

Cluster Analysis using Multiple Choice Question (MCQ) Data

Mark Buckley-Sharp and several colleagues published many papers on MCQ assessment. This included methods of scoring; methods of question analysis; use for quizzes and self-learning; and question banking. The theme running through this work was that computers could be used to help teachers and the assessment process, and that material could be reused in different ways to enhance efficiency.

The idea that the response data in an MCQ examination could be the source for a cluster analysis of the candidates (as opposed to merely scoring or question analysis) arose in our group in the late 1960s, but had to wait for adequate computer power before it could be tested. Even when tested, there was the difficulty of knowing how to report the finding. It is rash to claim a genuine first, but although cluster analysis is a standard technique elsewhere, it proved impossible to find that anyone else had reported using this idea for educational purposes.

Finally, two papers were drafted in 1993 and submitted to Medical Education where (also under its previous title of British Journal of Medical Education) most of our previous work had been reported. Leaving aside minor comments, the referees felt that the readership could not assimilate this work, and it was declined. The papers have remained unpublished in a peer-reviewed journal, but are presented alongside this commentary with minor revisions from the original.

1. M.D.Buckley-Sharp (1993) Numerical Taxonomy using MCQ data
2. M.D.Buckley-Sharp (1993) Classification of Students by Numerical Taxonomy

Combined Abstract

Students at medical school are generally trained in a single stream: no attempt is made to tailor training to particular strengths or weaknesses. This contrasts with policies in primary and secondary education, whether students are streamed or taught in multi-ability groups. If it were possible to distinguish groups within a medical school intake, then it could follow that training should reflect this grouping. To be useful, either the groupings must be identifiably stable, or the special training should demonstrably reduce group differences. Cluster analysis (numerical taxonomy) is a method of identifying groups.

There are four distinct stages in cluster analysis

- **Choose Data.** The data comprises MCQ answer patterns from the examiner, and all candidates. Each alternative in each question provides a binary attribute (mark versus no-mark). The examiner is arbitrarily called candidate zero, but otherwise examiner and candidates have equal status at the start of clustering.
- **Choose a Distance/Similarity Measure.** Numerous such measures have been proposed. One requirement is for symmetry ie, the distance/similarity for AvB must be the same as for BvA: interestingly, this rules out the standard MCQ scoring formulae. The simplest distance measure is the proportion of binary disagreements, which is the average squared Euclidean distance and also the Hamming distance.
- **Choose a Cluster method.** Many methods have been proposed, and a weighted average method was chosen. At each cluster cycle, one fusion of individual or group is made, where the distance of fusion is the minimum of the set of all possible fusions. After (N-1) cycles, only one cluster remains. An analytical solution is

- suggested when the fusion distance required in a cycle shows a major increase over previous cycles.
- **Interpret the Clusters.** Any realistic interpretation is only likely within the last few cluster cycles. Validation requires consideration of the clusters by external criteria, which may lead to labelling: much as may be done in factor analysis.

Programming the cluster method for MCQ data was done by adding modules to the UT3 and SCORE9 series of programs developed by our group, and used for our other published reports. Modules includes those for computing the distance/similarity matrix; for fusing the matrix into clusters; and for reporting the results as a dendrogram. The matrix is of order N square, and is hence large in these analyses.

Having evaluated the available methods, chosen appropriate ones, and developed the programs to run them, two examinations were studied - a Pre-Final practice examination in Medicine, and a Final MB examination in Medicine. The choice of two examinations in the same subject and very close in time was deliberate. Where any clusters are found using one examination, then to be useful, the same clusters (by membership) should be found in another examination. Failure to match in this case would suggest that the clusters would always be too unstable to be of any use in course design or teaching.

The Pre-Final examination clustered best into two clusters. The Final examination clustered best into three clusters. The membership of two of the Final clusters corresponded well with the two Pre-Final clusters. The membership of the third Final cluster corresponded well with those not taking the Pre-Final examination ie, with a null group.

This work shows that cluster analysis (numerical taxonomy) can be applied to MCQ data, using a choice from the standard methods, which appears to be valid. While cluster solutions must be treated with caution, two examinations close in both subject matter and timing demonstrated good cross-membership in cluster solutions obtained independently from each ie, there was some stability in the classifications. This in turn shows that cluster analysis (numerical taxonomy) is not ruled out as an analytical method in educational assessment. As far as can be determined, this method of analysis of MCQ data has not been previously described.