

Skrivbedömning och validitet

Fallstudier av skrivbedömning i svenskundervisning på gymnasiet

Gustaf Skar

Abstract

Skar, Gustaf (2013). Validitetsperspektiv på skrivbedömning. Fallstudier av skrivbedömning i svenskundervisning på gymnasiet. (Validity and classroom-based writing assessment: case studies of writing instruction and assessment in upper secondary school.) Stockholm University, Dissertations in Language Education 2. ISBN: 978-91-7447-694-1. Written in Swedish with a summary in English.

This doctoral dissertation reports on results from three explorative case studies of teacher assessment practice within upper secondary school writing instruction. In Sweden, almost all responsibility for constructing, administering, and scoring assessments lies with the individual teacher. Unfortunately, little is known about classroom-based writing assessment and even less is known about the validity of such assessment. The aim of this dissertation is to build validity arguments based on classroom assessment practice concerning achievement tests in upper secondary school. Three research questions were formulated in relation to this aim: (1): To what extent can interpretation of scores be argued for? (2): To what extent can students be said to have had equal opportunities to learn what is later assessed? (3) To what extent can suggested and observed usage of scores be argued for, given the relationship between instruction and assessment?

The data for the studies consists of audio recorded observations, student texts, teacher comments, and scoring rubrics, and was gathered within writing units in three upper secondary schools. Altogether the observations comprise 17 lessons (or 19.6 hours). Data was also collected in interviews with three teachers and their students.

The data on instructed and assessed writing was analyzed by conceptual tools related to a theoretical model of writing, the so-called Writing Wheel. The validity argument was built using Bachman's (2005) Assessment Use Argument (AUA) model.

On an aggregated level, the results indicate threats to the validity of interpretation of scores, to the validity of usage of scores, and threats associated with inequitable assessment. The first types of threats stem, for example, from scoring rubrics that are not aligned to the assessment tasks at hand, and a low degree of standardization in the administration of the assessment tasks. The second type of threat is related to this; for example, low standardization led to incomparable student marks. While some students could benefit from contacts with able peers (and/or parents) others could not. The third type relates to possibilities to learn what is later assessed, which was not fully evident in some cases. Finally, the results also implicate that the building of an AUA can serve as a syllabus-design tool for practitioners as well as a design tool in intervention studies.

The closing chapter of the dissertation presents a number of hypotheses based on the case study findings. Concluding remarks suggest how these could be tested.