

## Training for National Examiners: Where Does it Start?

*Graham A. Wagner*

In the South Pacific there is much debate about what will happen when the New Zealand School Certificate and University Entrance examinations are no longer available as indicators of success in schooling. National leaving examinations of these kinds, with their widespread local currency and universal acceptance serve the time-honoured task of certifying that candidates leaving secondary school are either up to standard or not. The standard is determined not so much by what the candidate knows (or has learnt) but by how well he, or she, does in comparison with other candidates who also sat that examination.

It is entirely conceivable that a situation could arise where, for example, in a particular national examination subject paper questions were asked that bore very little, if any, relevance to the syllabus taught in the schools. Candidates marks would not necessarily indicate what they knew about the content of the syllabus, but the actual spread of their marks probably pattern the normal distribution curve, with most marks grouped around the centre of the range, tailing off in a similar fashion on both sides of the curve. This means that examiners could award pass and fail grades, for that subject, by simply taking the mid-point, or beginning of the distribution of all candidates scores, and by using standard deviation or ranking methods award predetermined percentages of grades or marks.

This scenario is not that far-fetched, for it is essentially the approved practice in national examinations to write questions that discriminate between high and low achievers; unfortunately, in most cases, without automatic reference to the questions' relevance or the examinations' overall content validity. Why this can happen is easy to see when it is realised that it is usual for bright students to score higher on any given test than the not so bright students, given of course that those being compared have equal exposure (or lack of it), to prior subject knowledge, irrespective of the content being tested. The point being made here is that it is not necessarily what a candidate knows that counts, but how much more, or less, he or she knows than those they are being compared with.

This kind of examination is called a norm-referenced examination because it allows the examiners to compare the candidates' performance with other

candidates from the same reference (or norm) group, and is the dominant method of measuring and interpreting academic achievement in the South Pacific region.

### **Qualities of a national examiner**

While norm-referenced examinations have relevance as selection examinations in cases where only a few candidates, for example, can be selected for access to limited positions for further training, they are not the most appropriate means for assessing school-leaving levels of competence. They also have the disadvantage that they can be misapplied by ignorant, or less than adequately trained, examiners.

In this sense 'adequately' means more than just being a subject expert, it includes being a trained and experienced teacher as well. Someone, in other words, who knows intimately the subject matter to be examined, how students learn, what can be expected of them in an examination of the kind to be employed and, most importantly of all, someone who has a facility in the applied skills of planning, setting, marking and interpreting examinations, at least of the classroom variety. (For a national examiner, one might even say 'considerable' facility in such examination skills.)

That is not all, because the skills of a national examiner are built on, or extend, the skills of a good classroom teacher, so advanced examiner training, of the kind provided by the South Pacific Board for Educational Assessment, is a prerequisite before being 'let out' to 'practice for real'. If national examiners are properly selected and trained, the distinction between norm-referenced examinations and criterion- (or domain-) referenced examinations (the latter being concerned with what the students know or can do) will be placed in proper perspective, and the examiners will be more likely to be concerned about the content validity of their examination papers than they are at present.<sup>1</sup>

Generally speaking, national examination questions are related to national subject syllabuses so there should be no major concern that the subject-matter taught over all those years, leading up to the national leaving examination is not going to be tested in the end. Of course some of it is. However, not very much can be covered in one three hour examination paper. Therefore a justifiable criticism of national examinations is that examiners can only draw a very small sample of questions from a huge pool of potential questions. Through this limited

sample method examiners must estimate the overall competence of the candidates and their knowledge of the subject-matter they have been studying so diligently over a long period of time. If the examiners are not very astute at picking the key questions to indicate subject competency, or if they make the inexcusable mistake of including faulty examination questions which either disguise the examiner's real intent or give obvious clues to the right or wrong answers, then the examination results are as misleading about competency as in the scenario depicted in the second paragraph of this paper.

### **Current national examiner malpractices and inadequacies**

Unfortunately, it is easy to cover up examiner incompetence, or inadequacies in national examinations, by adjusting or scaling marks up or down, because the individual scores can usually be made to approximate a normal distribution curve, which is what one might expect from the results of any achievement test on a large population. (Remember the point already made about high ability and low ability students and the way they spread along an achievement continuum?). Incidentally, have you ever heard of candidates being asked to resit a totally inadequate national examination paper even though there may have been a public outcry by teachers and parents about the inappropriateness of the questions? The reason nothing often happens is that the really damning evidence is always kept in the hands of the examiners. They have been known to bluff their way out of such public furores by claiming that they have a greater understanding of such matters than laymen. All they have to do is to invoke the mystique of statistical manipulation to, they say, 'right any wrongs'. Such public disclosures do not happen very often, but often enough regrettably to indicate the tip of an iceberg and to show that, although we have reached a high level of sophistication in measurement evaluation theory, the level of practical expertise among many national examiners is probably well below par.

One of the principal reasons for invalid national examination papers lies in the calibre of the national examiners themselves. Most examiners have come from, or are still in school classrooms around the region. Whether or not they have been trained in measurement theory at university or teachers' college, it is likely that they have had little opportunity to try out any of the approved classroom measurement techniques, such as item-writing or item-analysis, before they were overwhelmed by the sheer magnitude of routine teaching tasks and classroom management. Small wonder then

that most classroom teachers in the USP region require in-service classroom measurement training courses which can do no more than revise what has been covered in pre-entry teacher training.

Unless potential national examiners have had some kind of advanced in-service training in classroom measurement, and have been participants in the SPBEA's regional examiner's training programme, it is likely that their measurement and evaluation skills will not be adequate for the task of being a national examiner.<sup>2</sup> They will often just get by (by training themselves at great effort and considerable sacrifice) but their examination results will probably tell us nothing more than that they have once again ranked candidates in a predictable fashion and that these candidates know something about the examination subject. If you were to ask one of these examiners to give you a breakdown, in an examiner's report, of the candidates' strengths and weaknesses then you would probably be disappointed with the information they came up with. The point is that all examiners, but especially national examiners, should be able to tell you not only how many candidates passed each examination question, how difficult the items were, how much the items discriminated between upper achievers and lower achievers, whether the test as a whole was reliable (consistent) enough but, more importantly, whether the test as a whole was valid or not. In other words, whether the test best measured the content, or subject matter, that it is supposed to measure in the time allowed.

### **Importance of content validity**

Content validity is the most important concept in all classroom testing, not least of all in national examinations. It is no good referring to the statistics of an item analysis print-out, good though these may be, without first checking on the validity of the examination questions. This can be done, in national examinations, by using what we call a two-way examination blueprint, or planning grid, to check the relationship between content and levels of thinking (that is, low-level thinking like 'recall' to high level thinking like 'evaluation'). It is hard to conceive of any other way of checking the content validity of an examination except by using this method. Yet how often do you hear about examination blueprints in national examinations, let alone see one?

In recent times the New Zealand Education Department has published examination specification data in its *Education Gazette* which, in most subjects over the years, has included a model examination blueprint and

thus a guide for chief examiners. It is doubtful that many examiners in the region can prepare examination blueprints, know how to interpret them, or use them to determine strengths and weaknesses in candidates' answers. Without the use of blueprints examination question setting will only be a 'hit-and-miss' affair with every likelihood of an unrepresentative sample of questions being selected, resulting in the production of an invalid paper (that is, a paper that will not be measuring precisely what it was supposed to measure). Yet, no one except the examiner will be aware of the real extent of the paper's invalidity in such cases.

The reason for misgivings about the general competency of national examiners is based upon the widespread observation in the region that the long held concern by teachers for their students to pass the approved school leaving examinations has stunted the teachers' development of a skill in a different kind of classroom measurement which is designed, first and foremost, to facilitate students' learning, not just to help them pass or fail examinations. This kind of classroom measurement is different from national school leaving norm-referenced examinations, but, let there be no doubt that they both serve complementary purposes at different stages in the schooling process. Without recourse to both methods of examining, the teacher/examiner is like a car mechanic without a full set of tools with which to repair a brokendown car. However, skill in one should precede the other in the proper preparation of competent national examiners.

### **Criterion-referenced examinations**

In contrast to norm-referenced examinations (which use the performance of the total group to measure the performance of one of the group), the classroom teacher's criterion-referenced examinations have more utility in school learning. Criterion-referenced examinations are designed to measure whether a student has reached a previously set level of achievement. A student's score on this kind of examination is independent of the scores of other members of the class or group that took that same examination. What is more important, from the educational point of view, is that criterion-referenced examinations tell the teacher or student precisely what is if they know, or do not know. In other words, they provide an assessment of the students, knowledge of the content which was, or should have been, covered during teaching and learning. They are thus far more content-related than the very general highly selective norm-referenced national leaving examinations. Because of this, criterion-referenced examination results have more meaning, for they can be used

to diagnose students' learning strengths and weaknesses, as well as to certify levels of competency.

By design, criterion-referenced examinations are narrower in scope than the very broad based norm-referenced national examinations. Nevertheless, unless teacher/examiners have been brought up on, and have fervently practised, criterion-referenced measurement principles in their classrooms, it is hardly likely that they will be good teachers who use their examination results judiciously to plan individual, and class, remedial or extension programmes. Furthermore, they can hardly become good national examiners who can be relied upon to select the best range of limited key questions to measure adequately examination candidates subject content knowledge and thus determine their suitability for a school leaving certificate.

Mention should be made here of the fact that criterion-referenced examinations are different from norm-referenced examinations in a number of significant ways from their purposes right through to how they are interpreted. These differences are summarised in Appendix I and elaborated on in texts in the reference section of this article. Suffice it to say that if an examination, be it a classroom one or a national one, does not cover an adequate or fair sample of the course or subject being examined then it is not a valid examination. Concommitantly, national examiners will not be able to produce valid examinations unless they have been properly trained in the principles of criterion-referenced examining prior to the specialised training they require for norm-referenced examining.

## **Conclusion**

We must cease perpetuating the notion that people who are proficient in subject content (that is, are knowledgeable about that subject) can also, by definition, teach or write suitable examinations in that subject. This is fallacious thinking, although there may be some rare exceptions. By and large, a national examiner needs to be trained to be a national examiner. Now that locally-based national school leaving examinations are on the increase in the South Pacific region, it is timely to look at the training of national examiners. In doing this, it is contended here that a qualified national examiner must be thoroughly trained in the latest classroom measurement theory and practice, and that central to their overall suitability is their ability to prepare and use valid criterion-referenced examinations.

## Notes

1. Most measurement and evaluation theorists and writers accept that as a term 'criterion-referenced' is misleading and that it would be more appropriate to use 'domain-referenced' to describe what we are really concerned with, which is the focus on content validity. Why there has not been a name change is explained, somewhat unconvincingly, in Propham (1981).
2. The USP has developed new courses in Educational Measurement, both on campus and at a distance, to meet the regional demand for more adequate measurement training. The University has also introduced a new Diploma in Educational Measurement which, it is assumed, will be taken by those preparing to be, or practising as, professional examiners.

## References

- Gage, N.L. and Berliner, David C. (1984) *Educational Psychology: Third Edition*. Houghton Mifflin: Boston (See pp. 654 - 661).
- Mehrens, William A. and Lehmann, Irvin J. (1984) *Measurement and Evaluation in Education and Psychology* (3rd ed.). Holt-Saunders International: Japan. (See Chapter 2).
- Propham, W.J. (1981) *Modern Educational Measurement*. Prentice-Hall: Englewood Cliffs: N.J. (See p. 30 in particular)
- Theobald, John (1974) *Classroom Testing*. Longman: Australia. (See Chapter 3, 5 and 6).

Appendix 1 — see over.