Summary: One of the important questions in the practice of educational testing is how a particular test should be scored. In this paper we consider what an appropriate simple scoring rule should be for the Dutch as a second language test consisting of listening and reading items. As in many other applications, here the Rasch model [G. Rasch, Probabilistic models for some intelligence and attainment tests. Chicago, Il: Univ. Chicago Press (1980)] which allows to score the test with a simple sumscore is too restrictive to adequately represent the data. In this study we propose an exploratory algorithm which clusters the items into subscales each fitting a Rasch model and thus provides a scoring rule based on observed data. The scoring rule produces either a weighted sumscore based on equal weights within each subscale or a set of sumscores (one for each of the subscales). An MCMC algorithm which enables to determine the number of Rasch scales constituting the test and to unmix these scales is introduced and evaluated in simulations. Using the results of unmixing, we conclude that the Dutch language test can be scored with a weighted sumscore with three different weights.

MSC:
62P15 Applications of statistics to psychology
62F15 Bayesian inference
62J12 Generalized linear models (logistic models)
62H30 Classification and discrimination; cluster analysis (statistical aspects)

Keywords:
educational testing; Markov chain Monte Carlo; mixture model; multidimensional IRT; one parameter logistic model; Rasch model; scoring rule

Software:
eRm

Full Text: DOI Euclid

References:
Soc. Ser. B 56 363-375. · Zbl 0796.62028


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.