Abstract

The National Educational Panel Study (NEPS) analyzes how competencies, e.g. mathematical competence, develop over the lifespan, how they influence educational careers and how far competence development is related to learning opportunities. For data collection a multicohort-sequence design with different instruments at each educational stage is used (Blossfeld & von Maurice, 2011). A requirement for appropriate data use is the valid interpretation of the test scores.

This dissertation contributes to the validation of the NEPS test score interpretations by an extensive analysis of the mathematics test for ninth graders (K9). Therefore, it was examined whether the differences in the observed performance reflect differences in the mathematics competence as defined in the NEPS framework. To validate this test score interpretation a chain of arguments based on the Argument Based Approach from Kane (2013) was developed for the NEPS K9 mathematics test. The analyses were based on the *scientific use file* of the NEPS main study in 2010. The sample consisted of 15.239 ninth graders. An additional validation study was conducted to examine the relationship of the NEPS mathematics tests with other measures of mathematical and non-mathematical competencies. For the assessment of mathematics and science competence, items from the Programme for International Student Assessment (PISA) and from the German National Assessment (LV) were used. In addition a cognitive ability test was applied. Overall 1.965 students from 80 schools were selected. To prepare the validity argument, literature reviews of already published analyses and an expert review of the NEPS mathematics item characteristics were conducted. Furthermore, the competence data were scaled using Item Response Theory. Additionally, dimensional analyses were performed and correlations with criteria of mathematics competence were calculated. The outcomes provided evidence for the validity of the NEPS test score interpretation. Yet, limitations and the need of further research were found. For example, there was no evidence supporting the assumption of one dimensionality of the NEPS mathematics test. Furthermore, additional analyses of the relationship with appropriate criteria of the target domain mathematics competence are desirable. The evidence found in this dissertation supports the test score interpretation "differences in the rank order of the observed performance point to differences in the rank order of mathematics competence, as defined in the NEPS framework". Relevant competence domains of

the target domain could be identified, the scoring criteria were adequate and the relationship with the PISA and LV mathematics test fulfilled the expectations. In this manner, a complete chain of arguments was developed and recommendations for strengthening and extending the interpretation of test scores were formulated. The Argument Based Approach from Kane was proven to be a beneficial approach, which also allowed taking into account the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014).