225-7730-0.ch008

format for measurement; fourth, having initial items reviewed by experts; fifth, considering including validation items; sixth, administering the items to a developmental sample; seventh, evaluating items; eighth, optimizing instrument length (American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME), 1999; DeVellis, 2012). To make those procedures concrete, first, researchers need to be clear about what they want to measure. More often than ever, social science and educational researchers are dealing with unobservable, abstract concepts and human constructs, such as love, happiness, motivation, and self-efficacy. Theoretical foundations for those concepts are important and substantive literature review helps clarify what researchers wish to measure. Relevant theories should always be considered before developing an instrument of this nature and preferably serve a guide for instrument development (Cronbach & Meehl, 1955; DeVellis, 2012). Lack of theoretical guidance in instrument development can lead to futility and frustration.. The second step of generating an item pool in instrument development is rather straightforward . In the early item development stage, generate more items than necessary and include apparently redundant items. If items are easy to write, generate three or four times as many items as you plan to include in the final instrument. If items are not easy to write, generate 50% more than what you plan for your final instrument. In writing items, use simple languages and avoid writing double-barreled items. Negatively worded items have its merits and yet however, need to be used sparsely and cautiously. Third, response formats used to rate an item vary. The most common one is Likert scale, five-point or seven-point. Other response formats include but not limited to semantic differential scales, rating scales, Thurston and Guttman scales. Five-point Likert response scales or higher are recommended to use to meet the assumption of continuity of variable distributions and a linear relationship between the variables and factors for the legitimacy of analysis of Pearson Product-Moment correlation matrices in exploratory factor analysis (Bandalos & Gerstner, 2016; Finney, DiStefano, & Kopp, 2016). Fourth, have the initial item pool reviewed by at least three subject-content experts for their relevancy and importance in measuring the concepts and constructs and for ease of comprehension and grammatical errors. Fifth, consider inclusion of items to measure social desirability associated with self-report nature of these instruments. Also consider including items measuring similar constructs or consequences to evaluate the construct validity. Sixth, administrate the revised items to a participant sample of adequate size. Depending on the number of the items, the general rule of thumb is 15 participants per item (Field, 2005) or 300 cases (Tabachnick & Fidell, 2001). Preferably these participants are a good representation of the population that the instrument is developed for. Seventh, when the data are collected, if applicable, some items are to be reversely coded and all the items are evaluated for their correlation with other items, their means, and variances. Factor analysis and Cronbach's alpha are used to derive and evaluate the dimensionality and the degree of inter-connectedness of items in an instrument and the reliability of each dimension is measured by the size of coefficient alpha.

Exploratory factor analysis, as a type of factor analysis, is extremely useful in generating the underlying factor structure of an instrument that instrument developers intend to measure when writing items. EFA is used in the beginning stages, is the "heart" of instrument development (Nunnally, 1978) and can be used to inform evaluations of score validity and in the development of operational representatives for the theoretical constructs (Gorsuch, 1983; Thompson, 2010). It is a data-reduction technique and can summarize the inter-correlational patterns among observed variables (items) by generating a few parsimonious underlying latent variables (factors) with the support of very little theory. It is more often data driven. This is the concept of principal axis factoring (PAF). It is different from the other major class of exploratory factor analysis, principal component analysis (PCA). The aim of the latter is data reduc-