



CAMBRIDGE ASSESSMENT

Piloting a method for comparing examination question paper demands

Paper presented at the AEA-Europe annual conference
Glasgow, Scotland, 5-7 November 2015.

Lucy Chambers, Jackie Greatorex, Filio Constantinou and Jo Ireland

Research Division
Cambridge Assessment
1, Regent Street, Cambridge
CB2 1GG

Chambers.L@cambridgeassessment.org.uk
www.cambridgeassessment.org.uk

Cambridge Assessment is the brand name of the University of Cambridge Local Examinations Syndicate, a department of the University of Cambridge. Cambridge Assessment is a not-for-profit organisation.

Abstract

Piloting a method for comparing examination question paper demands.

Background

Comparability is essential for a fair examination system: the assessments produced must be of equivalent standards for an examination system to be deemed just. One aspect of comparability concerns the cognitive demands which different examinations pose on candidates. It is important these demands are relatively similar in order to provide appropriately targeted question papers, to allow standards to be maintained and to present candidates with a question paper that does not differ from their expectations of demand.

International General Certificate of Secondary Education (IGCSE) qualifications are sat by candidates from around the world. To help maintain examination security the world is divided into administrative zones each with a different question paper. It is important that the demand of the question papers is similar over sessions and/or between administrative zones.

This study sought to pilot a method for comparing the demands of an academic qualification over sessions and/or between administrative zones.

Method

The research method builds on earlier work which compared the demands of vocational assessments using Thurstone paired comparisons in conjunction with a scale of cognitive demands (Crisp and Novaković 2009).

The research was conducted using IGCSE Geography question papers from four sessions and three administrative zones. Six experts familiarised themselves with the question papers and a demands framework. Each expert made three paired comparison judgements; these required participants to decide which question paper was more demanding according to a demands framework. This was to provide experience of the task and to generate a list of geography based examples of how the framework was used. Researchers summarised the exemplifications and this was used by the participants to inform the next stage of judgements. In the second stage, the experts made paired comparison judgements on all question papers under consideration in a complete design. The experts subsequently completed an evaluative questionnaire.

In a departure from Crisp and Novaković's (2009) original method, Rasch was used to model the paired comparison data and place the question papers on a logit scale from the most to the least demanding. Approximate 95% confidence intervals were used to indicate which question papers were significantly more or less demanding than average.

Findings

The analysis showed which question papers were significantly more or less demanding than average. No session or administrative zone was consistently affected by significantly demanding or undemanding question papers. Any variations in question paper demands can be adjusted for when grade boundaries are set.

The experts felt that the initial judgements familiarised them with the materials and that the exemplification summary supported subsequent judgements. Most experts felt that the method was satisfactory for comparing question paper demands in an academic qualification. The present study provides evidence which suggests that this method can be applied with success in the context of more traditional academic subjects such as Geography.

Reference

CRISP, V. & NOVAKOVIĆ, N. 2009. Are all assessments equal? The comparability of demands of college based assessments in a vocationally related qualification. *Research in Post-Compulsory Education*, 14, 1-18.