Title	How teachers can construct multiple choice questions with meaningful test results
Author(s)	Cheung K. C.
Source	*Teaching and Learning, 13(1), 71-76*
Published by	Institute of Education (Singapore)

How Teachers can Construct Multiple Choice Questions with Meaningful Test Results

K. C. CHEUNG

A. What teachers learn in traditional Test Construction courses?

Around the world, Test Construction is a core course in most teacher education programmes. Teachers typically learn the many *purposes of testing.* The significance of the *timing of testing,* whether conducted before, during, or after instruction, is underscored for its linkages with the specific purposes of testing. Five purposes which are commonly encountered by the teachers are listed below.

- *selecting* students for placement;
- *diagnosing* and *charting* student progress;
- *grading* student attainment level for promotion;
- *reporting* to parents;
- *certifying* to employers and institutions.

In Test Construction, *test validityis* a core concept. It is ensured through careful *test specification* and *standardization* of test administration procedures. Important considerations include the following five aspects.

- scope and depth of *content coverage;*
- item and test *format;*
- testing *time;*
- examination *setting;*
- invigilation and scoring *procedure.*

Content coverage can be a difficult decision for the teachers. For a given test length and amount of testing time, there are trade-offs between the range of content (ie. bandwidth) and precision and authenticity (ie. fidelity) of what is intended to be tested. An optimal balance between bandwidth and fidelity is required for the construction of a test plan. One test construction method commonly taught to the beginning teachers is to classify the topics/areas against categories of the Bloom's Taxonomy (eg. Recall, Comprehension, Application). Items are then constructed according to this blue-print and the specific instructional objectives as laid down in the programme of study. Amongst the various item formats, the multiple choice question (MCQ) remains very popular nowadays. This is because of its administrative convenience – it can be objectively scored as either right or wrong and a total score can be summed easily to represent the level of ability of the student. Unfortunately, some teachers do not realise that students' responses to the various options of the MCQ can provide useful diagnostic information, Instead of being scored dichotomously, the options can be graded according to their degree of correctness cr completeness as an answer.

Teachers do not routinely assess the quality of the items they are able to construct. Two popular item statistics, item difficulty and item discrimination, are useful to them. **Generally speaking, item discrimination should be as high as possible because this would contribute to the reliability of the total score as an estimate of a student's attainment level.** For a full-fledged understanding of item discrimination, teachers can compare the proportion of the high-scoring group (eg. top 30%) with the low-scoring group (eg. bottom 30%) choosing each option of an item. For most teachers, the most desirable range of item difficulties is a difficult decision because of its dependence upon the purpose of testing. For example, if the purpose is to rank students, the test should comprise of items with a full range of item difficulties, say from 0.2 to 0.8 (ie. indicating 20 to 80% of students passing an item). **As a general guide, the items should have levels of difficulty targeted at those students who should be measured accurately.**

B. What teachers do not learn in traditional Test Construction courses?

What teachers do not learn in traditional Test Construction courses and what they found most intricate can be summarised by the following two inter-related questions.

- What attributes make up item difficulty?
- How to control the difficulty level of an item?

Subject-based research, such as mathematics and science, has repeatedly shown that item difficulty is a generic concept and is very sensitive to how the questions are framed/phrased. This sensitivity may be a consequence of the following:

- a change of *question context* as perceived by the students (eg. use of commonsense when questions are phrased in everyday context);
- a change of *test-taking strategy* (eg. the question invites guessing because of the inconsistencies of MCQ option formats);
- inadequate consideration of students' *opportunity to learn* the material being tested (eg. content included in the test is not adequately covered in the course).

A resolution of the above two thorny questions poses the need for *meaningful testing* in the classroom – **a strive for test results being capable of affording meaningful interpretations for the purposes of understanding how a student progresses from a low level of understanding to the higher levels so as to help a student overcome learning difficulties.**

C. How can teachers move from traditional to meaningful testing?

At present, teachers use the total score to compare students' ability – the higher the total score, the more ability the student possesses. However, an item is found difficult for many reasons. Two of these are important for meaningful testing.

- a difficult item affords the use of higher order concepts and skills by the more mentally-developed students;.
- a difficult item affords adaptation of students' knowledge when applied to novel situations.

Consequently, descriptions on how concepts and skills progress from lower-order into higher-order ones are needed to design a test. If the MCQs are on application, features discriminating between experts and novices should be incorporated into the test as well.

In MCQ, the options which are not the key (ie. those options scored as zero) are commonly known as *distractors.* It is common to find that teachers do not know how to decide on the *optimal number of options* and how to *minimise guessing.* One advice found in traditional Test Construction guidebooks is to make all distractors equally plausible and attractive to examinees who lack the knowledge referenced by the question.

As discussed earlier, distractors contain diagnostic information for overcoming learning difficulties. One insight from recent cognitive science research is that the distractors can comprise of the *alternative conceptions* of the students, or in similar vein, the *alternative routes* taken to solve a problem task by the novices and experts. As such, the number of options required for each item is not uniform and is the same as the number of predominant alternative conceptions, and therein maximum qualitative item discrimination is achieved by outcome. Since by definition alternative conceptions are conceptions viable to students, guessing and perseverance are no longer key factors determining success on an item. Test performance no longer depends on luck and effort. Test validity, an evaluation of congruence between concepts and skills as applied by the students during testing and those intended to be afforded by the questions, is thus greatly enhanced.

It should be noted that meaningful testing does not preempt the traditional wisdom of test construction. Ideas of test plan, item difficulties and discrimination are still applicable. Meaningful testing serves to reiterate that MCQ testing is an art – in the design of items with a range of item difficulties in accordance with concept and skill development

and manifestations of alternative viable conceptions as distractors. This resolves the perennial dilemma between bandwidth and fidelity of a test score designed for informing how to overcome learning difficulties. Most important of all, meaningful testing capitalises on the professional expertise of the teachers.

D. How to construct a multiple choice test on the part-whole concept of fractions affording meaningful interpretations of test results?

An example is presented here to illustrate how a MCQ test on the part-whole concept of fractions can be constructed using principles of meaningful testing. Four sequential steps need to be considered.

1. Try to figure out what concrete instances of fractions in everyday settings are experienced by the students being tested (eg. sharing a pie between two brothers; dividing a box of pencils amongst a small group of children). Consider instances that are personally relevant to the students.

   Try to figure out what basic concepts and skills contribute to relating the parts to the whole in order to form the notion of a proper fraction (eg. What is the whole? What constitute the parts of the whole? How to partition the parts of the whole? Should the parts of the whole be of equal-sized? How to count the parts in relation to the whole?). At the same time, try to find out in what ways alternative concepts and skills are regarded as less viable1 correct for an understanding of the part-whole concept of fractions (eg. no recognition that the groupings of pencils, and partitions of a pie are of equal-sized).

3. Try to figure out how concrete instances of fractions can be represented both schematically and symbolically such that key concepts and skills can be identified (eg. sharing a pie between two brothers – use a circle to represent a pie and a diagonal line to cut it into two halves). At the same time, assign the number of equi-sized shaded parts as numerator and the total number of equi-sized parts as denominator of a fraction (eg. for a fair share of a pie between two brothers, each will get 1/2 of the pie when the cut passes through the centre of the circular pie).

4. Construct items that can be grouped at progressive levels of understanding (eg. a unit fraction is easier than a non-unit fraction; a representation that can be counted more readily is easier than one in which the partitions may not be readily seen as equal-sized; a representation with parts shaded in a random fashion is more difficult because of the difficulty in grouping the parts mentally together). Supply the alternative conceptions as distractors (eg. relating shaded parts to the unshaded parts showing misunderstanding of the concept of the whole; the partitions/groupings are not of equal-size showing misunderstanding of equal-partition).

In this way, items are of a range of difficulties which is in line with the progressive levels of understanding needed to solve them (see Figure 1). Although meaningful testing initially looks formidable to the teachers, it can be concluded that it provides meaningful diagnostic information for overcoming learning difficulties.

Reference

Cronbach, L.J. (1984). *Essentials of Psychological Testing (4th Edition).* New York, Harper and Row.

Cheung, K.C. (1991). *On Meaningful Measurement: Concepts, Technology, and Examples.* CARE Research Paper No. 3, National Institute of Education, Singapore.