Investigating experts' perceptions of examination question demand

Tom Bramley

Paper presented at the AEA-Europe conference,
Limassol, Cyprus, 3-5 November 2016

Research Division
Cambridge Assessment
1 Regent Street
Cambridge
CB2 1GG

Bramley.t@cambridgeassessment.org.uk

www.cambridgeassessment.org.uk

Abstract

The demand of exam questions is a concept that needs to be distinguished from empirical difficulty, intended difficulty and perceived difficulty. Demand is usually regarded as a qualitative concept. Pollitt, Ahmed & Crisp (2007) defined demands (plural) as "…separable, but not wholly discrete skills or skill sets that are presumed to determine the relative difficulty of examination tasks and are intentionally included in examinations." One reason for investigating demand is in order to be able to predict empirical difficulty. If this prediction could be made accurately, then it could have application in standard-setting, standard-maintaining, pre-testing and item-generation contexts. Except in some very restricted domains, however, it does not yet seem that there is a good predictive relationship between a priori demand(s) and question difficulty.

However, in the UK context, the main reason for wanting to define demand has been for the purpose of making comparisons between individual questions and examinations in terms of their demand (or demands). For this purpose, demands have been conceptualised and operationalised through the CRAS scale. This is a tool by which expert judges make ratings of exam questions on a 5-point scale with respect to their Complexity, Resource requirements, Abstractness, Task Strategy and Response Strategy.

This study followed the approach to social science research known as Facet Theory (e.g. Borg & Shye, 1995) to formalise the definition of the concept of demand and to investigate the multivariate structure of ratings of exam questions using the CRAS scale. Ratings of short- and extended-answer exam questions in two contrasting subjects (English as a 2nd language, and Physics) were analysed with the relatively little known POSAC method (Partial Order Scalogram Analysis with base Coordinates). This can be thought of as a form of ordinal principal component analysis that creates a 2-dimensional representation of a higher-dimensional structure. Profiles of ratings across the 5 variables can be classified in terms of 'comparability'. Two profiles are comparable if one has the same value or higher (lower) than the other for every element. Two profiles are incomparable if one is higher on some elements and lower on others. The POSAC analysis positions profiles on a 2-dimensional grid such that the maximum number of comparable and incomparable relationships are preserved in 2 dimensions. The original variables (the CRAS variables) can then be interpreted in terms of how they partition the 2-D representation into distinct regions (e.g. Shye, 1985).

The clearest finding was that demand is an approximately unidimensional concept. All CRAS dimensions correlated strongly with overall demand, and the majority of pairs of

questions were comparable in both A level Physics and IGCSE English as a 2nd language. In Physics there was also evidence of a consistent pattern in the second dimension, which reflected differences in the perceived Abstractness of the questions. These findings emerged despite relatively low inter-rater agreement on the raw ratings.

In conclusion, the POSAC technique seems a suitable and even useful method for investigating the internal structure of CRAS ratings of examination question demand. It takes account of the ordinal nature of the data, can help in understanding its dimensionality, and allows it to be visualised in 2-dimensional form.

Borg, I., & Shye, S. (1995). *Facet Theory: form and content.* Thousand Oaks: CA: SAGE.

Pollitt, A., Ahmed, A., & Crisp, V. (2007). *The demands of examinations syllabuses and question papers.* In P. E. Newton, J. Baird, H. Goldstein, H.Patrick & P. Tymms (Eds.), Techniques for monitoring the comparability of examination standards.London: Qualifications and Curriculum Authority.

Shye, S. (1985). *Multiple Scaling. The Theory and Application of Partial Order Scalogram Analysis.* Amsterdam: Elsevier.