Research Report

Differential Item Functioning analysed in the context of Multi-modal Learning

Dr Peter Lam, School of Education
Ms Chiang Lian Hwa, part-time staff, NIE

The following project is one of the on-going research in psychometric applications from facilities available in the NIECER-Research funded Psychometric Laboratory which was set up in 1995

Differences in performance in Mathematics between two groups of students, such as between males and females or between majority and minority groups have often been reported in research literature. Sometimes the differences in performance have been attributed to real group differences. At other times, some test items have been designed in such a way as to favour one group over another. The language in which a test item was written or even the context in which the question was set may make the question more difficult for one group than another. For example, a question such as, "Find the difference between 10 and the sum of 2 and 3" may present a higher level of difficulty to an examinee not as competent in the English language as another even though the arithmetic involved in this question is quite simple.

The present study attempts to detect patterns of differences in the Mathematics performance of Malay and Chinese Primary Four students. The study also attempts to identify the characteristics of differentially functioning items that may help to explain the difference in performance between the two groups of pupils. The test instrument used was the Test of Basic Numeracy designed by Lam (1994).

Differential item functioning (DIF) procedures have been used for detecting test items which tend to favour one group over another. Such procedures have been used in item bias screening procedures for detecting flawed items in tests.

Three statistical methods were applied to the data-set to flag items which exhibit DIF. The three methods were:

1 Mantel-Haenszel (MH)
2 Logistic Regression (LR)
3 IRT-Logistic Regression (IRT-LR)

Based on the results of the flagged items, a qualitative analysis was carried out. Specifically, interest was centred on the kinds of process skills assessed by each item. An attempt was made to address ethnic differences in the solution of these problems within the framework of the Theory of Multimodal Learning (Biggs and Collis, 1991; Collis, 1992) and to look beyond for possible implications for classroom instruction and test setting.

Results showed that the Mantel-Haenszel and Logistic Regression procedures were easier to perform than the IRT-LR procedure. The latter was more computer-intensive as each item had to be studied separately for the presence of DIF. However, the results obtained by the IRT-LR procedure were more informative. This was because the IRT-LR method did not just flag items for DIF. It was also able to produce ICCs for each item. Where DIF existed, the superimposition of the ICCs for both Focal and Reference groups gave the researcher a very clear visual display of the differential behaviour of any DIF item.

The following observations were made:

1. Chinese students in this sample performed better than Malay students on school-taught Mathematics at and beyond the concrete-symbolic mode.
2. DIF was observed to occur in favour of the Malay students on items that were judged to be solvable by examinees who were operating at the iconic mode.
3. DIF could also be due to other confounding factors such as language. In this respect, one may be able to distinguish between mathematical and non-mathematical terms.
4. DIF may also result from strict adherence to "school-taught" (concrete symbolic) methods, with a neglect of the iconic mode of representation.

The use of the Theory of Multimodal Learning was advantageous in that it was able to sift out a trend of differential item functioning which may otherwise have gone unnoticed. It did seem that one group of students was able to operate better at the concrete-symbolic mode. There also appeared to be some indication that students well-versed with school-taught mathematics may not necessarily be as adept at operating in the iconic mode. This certainly has implications for curriculum and teaching.

References

Biggs, J.B., and Collis, K.F. (1991). Multimodal Learning and the Quality of Intelligent Behavior. In Helga A.H. Rowe (ed.), *Intelligence — Reconceptualization and Measurement.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Collis, K.F. (1992). Curriculum and assessment: A basic cognitive model. In Gilah Leder (ed.), *Assessment and Learning of Mathematics*. Victoria, Australia: ACER.

Lam, Peter (1994). *Test of Basic Numeracy* (Unpublished test material).