Dries Debeer, **Item-position effects and missing responses in large-scale assessments: Models and applications**. Dissertation submitted to obtain the degree of PhD in Psychology, October 2016. Supervisor Prof. Dr. Rianne Janssen.

Psychological and educational assessments commonly consist of multiple items that are inevitably administered in a specific item order. Hence, effects related to the sequential ordering of items (i.e., item-position effects) may arise. Typical examples of reported item-position effects are a change in the difficulty of items in aptitude tests (i.e., so-called fatigue or practice effects) and an increase in the consistency of the responses in attitude and personality questionnaires (i.e., the so-called Socratic effect). Further, the serial ordering of items also plays a role in skipping items and dropping out before the end of the assessment, both of which are observed in large-scale educational assessments.

Common psychometric models assume that only properties of the test taker and the item contribute to the item response, and that these properties are invariant with respect to the position in which the item is administered. This dissertation focuses on effects of the sequential ordering of items and how these effects can be investigated and modeled using Item Response Theory (IRT). Models are proposed, evaluated and applied to empirical data.

In **Chapter 1**, an IRT framework is proposed to model and investigate item-position effects in achievement tests with dichotomous items. Within the proposed framework, a variety of functions of item position can be added to both the item discrimination and the item difficulty parameter. Further, by introducing individual differences, the position effect on item difficulty can be interpreted as a test taker's persistence. A simulation study indicates that ignoring this item-position effect can result in biased estimates. Further, using two empirical illustrations the applicability of the modeling framework is demonstrated.

In **Chapter 2**, a multilevel extension of the model with individual differences in ability and persistence is formulated and applied to the PSIA 2009 reading assessment data. Persistence, which can be related to a change in examinee effort, is investigated across all participating countries, and the individual differences in ability and persistence are decomposed into a within-school and between-school part. A negative average persistence is found consistently across all countries. Both the negative average persistence and the variance in persistence prove to be stronger in the lower performing countries.

Chapter 3 focuses on the Socratic effect in personality and attitude questionnaires. An IRT approach based on the Generalized Partial Credit Model is proposed, in which the Socratic effect is modeled as an item-position effect on the discrimination parameter. Evidence is found for a small linear Socratic effect in the CES-D, which is a commonly used screening instrument for depression.

In **Chapter 4**, skipped and not-reached responses in educational assessments are modeled using tree-based multidimensional IRT models with sequentially interconnected subprocesses (i.e., IRTrees). Responses and omissions are modeled jointly, taking into account that both test takers and items can contribute to the two types of omissions. A simulation study shows that when the missingness is not at random, ignoring missing responses can results in biased estimates, and that an IRTree can reduce this bias. The applicability of the IRTrees for response omissions is illustrated using the PISA 2009 reading assessment data in Argentina.