5

Testing the Rasch Model

Cees A. W. Glas and Norman D. Verhelst[1]

ABSTRACT In this chapter, it is shown that the problem of evaluating model fit can be solved within the framework of the general multinomial model, and it is shown how tests for this framework can be adapted to the Rasch model. Four types of tests are considered: generalized Pearson tests, likelihood ratio tests, Wald tests, and likelihood ratio tests. The statistics presented not only support the purpose of a global overall model test, but also provide information with respect to specific model violations, such as violation of sufficiency of the sum score, strictly monotone increasing and parallel item response functions, unidimensionality, and differential item functioning.

5.1 Introduction

One of the aims of this chapter is to create some order in the plethora of testing procedures for the Rasch model (RM). To present an overview of the various approaches to testing model fit, a taxonomy will be used that is based on three aspects: the assumptions and properties of the model which are tested, the type of statistic on which the test is based, and the mathematical sophistication of the procedure, particularly, the extent to which the distribution of the statistic is known. These three aspects will first be discussed in some detail.

The first aspect of the taxonomy of model tests concerns the assumptions and properties of the model to be tested. Although model tests are constructed to have power against specific alternatives, it is not really possible to completely separate the assumptions in order to test them one at a time. Some confounding will always occur. Given this restriction, the model tests can be ordered as follows.

First, there are tests that focus on the assumptions of sufficiency of the sum score and of strictly monotone increasing and parallel item response functions. To this class belong the Martin-Löf (1973) T-test, the Van den Wollenberg (1982) Q_1-test, the Glas (1988a, 1989) R_1-test, the Molenaar (1983) U_i-test, and the S_i- and M-tests (Verhelst & Eggen, 1989; Verhelst, Glas, & Verstralen, 1994). An important property of the Rasch model is that, under very mild regularity assumptions (see Pfanzagl, 1994), consistent item parameter estimates can be obtained from a sample of any subgroup of the population where the model holds. So item parameter estimates obtained using different samples from different subgroups (say, gender or ethnic subgroups) of the population should, apart from

[1]National Institute for Educational Measurement, P.O.-Box 1034, 6801 MG Arnhem, The Netherlands; fax: (+31)85-521356

G. H. Fischer et al. (eds.), *Rasch Models*

random fluctuations, be equal. Not only the estimates obtained from subgroups formed on the basis of background variables should be approximately equal, also estimates obtained at different score levels should, within chance limits, be equal. Andersen's likelihood ratio test (Andersen, 1973b), the Fischer-Scheiblechner test (Fischer & Scheiblechner, 1970; Fischer, 1974), and the Wald test proposed in this chapter are constructed to be sensitive to violation of this property.

The assumptions of unidimensionality of the parameter space and of local stochastic independence are the focus of the Martin-Löf (1973, 1974) likelihood ratio test, the Van den Wollenberg (1982) Q_2-test, and the Glas (1988a, 1989) R_2-test.

The additional assumptions with respect to the distribution of ability, which are made in MML estimation, are the focus of the R_0-test by Glas and Verhelst (1989).

Finally, tests for the RM can also be based on explicit extensions of the RM, such as the class of log-linear models by Kelderman (1984, 1989), the OPLM presented in Chapter 12, and the RM with a multivariate ability distribution by Glas (1989, 1992).

The second aspect of the taxonomy of model tests concerns the type of statistic used. The classification related to this aspect generally follows the usual classification in discrete statistical models. So tests can be based on Pearson-type statistics, that is, statistics involving differences between observed frequencies and their expected values, likelihood ratio statistics, and Wald statistics. In addition, in Section 5.5 the application of Lagrange multiplier statistics to the RM will be discussed. To the practitioner, the first aspect of the taxonomy, that is, the assumptions that are tested, will be much more important than the second aspect, that is, the type of statistic used. However, from a theoretical point of view and for reasons of presentation, the second aspect has gained priority in this chapter.

The third aspect of the taxonomy of model tests concerns the mathematical sophistication of the procedure, especially the extent to which the (asymptotic) distribution of the statistic is known. In most instances, the tests for model fit can be arranged into two classes: a class of statistics where the (asymptotic) distribution is known, and a class of statistics which can be viewed as approximations of the statistics in the first class. It must be stressed that it is perfectly feasible to construct a test statistic which does have a known asymptotic distribution, but which is completely uninformative with respect to model violations and has little power. So the requirement of having a test focussed on specific alternatives precedes the mathematical sophistication of the approach. But from the viewpoint of interpretation of the outcome, test statistics where the (asymptotic) distribution is known, are far preferable to statistics where this is not the case. Also generalizations of the model, such as the OPLM, and specializations of the model such as the LLTM (Chapter 8 of this book), are fostered by a firm statistical framework.