CHAPTER 1. INTRODUCTION

1.1 The Problem of Measurement in the Social Sciences

In our daily conversations we often refer to characteristics of people that are difficult to observe. We may say of a person that "he is very active," "she is very intelligent," or "he is more conservative than his wife." We conclude this after observing a specific instance of activity, intelligence, or conservatism. But we do not leave it at that; we generalize. We assume that these observations tell us something about characteristics, or traits, of people that go beyond specific instances. A person who behaves actively, intelligently, or conservatively now will (we believe) also behave this way later. Some people are (we believe) more active, more intelligent, or more conservative than others in general. And social scientists would like to be able to measure such personality traits or characteristics. Because these traits can only be inferred by observing specific instances of behavior, we call them *latent* traits. Latent traits stand in contrast to *manifest* traits, that is, characteristics that can be observed directly and (usually) do not change, such as gender or ethnic background.

Social scientists are often interested in latent traits such as abilities (e.g., level of activity, knowledge, or intelligence) or attitudes (e.g., conservatism, trust, or religious beliefs). They believe that abilities and attitudes are important in explaining why people behave the way they do, why they make the choices they make. This makes the social scientist's work difficult because these characteristics of people are hard to measure. Social science would be much easier if the units of measurement used were simpler, along the lines of monetary units or units from the natural sciences, such as blood pressure, electric potential in the brain, or the chemical composition of blood.

An additional problem is that most latent traits that social scientists want to measure do not have well-established units of measurement. For instance, there is no such thing as the milli-Marx as a unit of alienation or the kilo-Freud as unit of frustration. Social scientists have not been able to build social microscopes or telescopes. They can use numerical (or cardinal, or interval level) measurements only if they are willing to accept a number of restricting assumptions. They may need to settle for measurement at the ordinal level, that is, ordering people with respect to their ability or attitude. In such a case we will call the measurement instrument an *ordinal scale*, sometimes also called a nonmetric or nonparametric scale.

1.2 Measurement Theory and Item Response Theory for Dichotomous Items

We may even wonder whether it is possible to measure latent traits such as abilities and attitudes. The answer to this question is yes, but only if we are prepared to accept certain assumptions. These assumptions are formulated in a theory. Because most of our measurement depends on the interpretation of answers to test or survey questions, *our assumptions deal with a theory about how people answer questions.* One such theory is known as item response theory (IRT). The questions are called items, and the behavior of a person (i.e., his or her answer to a question) is called the response. The simplest behavior is the dichotomous response, which allows only two options. For instance, a person either has or has not performed a certain activity; she has given the right or the wrong answer to a knowledge question; he has agreed or disagreed with a conservative statement. IRT is formalized in a mathematical (measurement) model. This is not as ominous as it sounds. In fact, every step in understanding this model is quite simple.

We will first pursue the possibilities of measuring latent traits of people on the basis of only dichotomous responses. Once the measurement model for dichotomous responses is explained, we will extend it to a measurement model that allows more than two possible responses. For ease of exposition we will refer to a dichotomous item as having two responses: a positive response and a negative response. Which of the two responses the researcher calls positive depends on the direction of the latent trait she wants to measure (e.g., activity rather than inactivity, knowledge rather than lack of knowledge, conservatism rather than liberalism) and has no moral connotations.

1.3 Two Fundamentally Different IRT Models

We can distinguish between two fundamental reasons why people may give the positive (“yes”) response to an item. Let me explain this with an example of two different sets of questions that can both be answered with “yes” or “no.”

1a. Are you approximately 1.80 m (5'11") tall?	yes/no
1b. Are you approximately 1.70 m (5'7") tall?	yes/no
2a. Are you at least 1.80 m (5'11") tall?	yes/no
2b. Are you at least 1.70 m (5'7") tall?	yes/no

All four questions include a size as some kind of benchmark. To questions 1a and 1b a person gives the positive response only if indeed he is approximately 1.80 m or 1.70 m tall. The height of the person is the same as the height that is mentioned in the question. Or, to say it slightly differently, the difference between the height of the person and the height mentioned in the question is negligible, and their similarity or proximity is high. In contrast, the negative response is ambiguous: A person who is not approximately 1.70 tall is either shorter or taller; these two possibilities are each other's opposite.

Each person who says "yes" to question 2a or 2b can have any height greater than 1.80 m or 1.70 m. Moreover, there is an overlap in the groups of people who give the positive response to questions 2a and 2b because anyone who is taller than 1.80 m necessarily is taller than 1.70 m. Relations such as "greater than" or "taller than" are called dominance relations. Similarity relations, as described in the previous paragraph, are called proximity relations.

On the variable height, all people can be represented with a number: their value, that is, their height, in centimeters or in feet and inches. But—and this may be difficult to grasp at first—each of the questions, or items, can also be represented with a number or value: the benchmark value implied in the question. Questions 1a and 1b receive the positive response from a person who has (approximately) the same number as the item. Questions 2a and 2b receive the positive response from a person who has a higher number than the item. A positive response to questions 1a and 1b is called a *proximity* response, and a positive response to questions 2a and 2b is called a *dominance* response.[1]

[1]Proximity responses may seem useless because of the ambiguity of the negative response. But there are situations—for instance in analyses of preferences, child development, or politics—in which it is useful to ask questions that lead to responses that are best analyzed as proximity responses. In those cases the negative response will have one of two opposite meanings. For example, "Do you like coffee with one lump of sugar?" ("No, I like coffee without sugar," or "No, I like coffee with more sugar"). "Does the baby crawl?" ("No, the baby cannot crawl yet," or "No, the baby does not crawl anymore"). "Is voting the only way for people to have a say in government?" ("No, voting is not a way," or "No, there are more ways"). Such data cannot be analyzed with the model described in this monograph. However, they can be analyzed with an IRT measurement model called the unfolding model, which was invented by Coombs (see Coombs, 1950, 1964; and some of his successors, e.g., Andrich & Luo 1993; Roberts, Donoghue, & Laughlin, 2000). An ordinal unfolding model, analogous to the model discussed in this monograph, was developed by van Schuur (1993) and van Schuur and Kiers (1994). Circular unfolding models, also known as the circumplex (e.g., Plutchik & Conte, 1997), have been developed by Browne (1992) and others. An ordinal circular unfolding model has also been developed (Mokken, van Schuur, & Leeferink, 2001).

This chapter will deal only with dominance responses and with an IRT model that represents them: *the dominance model*. It is sometimes called the Guttman scale, after its originator, Louis Guttman (1950).

There are two classes of dominance models: metric and ordinal. The difference is that in metric models we take the values with which people and items are represented as numbers on an interval scale, whereas in ordinal models we interpret these values only in terms of their rank order, that is, as numbers on an ordinal scale. This monograph deals only with the ordinal dominance model (for metric models, see Andrich, 1988; Bond & Fox, 2007; Embretson & Reise, 2000; Ostini & Nering, 2006; Smith & Stone, 2009). We will get back to the implications the ordinal model has for the use of scale scores in linear multivariate statistical techniques, such as regression analysis or analysis of variance, in chapter 9.