

# On the statistical meaning of the item parameters in IRT models

Ernesto San Martín

*Faculty of Mathematics, Pontificia Universidad Católica de Chile, Chile*

*Interdisciplinary Laboratory of Social Statistics LIES, Pontificia Universidad Católica de Chile, Chile*

*The Economics School of Louvain, Université catholique de Louvain, Belgium*

IRT models are widely used in international and national educational monitoring systems, as PISA and TIMMS, or SABER PRO (Colombia) and SIMCE (Chile). These models are palatable because the probability of correctly answering an item is due to two factors: one characterizing a person; the other characterizing an item. The first factor is called *ability*; the second factor is composed of properties of the item such as *difficulty*, *discrimination* and *guessing*. However, what is the meaning of those terms? When we say that an item have a *difficulty equal to 0.9*, are we providing the same information if we use a 1PL, 2PL or 3PL model? When we say that one of the nice properties of IRT models is the joint representation of abilities and difficulties, is this correct if the abilities are considered either as a random effect or as a fixed effect?

The previous questions can be answered if identifiability is understood as a concept related to model construction. We argue that a parameter can be statistically interpreted if it is identified. In any other case, the model is meaningless and therefore not useful. In order to justify these statements, we will discuss the concept of identification; we will show why this concept is fundamental even if the inference is done under a Bayesian perspective. Thereafter, we will apply this conceptual framework to the following cases:

1. The meaning of *difficulty* in two specifications of the Rasch model, namely a fixed-effect model and a semi-parametric model.
2. To show that the joint representation of difficulties and abilities is only valid in a fixed-effect specification of the Rasch model.

3. The meaning of *difficulty* and *guessing* in a 1PL-G model (that can be viewed as a 3PL model with equal “discriminations”).