

**The impact of item response theory in educational testing and assessment:  
A practical point of view**

**Cees A.W. Glas**

**Department of Methodology, Measurement and Data Analysis,  
University of Twente, the Netherlands**

Educational assessment and evaluation can be targeted at many levels. It can be targeted at the achievements of individual students, at schools, at school districts, at school systems, and even at countries. It can serve the purpose of certification and accreditation, the purpose of diagnosis and improvement, or it can support accountability. It can take the form of an examination system, a pupil monitoring system, a school evaluation system, or a national assessment. Educational assessment and evaluation are based on several data sources, such as the data from achievement and ability tests, students' and parents' background variables, such as socio-economic status, intelligence or cultural capital, school variables, and features of the schooling system.

It will be shown how these various measurements can be combined and related to each other. It will be shown that item response theory (IRT) provides a useful and theoretically well-founded framework for educational measurement. It supports such activities as the construction of measurement instruments, linking and equating measurements, and evaluation of test bias and differential item functioning. Further, IRT has provides the underpinnings for item banking, optimal test construction and various flexible test administration designs, such as multiple matrix sampling, flexi-level testing and computerized adaptive testing.

In this presentation a concise introduction to the principles of IRT models will be given. The models pertain to dichotomous items (items that are either scored as correct or incorrect) and polytomous items (items with partial credit scoring, such as most types of open-ended questions and performance assessments). It will be outlined how the models are estimated and tested, how tests are scored using IRT, how items are selected given specific test purposes.

Finally, some of applications are presented, such as the use of incomplete assessment designs, equating and linking of assessments, evaluation of differences between groups, and applications to multilevel analyses as used in school effectiveness research.