CAMBRIDGE ASSESSMENT

Item difficulty modelling: exploring the usefulness of this technique in a European context

Rebecca Hopkin and Victoria Crisp
Research Division, Cambridge Assessment
*Hopkin.R@cambridgeassessment.org.uk

Paper presented at the AEA-Europe annual conference, Belfast, November 2011.

Abstract

Item difficulty modelling is a research technique that involves attempting to statistically model the difficulty of examination questions using variables that represent key observable question features. The resultant statistical model identifies the amount of variability in question difficulty that can be accounted for by the question features. This model may then be used to explore question validity, to support question writing, and to predict question difficulty. Item difficulty modelling is a popular technique in the USA, where it has most commonly been used to examine the difficulty of items in multiple choice tests of tightly-focussed aptitudes or skills. This research approach appears to be **new** to the European assessment context.

This paper presents the results of a recent study which explored the **benefits and limitations** of item difficulty modelling in the UK general qualification context. The difficulty of 38 questions in a multiple-choice A-level Physics examination was modelled, using codeable attributes of the questions and processes thought to be involved in answering. The **statistical outcomes of this process were successful**: using multiple linear regression and Linear Logistic Test Modelling at least 66% of the variability in question difficulty could be accounted for by the question features. However, despite the success of the statistical prediction, **a number of limitations** of the item difficulty modelling approach were identified in this context. Foremost was a sample size limitation. With a small sample of 38 questions, the robustness and generalisability of the difficulty model was low. For a general qualification context where questions are rarely banked or re-used, this limitation potentially constrains the usefulness of the item difficulty modelling approach unless data from more items can be linked. For systems where there is significant re-use of items there is potential for item difficulty modelling to provide greater insights.