



CAMBRIDGE ASSESSMENT

How valid are A levels? Findings from a multi-method validation study of an international A level in geography

A paper presented at the Association for Educational Assessment in Europe, 11th Annual Conference, Oslo, November 2010.

Stuart Shaw and Victoria Crisp

Email: crisp.v@cambridgeassessment.org.uk

Abstract

Validity is a key principle of assessment and yet it is remarkably difficult to provide comprehensive evidence in relation to all aspects of an assessment's validity. Last year at AEA-Europe we described the development of a framework to support validation of traditional examinations and the types of methods and analyses that we envisioned being used. This paper will report on the findings from piloting the framework and linked methods and evaluating the validity of an International A level in geography. Evidence was gathered with respect to each component of validity in the framework via a number of methods. These included:

- documentary review of syllabus support materials;
- analysis by subject experts of the assessment purpose(s), intended constructs to be assessed, skills used by students in answering, and representation of skills across examination papers;
- traditional statistical analysis, and Rasch analysis, of student scores to consider item functioning;
- reliability analyses using data from a multiple re-marking exercise;
- analysis of think aloud protocols of students answering exam questions;
- a questionnaire to teachers to regarding advice on teaching practice and on how test scores/grades should be used.

The findings were structured in terms of an underpinning validity argument, with 'evidence for validity', and a number of potential 'threats to validity'. For the geography A level assessments, the evidence gathered suggested considerable support for the appropriacy of the interpretations of exam results. For example, the expert analysis suggested good coverage of topics across exam papers over time, Rasch analysis and student verbal protocols suggested that all but one of the items measured construct relevant skills, and questionnaires suggested that the assessments do not have a negative backwash on classroom practice. Some potential threats to validity were identified and these should feed into improvements to the assessments.



UNIVERSITY of CAMBRIDGE
Local Examinations Syndicate