

EXAMINERS AND CURRICULUM DEVELOPERS

Trevor Rees

In attempting to meet the criteria of validity and reliability for national examination questions, examiners have no problem, in principle, in acknowledging the extensive earlier input into the learning process by experts in curriculum development. It is unfortunate, therefore, that when it comes to setting classroom or national test papers, teachers and examiners may feel themselves in some difficulty due to thoughtless earlier curriculum-led statements in the area of assessment guidelines.

There are a number of reasons for this tension, but probably three main ones. In the first place, many syllabuses are too long. This is probably attributable in part to the almost exponential growth in contemporary knowledge, which tends to put increasing pressure on particular segments of organised knowledge. Instead of keeping the amount of content manageable (ie appropriate to the time available in the average school) syllabuses tend to get longer regardless. One consequence in summative examinations is that students may frequently be asked to choose between a variety of questions on the same test paper. Any such series of choices carries the technical disadvantage that the validity of the ensuing rank order of students is adversely affected, since several different tests may be at work at the same time. The obvious remedy would be to keep the content of syllabuses within manageable limits so that a proper sampling can be made by examiners. My own best solution would be to discourage choices of any sort in examination papers, and to insist that test papers sample syllabuses properly. I believe that, to be valid, a test instrument, especially where the results lead to a rank order, will set the same tasks for all can-

didates. If choices are allowed, validity (and reliability!) is undermined and, of course, will be most unfair to those students unfortunate enough to have chosen the more difficult options.

Secondly it is a matter of slight debate that examination prescription statements are often flawed. Most Pacific educators are familiar with the sort of "prescriptions" which tell the examiner in advance that an examination is to be two hours in length; shall consist of 25 multiple choice questions (each carrying one "mark"), two matching items and an essay. The paper shall be out of 100 "marks". And so on... Quite frankly, this is a sure recipe for an unbalanced and distorted test instrument. A good examiner will not know in advance such specifics until a test plan or "blueprint" has been prepared. Only after this has been done, and a careful sampling of course content undertaken, will an examiner be in a position to determine such critical factors as paper length, item types, cognitive and other levels, skills/ content to be tested, weightings and then (and only then) the number of "marks" the overall paper will carry. Incidentally, there is nothing magical about the number 100. A test can be "out of" any number of "marks", even 129 or 251! You won't be surprised to find a hardened assessment Board Director like myself suggesting that other professionals should leave the planning of the examinations to the examiner.

"Most Pacific educators are familiar with the sort of "prescriptions" which tell the examiner in advance that an examination is to be two hours in length; shall consist of 25 multiple choice questions (each carrying one "mark"), two matching items and an essay..."

A third cause of disarray is that syllabus statements can and sometimes do confuse norm and criterion referenced approaches to testing. Currently there is a clear move away from the practice of assessing a

student's acquisition of syllabus-defined skills and knowledge with reference to other students, a procedure that implies norm-referencing. Alternatively, student performance may be measured against stated objectives with the level of achievement determined by reference to explicitly stated sets of criteria. This, of course, is a standard account of criterion-referencing. Unless curriculum writers and teachers not only distinguish between the two approaches to assessment, but also have a sense of when both might be deemed appropriate, a good deal of confusion is likely to occur. At worst, an entirely invalid alleged summative assessment will be perpetrated if the two disparate forms are unwittingly combined into a single so-called "result". The problem here, properly under-

stood, carries its own solution: reports of student assessment should never mix the two approaches; norm and criterion referenced testing need to be kept separate.

Space precludes my offering further examples of an apparently growing mismatch between the syllabuses of the curriculum developers and the ways in which such schemes are assessed in classrooms and national examinations. But even this brief analysis may be correctly interpreted as an appeal for much greater professional contact between those people who determine what students are to learn and those who have to assess validly how much of that learning has taken place.