Title	Test planning and construction
Author(s)	Tan Wee Kiat
Source	Typescript of an unpublished manuscript

Citation: Tan, W. K. (1991). Test planning and construction. Unpublished manuscript.

TEST PLANNING AND CONSTRUCTION

Dr Tan Wee Kiat
Institute of Education
Singapore

Examinations are formidable, even to the best prepared, for the greatest fool may ask more than the wisest man can answer.

C.C. Colton 1836

In the informative book "Teaching and Learning in Higher Education" (Beard, 1970) published by Penguin, there is mention of different examination systems used in various institutions of higher learning. One examination, carried out by the Australian College of General Practitioners, consists of two parts. In the first part, which is written, there is a 3-hour essay on medical practice, a 2½-hour paper on a speciality of the doctor's choice (eg obstetrics, industrial medicine), a 2-hour multiple-choice test of essential information and a 3-hour programmed simulation test of patient management. The second part of the ACGP examination tests the doctor in action. In a simulated diagnostic interview a member of the college plays the part of a normally difficult patient. This is followed by a simulated patient management conference in which the contributions of each examinee are assessed both for effectiveness in working with other participants and with respect to medical knowledge. Finally, there is a practical test in which X-rays, colour transparencies, etc, are viewed and the examinee is required to comment on diagnosis, treatment and prognosis for a patient.

An example of a different type of examination system is that carried out by the English Department in a U.K. institution. Here the examination system consist of 5 types. (i) Three-hour examinations with textbooks and dictionaries allows candidates to work against the clock and provide evidence that the holder of a good honours degree is not "paralytically incompetent to produce respectable work in a hurry". (ii) A paper of stated length prepared during 14 days allows unrestricted use of books. Some students are said not to sleep for the last two or three days because of this paper. (iii) Long essays are used to allow students to write about their optional topic: this, it is felt, gives students opportunity to make significant and fresh discoveries whilst working in a chosen field. (iv) His five best tutorial essays are also selected by each student and three of them are assessed by four wholly independent examiners; the remaining two are studied in the event of doubt or disagreement. (v) Finally a form of viva is employed which consists in a continuous assessment of oral contributions to seminars and tutorials throughout the term. The effect of this on seminars has proved entirely good since students realized that nothing would gain marks except responsive contribution to relevant discussion. It is reported that the assessors found it necessary and possible to give credit for 'the ability to hold one's peace' and that it was even possible to distinguish between different qualities of silence during seminars!

The two examples used here serve to illustrate the relationship between the objectives, the curriculum, and the examination system. Firstly, the objectives must be clear, definable, desirable and attainable. Secondly, the curriculum is designed to achieve these objectives. Thirdly, the examination system is used to find out whether course participants, through this curriculum, have reached the objectives.

Assuming that the objectives and the curriculum are well-thought out, the following questions are then asked:

What will the examination system be comprised of?
What is the weight that will be attached to each component?
In the overall assessment what constitutes a Distinction, Credit, Pass and Failure?

Whatever kind of examination system is considered the following criteria should be applied to it.

Criteria for judging a test

1 Relevance

The test items should be directly related to the course objectives and curriculum used. The consideration of relevance increases the probability that the test is valid by posing the question:

Is the test measuring what it is supposed to measure? If a test item is in line with the objectives but was not covered in the curriculum there must be strong justification for its inclusion.

2 Balance

A test shows balance if the proportions of test items dealing with each aspect of achievement conform with the test constructor's specifications. To meet this criterion, a Table of Specifications is used. In this table the test topics are listed vertically while the desired levels of achievement are listed horizontally. The number of questions that are to be generated for each cell should reflect the importance of that cell (eg as indicated by the time spent on it during the course).

3 Efficiency

A test programme that yields a large number of independent test scores per unit time is an efficient programme. However, some compromise between planning, construction, test administration, scoring and cost has to be made.

4 Objectivity

The test constructors must agree on the "right" or "best" answer to each particular test question. Ideally, the "right" answer should be worked out beforehand so that it can be used as a model for assessing the students' answers. Thus, objectivity is more a function of the answers rather than the mode of examination. In this context "objective tests" are more objective than "essay tests".

5 Specificity

The test questions should be such that subject-matter experts receive near-perfect scores while course-naive students receive poor scores. Such differences should not be because the test questions are examining trivia.

6 Difficulty

The test questions should be appropriate in difficulty level for the group being tested. Generally, the test constructor aims at questions of average difficulty, ie, where fifty-percent of the students can answer each question correctly.

7 Discrimination

This refers to the ability of a test question in differentiating between high scorers and low scorers. For the purpose of classifying students in order of merit the test must contain many questions of a discriminating nature.

The criteria of difficulty and discrimination form the basis of test analysis.

8 Reliability

This concerns the consistency of measurement especially when different markers are assigned to mark the same question. In this context the objective test is a more reliable measure than the essay test.

9 Fairness

The test is such that it will allow each student an equal chance to demonstrate his ability. For example, requiring a science student whose dominant language is Chinese to take a viva in English is likely to place the student at a disadvantage compared to students with an English school background.

10 Speededness

The time allowed for a test should be such that all or, at least, most students to finish the test.

It is likely that the present examination system involving oral, written and practical tests will prevail for some time. In any case even when the examination system is modified, such as the ACGP examinations, oral, written and practical tests are still included. It is useful to consider the principles underlying such tests.

Principles underlying the use of Oral Tests (Viva Voce)

1 Oral tests should be used (a) to obtain information as to the depth of student's knowledge, (b) where oral presentation is clearly a purpose of the course, and (c) where other means of assessment are simply inappropriate.

2 Prepare in advance a detailed outline of materials to be sampled in the examination, even to the extent of writing questions which will be asked.

3 Pose questions which students with the training which has preceded a particular examination can reasonably be expected to know. An examination is not the place for an instructor to demonstrate his own erudition.

4 Where several examiners are involved, each one should be responsible for questions on a specified part of the full examination.

5 Determine in advance how the scoring system and records of student performance will be kept and what weights will be assigned various factors.

6 Develop some facility with several basic techniques for successful oral examining, such as (a) creating a friendly atmosphere, (b) asking questions, (c) recording responses unobtrusively, (d) avoiding arguments with the student.

7 Do not spend a disproportionate time probing for the answer to one question. If the first several questions do not elicit the desired response, move on to some other matter.

Suggestions for constructing and assessing essay examinations

1 Use an essay question for the purposes it best serves, ie

- 1.1 Explain cause-effect relationships
- 1.2 Describe applications of principles
- 1.3 Present relevant arguments
- 1.4 Formulate tenable hypotheses
- 1.5 Formulate valid conclusions
- 1.6 State necessary assumptions
- 1.7 Describe the limitations of data
- 1.8 Explain methods and procedures
- 1.9 Produce, organize, and express ideas
- 1.10 Integrate learnings in different areas
- 1.11 Create original forms (eg designing an experiment)
- 1.12 Evaluate the worth of ideas

2 Prepare enough questions to sample the material of the course broadly, within a reasonable time limit.

3 Require all students to answer all questions on the test.

4 The questions should be structured.

5 Determine in advance how much weight will be accorded each of the various elements expected in a complete answer.

6 Do not judge papers on the basis of external factors unless those have been clearly stipulated.

7 Without knowledge of students' names, score each question for all students. Use several scores and scorers if possible.

Checklist for reviewing objective test questions

An objective test may be defined as a series of questions each of which has a pre-determined correct answer so that subjective judgement in the marking of each item is eliminated.

1 Content is trivial. Does the item deal with facts, principles, understandings, skills, etc, of sufficient importance to merit inclusion in an achievement test?

2 Content is too factual. Does the item measure sheer recall of relatively unimportant material or does it call for knowledge and understanding of important facts, concepts, etc?

3 Item is too easy or too difficult.

4 Vocabulary level is not appropriate for designated grade levels. Is vocabulary too technical, too literary, too textbookish, etc?

5 The statement of the problem is badly worded. Does the item contain superfluous information? Might a student who understands the issue involved be likely to miss the item because of its wording? Is the wording ambiguous?

6 Optional responses are not plausible. Can the student quickly eliminate one or more responses because they are unrelated to the problem?

Performance testing

As an example consider the type of practical work which can be divided into four main points:

i) Use of tools or equipment
ii) Safety procedures
iii) Speed at which a man does the job
iv) Quality of the completed job

Suppose the job is worth 20 marks. Assume each of the above points is worth a maximum of 5 marks. Each point to be assessed can therefore be allocated a mark ranging from 0 to 5.

Examiners Assessment Guide

USE OF TOOLS	Does not start job; abuses tools and equipment.	0 mark
	Occasionally misuses a tool; or uses the wrong tool.	1 mark
	Uses right tools, but incompetently, inaccurately and hesitantly.	2 marks
	Uses right tools with reasonable accuracy and some hesitancy.	3 marks
	Uses right tools, handled correctly, occasional hesitancy.	4 marks
	Uses right tools, handled correctly, accurately, confidently and positively.	5 marks

SAFETY PROCEDURES	Does not start job; fails to observe two or more safety procedures which could endanger the public, workmates, himself or any equipment, plant or property.	8 marks
	Fails to observe one safety procedure which could endanger the public, workmates, himself or any equipment, plant or property.	1 mark
	Observes all safety procedures but with four or more minor breaches which would not endanger the public, workmates, himself or any equipment, plant or peoperty.	2 marks
	Observes all safety procedures but with two or three minor breaches which would not endanger the public, workmates, himself or any equipment, plant or property.	3 marks
	Observes all safety procedures with an occasional minor breach which would not endanger the public, workmates, himself or any equipment, plant or property.	4 marks
	Observes all safety procedures.	5 marks
SPEED	Does not start job.	0 mark
	Disorganised approach, frequent halts or hold-ups, erratic.	1 mark
	Little method in approach, working rate inconsistent.	2 marks
	Methodical, maintains steady pace.	3 marks
	Methodical, works consistently at brisk tempo.	4 marks
	Methodical, works consistently at fast tempo.	5 marks
QUALITY	Does not start job, attempts job but two or more defects which could affect electrical and mechanical standards.	0 mark
	Attempts job but one defect which could affect electrical and mechanical standards.	1 mark
	Attempts job but four or more minor defects which would not affect electrical and mechanical standards.	2 marks
	Two or three minor defects which would not affect electrical and mechanical standards.	3 marks
	Occasional minor defect which would not affect the electrical or mechanical standards.	4 marks
	Faultless job.	5 marks