



UMEÅ UNIVERSITET

# Theory and Validity Evidence for a Large-Scale Test for Selection to Higher Education

**Jonathan Wedman**

## Akademisk avhandling

som med vederbörligt tillstånd av Rektor vid Umeå universitet för  
avläggande av filosofie doktorsexamen framläggs till offentligt  
försvar i Hörsal 1031, Norra beteendevetarhuset,  
fredagen den 22 september, kl. 10:00.  
Avhandlingen kommer att försvaras på engelska.

Fakultetsopponent: Professor, Jo-Anne Baird,  
Department of Education, University of Oxford, England.

Department of Applied Educational Science  
Educational Measurement  
Umeå 2017

**Organization**  
Umeå University  
Department of Applied  
Educational Science

**Document type**  
Doctoral thesis

**Date of publication**  
1 September 2017

**Author**  
Jonathan Wedman

**Title**  
Theory and Validity Evidence for a Large-Scale Test for Selection to Higher Education.

### **Abstract**

Validity is a crucial part of all forms of measurement, and especially in instruments that are high-stakes to the test takers. The aim of this thesis was to examine theory and validity evidence for a recently revised large-scale instrument used for selection to higher education in Sweden, the Swedish Scholastic Assessment Test (SweSAT), as well as identify threats to its validity. Previous versions of the SweSAT have been intensely studied but when it was revised in 2011, further research was needed to strengthen the validity arguments for the test. The validity approach suggested in the most recent version of the Standards for education and psychological testing, in which the theoretical basis and five sources of validity evidence are the key aspects of validity, was adopted in this thesis.

The four studies that are presented in this thesis focus on different aspects of the SweSAT, including theory, score reporting, item functioning and linking of test forms. These studies examine validity evidence from four of the five sources of validity: evidence based on test content, response processes, internal structure and consequences of testing.

The results from the thesis as a whole show that there is validity evidence that supports some of the validity arguments for the intended interpretations and uses of SweSAT scores, and that there are potential threats to validity that require further attention. Empirical evidence supports the two-dimensional structure of the construct scholastic proficiency, but the construct requires a more thorough definition in order to better examine validity evidence based on content and consequences for test takers. Section scores provide more information about test takers' strengths and weaknesses than what is already provided by the total score and can therefore be reported, but subtest scores do not provide additional information and should not be reported. All four quantitative subtests, as well as the Swedish reading comprehension subtest, are essentially free of differential item functioning (DIF) but there is moderate DIF that could be bias in two of the four verbal subtests. Finally, the equating procedure, although it appears to be appropriate, needs to be examined further in order to determine whether it is the best practice available or not for the SweSAT.

Some of the results in this thesis are specific to the SweSAT because only SweSAT data was used but the design of the studies and the methods that were applied serve as practical examples of validating a test and are therefore likely useful to different populations of people involved in test development, test use and psychometric research.

Suggestions for further research include: (1) a study to create a more clear and elaborate definition of the construct, scholastic proficiency; (2) a large and empirically focused study of subscore value in the SweSAT using repeat test takers and applying Haberman's method along with recently proposed effect size measures; (3) a cross-validation DIF-study using more recently administered test forms; (4) a study that examines the causes for the recurring score differences between women and men on the SweSAT; and (5) a study that re-examines the best practice for equating the current version of the SweSAT, using simulated data in addition to empirical data.

### **Keywords**

SweSAT, validity, theoretical model, score reporting, subscores, DIF, equating, linking

**Language**  
English

**ISBN**  
978-91-7601-732-6

**ISSN**  
1652-9650

**Number of pages**  
51 + 4 papers