

test and measurement techniques

4

the practice of test and measurement techniques

This fourth chapter reviews a certain number of measuring instruments, their advantages and limitations, and how they should be used. It shows the wide range of such instruments and the importance of choosing them in accordance with the educational objectives to be measured. It is stressed that every teacher should make a certain minimum of statistical calculations when he carries out an item analysis of a test.

Those with a deeper interest in these problems are strongly advised to consult the following publications:

Public Health Papers – WHO

- No. 36 A review of the nature and uses of examinations in medical education, 1968
- No. 52 Development of educational programmes for the health professions, 1973
- No. 72 Assessing health workers' performance: A manual for training and supervision, 1980.

After having studied this chapter and the references indicated, you should be able to:

1. Indicate the different elements that should be considered in the evaluation of a teaching programme.
2. Indicate the different elements that should be considered in the evaluation of the educational objectives of a teaching programme.
3. Define the advantages and limitations of a system of evaluation of teaching *by* the students.
4. Construct an observational rating scale and/or a practical test to evaluate the behaviour of a student in the domain of communication and/or practical skills.
5. Propose a question for a written (open-book) examination of the "essay" type or a series of six short, open-answer questions and indicate the norms of performance permitting objective marking (marking table).
6. Draw up three multiple choice questions (MCQ) in the domain of intellectual skills – at least two of the objectives must measure an intellectual process superior to level 1 "simple

recall" (either level 2 "interpretation of data" or level 3 "problem-solving")

7. Indicate the advantages and limitations of a programmed examination.

8. Define the following terms: prerequisite level test, pre-test, interval test, comprehensive, pre-final, indicate their purpose and the stages at which they are set

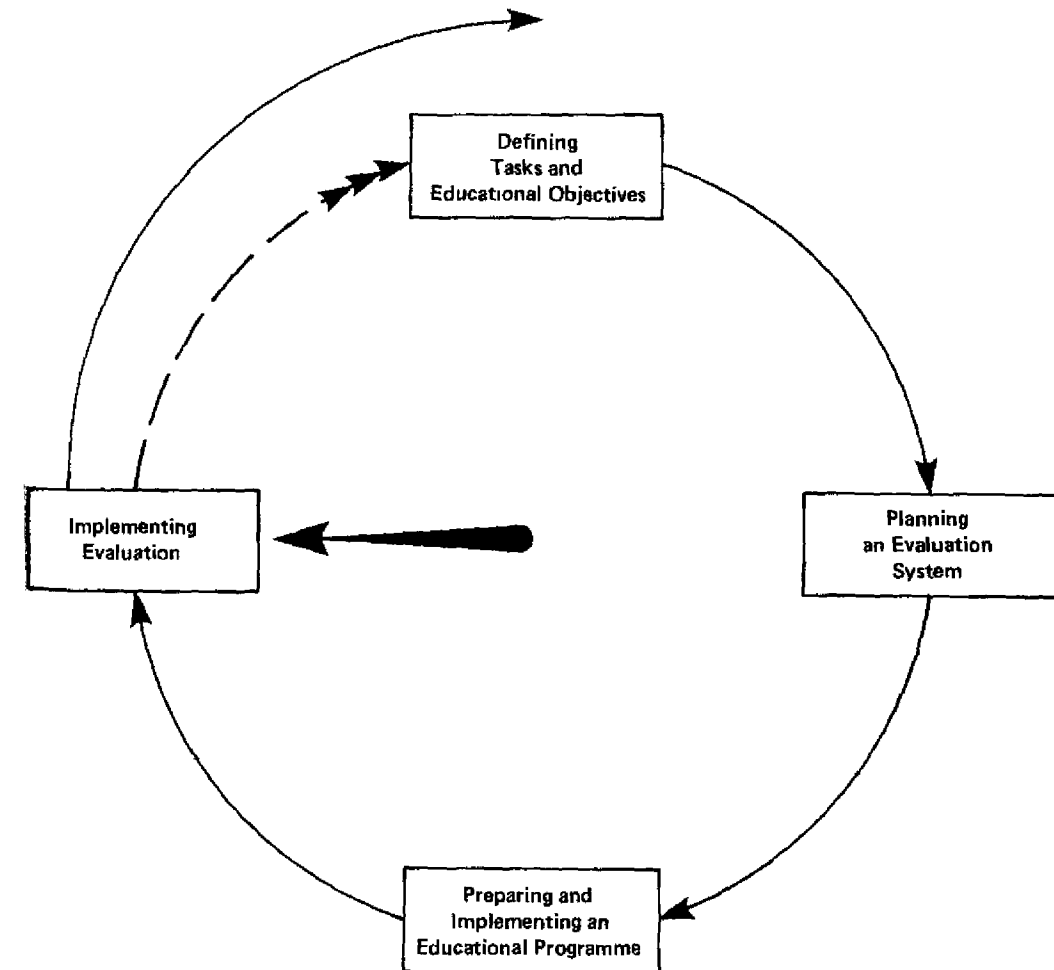
9. Explain the difference between a *relative* and an *absolute* criteria test.

10. Calculate the acceptable pass level for a MCQ examination and establish the scoring criteria and norms which permit determination of the passing grade of a mini-test (made up of the questions mentioned in objectives 5 and 6).

11. Do an item analysis of a question (calculate the difficulty index and the discrimination index) and draw the relevant conclusions.

12. Construct a specification table for an examination

the educational spiral



why evaluate?

Society, which finally pays the bill for health activities, relies on us to train health personnel to perform a social function corresponding to the health needs and demands of the community they are to serve.

Training centres for health personnel form an integral part of society, and must be prepared to operate *within* and *for* that society

Health activities are of a nature to stimulate social awareness and to provide leverage for social development. The following are some of the questions that must be asked in any attempt to discover whether training centres go to sufficient lengths in preparing the different kinds of health personnel and giving them a training that is "pitched" to meet the health needs of our societies.

- Do the graduates think and behave in terms of *health* rather than of *disease*? That is to say, do they apply techniques of prevention and health promotion and not only those of cure and rehabilitation?
- Do the graduates think and behave in terms of family and *community*, rather than in terms of the individual sick patient?
- Do the graduates think and behave in terms of membership of a *health team*

consisting of doctors, nurses and other health workers as well as social scientists and others?

- Do the graduates think and behave in terms of making the best and *most effective* use of the financial and material resources available?
- Do the graduates think and behave in terms of *their country's* patterns of health and disease, and the relevant *priorities*?

Consider the institution where you are working – and reflect on the above questions. If you can reply "yes" to all of them, then the essential has been achieved.

If, on the other hand, you cannot unreservedly say "yes" to certain of them, it is urgent for you to reconsider the orientation and the training programme of the school concerned.

To that end, several steps can be suggested

- 1 Evaluate the programme as a whole
2. Evaluate the general and intermediate educational objectives.
3. Have the students evaluate certain aspects of the curriculum.
- 4 Evaluate the students' level of performance.

guidelines for evaluating a health personnel training programme — summary description

The following guidelines are meant to be both general and comprehensive. The evaluator must adopt a realistic approach and decide in each case what is essential in any particular situation.

The guidelines take the form of a series of operations but, obviously, in practice these will often be carried out in a different order. What is essential, in any evaluation study, is for the evaluator to be constantly re-examining the information obtained, reformulating his questions and, even more important, re-vising his judgments. He is thus often obliged

to retrace his steps and, ideally, should review his whole orientation

The guidelines are presented under four main headings corresponding to the four phases of the evaluation:

- I Orientation
- II Design of the evaluation
- III Gathering information on the programme and its effects
- IV Analysis and reporting

Guidelines

Phase I — Orientation

1. Determine the general characteristics of the teaching programme.
- 2 Ascertain the general characteristics of the administrators, teachers and students.
3. Determine the financial resources available for the programme and its physical facilities.
4. Clarify the aims of the proposed evaluation.
- 5 Make an inventory of what information is readily available about the teaching programme
6. Determine the resources available for the evaluation and fix a time limit for its completion
- 7 Clarify the evaluator's role and that of the other persons taking part in the evaluation
- 8 Make a preliminary appraisal of the nature and the feasibility of the proposed evaluation.

Phase II — Design of the Evaluation

- 9 List the questions and issues to be considered,

10. Determine the appropriate sources of information and the procedures for its collection

- 11 Design the evaluation
12. Draw up an evaluation schedule.
13. Draw up an evaluation budget.

- 14 Obtain feedback check on the evaluation design, schedule and budget

Phase III — Gathering information on the programme and its effects

(a) Context and objectives

- 15 Describe briefly the context in which the programme operates.

- 16 Define the objectives by drawing up a list

- (i) of the future functions, taking account of the main health problems and the tasks and responsibilities deriving from each, *and*

- (ii) of the training objectives, taking account of the principal tasks the student should be able to perform satisfactorily at the end of his training

(b) *The characteristics of the students on enrolment*

17. Ascertain the students' characteristics that are likely to affect their performance as learners and subsequently as health workers

18. Gather information about the students at the beginning of their training.

19. Make a summary of the characteristics of the student group(s).

(c) *Resources and training processes*

20. Make an inventory of the resources available for the programme and describe how they are allocated

21. Examine and record the training processes used in the programme.

22. Make a summary of information relating to students' experience, using as themes the principal tasks and role models.

(d) *Effects and impact of the programme*

23. Ascertain the effects of the programme that are to be evaluated.

24. Describe and make a critical examination of the assessment procedures already in use in the programme.

25. If those methods are not adequate for the purposes of the present evaluation, develop and apply others.

26. Work out a set of procedures for observing and describing the long-term impact of the programme, if that is to be evaluated.

Phase IV — Analysis and reporting

27. Prepare a concise description of the programme using the information gathered in the course of Phase III.

28. Analyse the relationship between the various aspects of the programme.

29. Prepare an evaluation report recapitulating the qualities and achievements of the programme, the problems and difficulties encountered in carrying it out, and the available options or courses of action for its improvement.

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

For further details you are advised to read WHO Offset Publication No. 38 *Guidelines For Evaluating A Training Programme For Health Personnel* by F. M. Katz (WHO), Geneva, 1978)

or

Evaluation of Educational Programmes in Nursing by Moira Allen (WHO, Geneva, 1977)

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

points to consider in assessing the extent to which programme changes foster closer relationships between schools for health personnel and the wider society

1. Are changes in the government's priorities concerning health care more clearly understood by the programme planners inside the school?

2. Does the school now respond more swiftly and appropriately to any changes in national planning and priorities?

3. Does the school now work more harmoniously with the "consumers" of the trained personnel, i.e. with the relevant government agencies and the population?

4. Is direct feedback from (recent) graduates' "on-site" performance used as a criterion in assessing the strengths and weaknesses of the (new) curriculum? Is the school responsive to this type of feedback?

5. Are the various schools for health personnel within the whole national system learning from one another and assisting each other to replicate successful changes and progressively eliminate those which do not seem so successful?

6. Is the course content of the (new) programme *really relevant* to the common health

problems of the country? How could it be made more relevant?

7. What observable changes have taken place in

(i) patterns of mortality and morbidity
or

(ii) the responsiveness of people to the prevention of disease and disability?

Could such changes be the result of the programme changes?

8. Are consumers more satisfied with the quality of health care delivered by the health personnel trained according to the (new) programme?

9. What are the principal forces that have accelerated the processes of realistic, valuable programme change and development? What have been the major restraining forces on development? How might accelerating forces be strengthened and restraining forces altered?

Based on a paper by P. Blizard, WHO, New Delhi.

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

For each main question try to reply in one of the following ways

Yes — I have good reason to believe so, and I have even obtained some evidence to that effect.

No — I do not think that it is true

+/- — It is very possible but I have no facts to prove it

? — I am unable to reply to the question

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

points to consider in assessing the effectiveness of programme changes

— in terms of improvement in the administration and functioning of schools for health personnel

1. Has the (new) programme been based on a careful study of the skills needed (list of tasks) by health personnel in the community they are destined to serve?
 2. Have student failure rates and drop-out rates been reduced? To what extent?
 3. Has the time from initial enrolment to final graduation been reduced? To what extent?
 4. Has the volume of content in the (new) programme been reduced in comparison to what it was previously? To what extent?
 5. Have there been reductions in the financial and economic cost "per graduate"? To what extent?
 6. Are scarce teaching staff and associated facilities more effectively and efficiently used? In which particular respects?
 7. Is the content of the various subject areas vertically and horizontally integrated? What are the costs and benefits of that integration?
 8. Are collaboration and cooperation between the various departments seen as an easier, simpler process than before the programme changes were made?
 9. Have positive (or negative) effects been observed in the allocation of resources between teaching, research and administration?
 10. Have changes taken place in the teachers' conditions of work so that they can now spend more time and energy on their teaching and related responsibilities? Have such changes been an improvement? If so, why?
 11. Have locally organised teacher education programmes had some effect? If so, which?
 12. Has the local education bureau (if one exists) contributed to the changes that have taken place? In what ways could the functioning of the bureau be improved?
 13. Have administrative tasks become simpler and easier or have they become more difficult, time-consuming and complex?
 14. Are the roles of the chief administrator (Director, Dean, etc.) and other senior administrators now clearer, easier and more precise or have they become less precise and more difficult?
- Based on a paper by P. Blizard, WHO, New Delhi

Based on a paper by P. Blizard, WHO, New Delhi

For each main question try to reply in one of the following ways

Yes — I have good reason to believe so, and I have even obtained some evidence to that effect.

No — I do not think that it is true

+/- — It is very possible, but I have no facts to prove
it.

2 - I am unable to reply to the question.

A horizontal row of 20 small square icons. Each icon contains a unique black-and-white geometric design, such as triangles, squares, circles, and abstract patterns.

points to consider in assessing the effectiveness of programme change

- in terms of improvements in teachers' performance

1. Are the tasks involved in teaching the new programme seen by teachers as more interesting, relevant and rewarding?
2. Is the preparation of course content seen as a clearer and easier task, notwithstanding the possibility that it may be more time-consuming?
3. Have teachers improved their capacity to "facilitate" students' learning, i.e. have they developed the skills necessary to help students learn effectively?
4. Have there been demonstrable improvements in the programme for training new teachers to teach more effectively?
5. Do teachers now spend more time, effort and energy (in productive ways) on their tasks as teachers?
6. Do teachers actually *use* the skills they have acquired during training programmes when they are working with students?
7. Have there been any negative effects on teachers' other roles?

— and in terms of students' performance

1. Are students provided, during their studies, with a greater volume of information and variety of learning activities?
 2. Do students retain a greater proportion of the information and skills they have acquired?
 3. Do students show an improvement in their ability to use these skills in practical health care situations?
 4. Do students now show more initiative and independence in their learning habits, during their time in school, and in the later practice of their profession?
 5. Do students show an improvement in their willingness and ability to keep up to date as regards their professional competence (and under adverse conditions)?
 6. Are students now more actively involved in the tasks and processes of learning within the new programme?
 7. Do students now spend more time, energy and effort in their learning activities, and generally work harder?
 8. Does the working (and content) of the new programme provide a more interesting, exciting and rewarding experience for students?
 9. Do students play a fuller, more active, more rewarding and more useful role in programme planning, implementation and evaluation?
- Based on a paper by P. Blizard, WHO, New Delhi.

Based on a paper by P. Blizard, WHO, New Delhi.

guidelines for evaluating general and intermediate educational objectives

How can a training school for health personnel make sure that the teaching provided meets the needs and expectations of the population that the future graduates will have to serve? One of the best, but under-used, tools to ensure such relevance in teaching is the educational objective

A WHO Study Group met in 1976 to consider the place of educational objectives in the training of the different categories of health personnel, and in particular to devise guidelines for evaluating such objectives.

The guidelines given below are concerned not only with the way in which the objectives are stated but also with the process by which they were developed. Were the objectives defined so as to reflect, for instance, the needs of health workers and of the consumers of their services? Does the statement of each

objective contain an action verb clearly indicating what the student must do to show that he has attained the objective? Are the proposed objectives consistent with one another?

The guidelines are arranged under 20 headings. In most cases, the user can simply check one of the answers indicated. A space has been left at the end of each item for the user's remarks.

The report* of the Study Group, and especially the proposed guidelines, should be useful to administrators and teachers of schools for health personnel by helping them to define and evaluate educational objectives for their own use and for the benefit of their students.

*Criteria for the evaluation of learning objectives in the education of health personnel. Report of a WHO Study Group. WHO Technical Report Series No. 608, 1977, 47 pp. (The Report is also published in French and Spanish.)

A. Persons involved in drawing up the objectives

1. Which of the following persons were involved in either developing or reviewing the objectives?

- Teachers (if so please state what disciplines or specialties)
- Practitioners (if so, please state what disciplines or specialties)
- Education specialists
- Students
- Administrators (education, health, economy)
- Consumers of health care (i.e. "patients")
- Other (please specify)

2. How was the final decision made concerning the adoption of the objectives?

- By me alone (if so, please explain why)

- By my administrative supervisor (if so, please explain why)
- By the chief institutional administrator (if so, please explain why)
- By an interdepartmental committee (if so, please explain why)
- Jointly (if so, please indicate who was involved)
- Other (please specify)

(Circle the answer(s) you wish to give, and add any remarks).

3. Please describe any steps taken to avoid the objectives being distorted as a result of pressure exerted by certain interested groups (specialties, teaching disciplines, or others to be specified).

Remarks:

B. Basic data and other reference information

4. Given that the objectives should reflect the actual health needs of the community, which of the following needs are reflected?

- Present and future health problems
- Present and future functions of this category of health worker
- Functions of other health workers
- Expectations of health care consumers
- Expectations of health care workers
- Interrelationship between health needs and other needs of society
- Official government health policies
- Structure of the existing and future health care system
- Available data and research on health care and health manpower
- Current health manpower planning

5. When the objectives were developed, which of the following factors were considered? (In each case, please explain how.)

- Cost to the community (in relation to community resources)
- Characteristics of students
- Available teaching facilities

- Cultural setting of the educational institution
- Existing educational system(s) and philosophies
- Orientation of political and professional organisations
- Other (please specify)

6. Does each objective describe a specific competence that is significantly related to the performance of one or more of the tasks of that health worker?

Yes or No

If "yes" please give one or more examples.

Examples:

7. Do the objectives represent an adequate sample of the expected professional competences?

Yes or No

If "yes", please indicate how the sample was derived.

Remarks:

Circle the answer(s) you wish to give, and add any remarks.

C. Characteristics of the objectives

8. Given that the educational objectives are by definition student-oriented

■ Do they relate to the actual work the student is going to do after graduation?
Yes or No

■ Do they describe what the graduate will be able to do (*using an action verb*)?

Yes or No

9. Given that the objectives should be consistent with one another.

■ Do some objectives contradict others?
Yes or No

If "yes" please give an example.

■ Do some objectives support others?
Yes or No

If "yes" please give an example.

Remarks

10. Are the objectives realistic with regard to:

- Characteristics of students?

Yes or No

- Characteristics of teachers?

Yes or No

- Facilities available?

Yes or No

- Time available for learning?

Yes or No

11. Given that the objectives should be directed towards significant results of learning in all domains (i.e. knowledge, skills and attitudes)

- Are some of them specifically designed to facilitate personal affective development?

Yes or No

- Are some of them specifically designed to facilitate the development of intellectual enquiry?

Yes or No

- Are some of them specifically designed to facilitate development of the "total person"?

Yes or No

- Are some of them specifically designed to promote the development of competence (and, where appropriate, leadership competence) for team work?

Yes or No

D. Intended use of the objectives

12 Has consideration been given to how the objectives will be used in your institution?

Yes or No

If "yes", please indicate what plans were made for their use.

Remarks:

- Do some refer to development of research skills?

Yes or No

- Does the set of objectives refer to the need for life-long self-learning?

Yes or No

- Does the set of objectives refer to the need for life-long (continuous) self-assessment?

Yes or No

- Does the set of objectives reflect the scientific base for problem-solving needed by that category of health worker?

Yes or No

- Does the set of objectives reflect adequately the range of technical skills required from that category of health worker?

Yes or No

- Do some objectives reflect the importance of the health worker's responsibility to society?

Yes or No

- Do some objectives recognise the need to prepare the health worker to be responsive to new methods?

Yes or No.

If "yes", please give examples

13 If so, how was the final decision made concerning the use of the objectives?

- By me alone (if so, please explain why)

- By my administrative supervisor (if so, please explain why)

- By the chief institutional administrator (if so, please explain why)

- By an interdepartmental committee (if so, please explain why)

- Jointly (if so, please indicate who was involved)

- Other (please specify)

Circle the answer(s) you wish to give, and add any remarks.

14. Are students informed about the objectives?

- Students were given copies of the objectives.

Yes or No

- Students have discussed them

Yes or No

If so, under what circumstances?

- Students have not seen them

Yes or No

If not, why not?

E. Usefulness of the objectives

16 As you examine the set of objectives, which among the following results do you believe *likely to be produced*?

- They will be achieved.
- They will facilitate student learning.
- They will facilitate an integrated approach to learning.
- They will facilitate a problem-solving approach to learning.
- They will facilitate evaluation
- Other (please specify)
- They may not even be used.

Circle the answer(s) you wish to give, and, under "remarks", mention any facts in support of your opinion.

- Some of the students have probably seen them

Yes or No

If so, why only some of the students?

Remarks:

15. Are teachers informed about the objectives?

- All teachers were given copies of the objectives.

Yes or No

- Teachers have discussed them

Yes or No

If so, under what circumstances?

- Teachers have seen them.

Yes or No

If not, why not?

- Some of the teachers have probably seen them.

Yes or No

If so, why only some of the teachers?

17. Has the existence of the objectives had any impact on any of the following aspects of the teaching programme?

- Teaching/learning activities (if so, please give examples)
- Evaluation methods (if so, please give examples)
- Other (please specify)

Circle the answer(s) you wish to give, and add any remarks

F. Evaluation and revision of objectives

18. Are there provisions for a periodic review of the objectives?

Yes or No

If so, what is the review based on?

- Students' performance data
- Operational research data.
- Evidence of changing needs.
- Other (please specify)

Circle the answer(s) you wish to give, and add any remarks.

19. If so, who is involved in the review?

- Teachers
- Practitioners

- Education specialists
- Students
- Administrators
- Consumers (i.e. "patients")
- Other (please specify)

Circle the answer(s) you wish to give, and add any remarks

20. If the answer to question 18 is "yes",

- Does such a review facilitate programme changes?

Yes or No

(If so, please specify how.)

evaluation *by* students of programmes, teaching techniques and teachers¹

• Introduction

An evaluation made by students can provide the teacher with a most useful feedback of information on the quality of his teaching. Anyone who genuinely wishes to teach better should therefore seek *his students' opinions*. That may be done simply by a friendly talk with a few students, but it is preferable to prepare and distribute a questionnaire. Such an evaluation questionnaire may cover the whole or only part of the teaching, and it is for the teacher together with the students to decide just what should be evaluated by them. It would, for instance, be very valuable for the teacher to know how the students regard any changes in the teaching routine such as a new organisation or the introduction of audiovisual material.

It may be noted that anyone who asks students to evaluate his teaching need not doubt the validity of their judgement. For a decade or so many psychometric studies have revealed the validity and the accuracy of student opinions as well as their close correlation with "objective" measurements of the instructor's effectiveness. The many biases which were ascribed to the evaluators, for example, the influence of sex, academic efficiency, (poor students/good students), level of studies (beginners, finishers) the status of the course (compulsory/optional) have all proved to be of *negligible importance*.

• Preparation of "simple" evaluation questionnaires

Once the teaching aspects which are to be evaluated have been determined, the evaluation questionnaire is drawn up.

This questionnaire must comply with several essential requirements as concerns the wording of the statements, the scale of the answers and the method of "administration"

1 Wording of the statements

- The statements should be clear, simple and directly related to what it is desired to measure,
- statements which would be accepted straight away by everyone should not be included,
- statements containing double negatives should be rejected,
- each statement should contain a complete thought;
- the terms "uniquely", "solely", or "most of the time" should be avoided or used in moderation,
- a statement should preferably be in the form of an affirmation concerning which the student is asked to express a favourable, neutral or unfavourable opinion;
- the number of statements should be limited. Too long a questionnaire wearies the evaluators and this inevitably affects the validity of the answers. It is estimated that a questionnaire should not exceed 60 statements.

¹ Summary of a paper presented by J.-F. d'Ivernois, 21 May, 1975, Séminaire de Pédagogie universitaire, Laval University, Québec

1. Examples of statements

"Specific objectives make it possible to identify what is most important in the content to be learned".

"The time allocated for a learning activity is sufficient for you to achieve all the objectives aimed at by that activity"

2. Scale of answers

The student should not have to answer merely "yes" or "no" to a statement, for such an answer provides only scanty or dubious information.

That is why preference is given to the use of scales of answers comprising 5 degrees (and sometimes 7 or 9) (see below).

Students should enter against each statement (or on an answer-form or card) the numbers corresponding to their opinion. It is advisable to provide for an additional answer figure (0 to 9) corresponding to a *lack of opinion* concerning the statement.

The answers scale should be explained to the students at the beginning of the questionnaire, or mentioned against each statement. It is also as well to leave a space below each statement for comments by the student.

3. Method of administration

A person's reaction to a given stimulus is first immediate and then delayed. If the questionnaire is handed out immediately after the event the results observed will be different from those obtained if the questionnaire is administered later. These differences in distribution are attributable to the *effect of experience* (also called the "practice effect"). The teacher should therefore take this factor into account when he envisages the administration of an evaluation questionnaire.

4. Analysis and interpretation of the answers

Processing (machine or manual) of the answers is simple. The answers to a given statement are then analysed by calculating the number of students (frequency) who have replied by 1, 2, 3, 4, 5 or 9 to that statement. The *mean* of the answers to the statement can also be calculated. To do this one point is assigned to an answer "1", two points to an answer "2"; three points to an answer "3"; etc. The total of the number of points obtained for the statement is calculated and then divided by the number of students answering. To make interpretation easier, the trends (favourable, unfavourable) should be grouped together.

Example

Statement "The time allocated for a learning activity is sufficient for you to achieve all the objectives aimed at by that activity".

Numbers of answers: 100

Distribution: 1 : 3 3 : 5 5 : 69
2 : 2 4 : 21 9 : 0

Trends: 1 - 2 (disagreement) 5
4 - 5 (agreement) 90

Mean of answers : 4.51

subject matter. On the other hand, what proportion of the 50% of students who *agree* with the statement think that the subject matter covered is not only adequate but *even too much*?

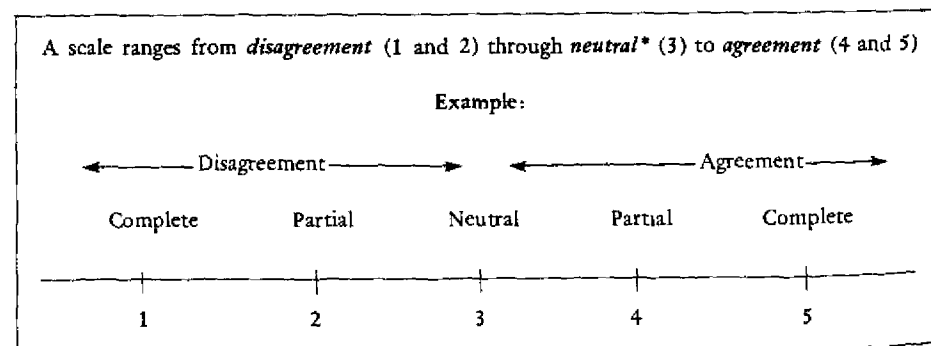
"Complex" evaluation questionnaires, such as the one devised by F. Gagné, covering the perception by the students of the teacher/students relationship have the aim of measuring as unambiguously as possible both the students' perception of *reality* and their level of expectations.

In a questionnaire of the Gagné type, each statement should include:

• Construction of complex questionnaires

1. "Simple" questionnaires enable the students' perception of *reality* to be measured. It is for the teacher who analyses the answers to deduce from them what the students desire or expect, but this interpretation may leave certain grey zones. If, for example, the statement submitted is: "the course provides an adequate coverage of the subject matter" and 40% of the students disagree with that statement then it is clear that these students feel that the course does not cover *sufficient*

- a *title*,
- a *detailed description* of the aspect to be measured;
- A 5-degree scale with *qualitative expression* of each degree;
- *Two questions*: the first (a) measuring the student's perception of reality and the second (b) the level of his expectations.



*For certain questions, it may be considered that a participant in an educational workshop, for instance, should be obliged to adopt a definite position and not take refuge in neutrality: the same applies to the position "no opinion".

Frequency of examples. What is required is to measure the frequency of examples putting in concrete form the more or less abstract concepts presented in the course.

- *Very rare* examples
- *Rare* examples
- *More or less frequent* examples
- *Frequent* examples
- *Very frequent* examples

Question A. Where would you place this course on the evaluation scale?

Question B. Where should it be in order to satisfy you?

2. Analysis and interpretation

The *means* of the answers to questions A (Reality) and B (Expectations) are calculated for each statement. Interpretation of these two means is easy since degree 1 on the scale shown usually corresponds to the minimum frequency or intensity, while degree 5 corresponds to maximum frequency or intensity (see example). The *deviation* between these two means is then calculated (S score). This deviation shows the dissatisfaction of the students, a dissatisfaction expressed in terms of a *lack* or an *excess*. The lower the value of the S score (the nearer it is to zero) the more

the particular aspect measured is deemed satisfactory. The S scores of several statements can be compared with one another for one and the same group of students. Furthermore, all the S scores in the questionnaire can be added together so as to give an overall and valid measurement of satisfaction or dissatisfaction. Finally, it should be noted that a questionnaire of the Gagné type can be adapted to different educational methods, e.g. lecture courses, small group activities, teaching by computer, by television, etc.

Example of Interpretation of a Questionnaire of the Gagné type.

Statement	Title of statement	Mean : Reality	Mean : Desires	Deviation (S score)	Interpretation
4	Variation in educational approach	2.39	3.74	1.52	Dissatisfaction
13	Active student participation (+ or -)	4.26	4.65	0.39	Satisfaction
27	Number of references to be consulted	4.51	3.27	1.24	Dissatisfaction (excess)

□
On this subject, see also *Evaluation of teachers and teaching effectiveness* by Christine H. McGuire in WHO Public Health Paper No. 61, Geneva, 1974.
□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □

EXERCISE

List the advantages and limitations of this type of evaluation of the educational process by students.

Advantages	Limitations

evaluation of students' level of performance

The following pages contain some examples of tests. In each case you are advised to practise making up a test on a subject with which you are familiar, after having selected a specific educational objective whose achievement can be effectively measured by the test in question (i.e., the test must be valid).

	pages
Real or simulated practical test	4.22
Execution of a project	4.22
Observational rating scale	4.23
Oral test	4.27
Essay	4.28
Short, open answer questions (modified essay question, and restricted response test)	4.28
Multiple choice questions (MCQ)	4.31
Programmed examination	4.41

Excellent examples may be found in
WHO Public Health Paper No. 72 "Assessing Health Workers' Performance: A Manual for
Training and Supervision", Geneva, 1980

and

Catalogue of Instruments for the performance assessment of Health Workers , HMD/81.6, WHO,
Geneva, 1981.

assessment of professional skills

Real or simulated practical tests

These tests are based on the *direct observation* of a professional task (cf. page 2 22).

A practical test is one that requires the student to perform a professional task in an environment and under conditions the same as or similar to those in which he will have to perform it in his future professional life.

For example:

The student is required to weigh a baby in an MCH centre and note the result on the appropriate record card

The student measures the blood pressure of another student.

The student is required to prepare a blood slide.

It is advisable to use this technique when the main component of an educational objective is a practical skill or an interpersonal relationship.

The disadvantages of this type of test are the relatively high cost in teaching personnel in view of the time required to observe each student's work, the risk of the patient being placed in a disagreeable situation, the difficulty in standardising the test conditions; and the heavy administrative task of coordinating the time schedules of teachers, students and services.

EXERCISE

Draw up a practical test, either real or simulated, or else a project test, designed to assess a professional task (pp. 1 41 or 1.52 or any task chosen by you), and keeping in view the criteria of validity, objectivity and practicability (re-read p 2.38).

Evaluate your results by constructing a specification table (see p 3.73).

Execution of a project

These tests are based on the *indirect observation* of a professional task.

A project execution test is one that requires the student to carry out an activity, in a variable period of time, that results in a *product* which is to be evaluated by the teacher

For example, the product might be a concrete piece of work such as a dissection, a dental impression or a histopathological slide, in addition to practical skills, certain intellectual processes can be evaluated by projects in a written form, such as a research report or a bibliography.

It is advisable to use this technique when the main component of an educational objective is a complex practical or intellectual skill, and when the product is more important than the student's manner of working.

The disadvantages of this type of test are the relatively high cost in teaching personnel responsible for evaluating the result of the project, and the need to establish a relationship of confidence with the student to avoid cheating.

assessing attitudes by observational rating scale

Everybody agrees that the attitudes of the physician and the other members of the health team to the patient and the patient's family are of the greatest importance.

Yet when one observes the way in which students' skills in this matter are assessed, it is seen to be so inadequate that one cannot help being struck by the paradox of the situation.

It must be recognised, in all humility, that this is the most difficult domain in which to make assessments, and the efforts of research workers in this field deserve every support. In a book such as the present one, only a superficial treatment of this subject can be given, and the reader should refer to more specialised publications (see bibliography).

1 Selection of students and assessment of attitudes

The term "attitude" denotes certain constant traits in an individual's ways of feeling and of thinking, and his predispositions towards action with regard to another person such as a patient or collaborator. An attitude is generally considered to be a *hypothetical construct* which is not directly observable but can be inferred from speech or outward behaviour. The inventory of attitudes constitutes the *operational definition* of the attitude.

Attitudes are probably not innate. The whole personality structure of an individual, and thus the whole of his behaviour, is constituted by a complex of interlinked attitudes. In its present state, research seems to indicate that it is illusory to expect to be able to change attitudes in the relatively short period of a programme of higher studies. Perhaps then it would be preferable to make sure, by a suitable selection process, that students *at the beginning of their studies already possessed* the appropriate attitudes for their future profession. It must however be remembered that the definition of those attitudes may easily be contestable, the preparation of such definitions must therefore be a *group activity*

free from arbitrary influences. It would be wise, moreover, to conduct carefully controlled experiments before taking any action based on the results of attitude assessment.

The situation being what it is, *the drawing of lots may not unreasonably be considered as a method of selection.**

2. Observational rating scales during studies

The method of observational rating scales requires the repeated and standardised *direct observation* of students' activity over a long period (several months) and in natural professional situations such as an outpatients clinic, a consulting room, a laboratory, or a hospital ward. The rating scales are used in evaluating the students' reactions, attitudes and activities. The fact that the scales are very easy to use often leads people to forget the many possible sources of error in such evaluations.

If a student is being observed when examining a patient, and if it is wished to make a separate evaluation of how he gains the patient's confidence, the following rating scale may be used.

- The student has taken all the necessary precautions, and the patient appears completely relaxed 5
- The student has taken the necessary precautions and has reassured the patient several times..... 4
- The student has made an effort, and has followed it up..... 3
- The student has made an effort, without following it up..... 2
- The student seems to be quite unaware of the problem..... 1

*The selection of students is made by drawing lots in the Netherlands

an example of an attitude rating scale

Task/Educational Objective: During a telephone conversation, to reassure the mother of a newly hospitalised child

Attitudes	- 2	- 1	0	+ 1	+ 2
1. When giving the mother information on her child's condition	refuses to reply to the mother's questions	gives no information	gives inaccurate information	gives accurate information but does not reply to mother's questions	gives accurate information and replies to mother's questions
2. When giving a clear explanation of what has been done for the child	often uses medical terms without ever explaining their meaning	often uses medical terms and seldom explains their meaning	seldom uses medical terms but does not always explain their meaning	seldom uses medical terms and always explains their meaning	uses only terms that the mother can understand
3. When suggesting that the mother should see her child	refuses the mother's request to see her child	does not suggest that the mother should see her child	agrees when the mother makes the request	spontaneously suggests that the mother should see her child	spontaneously makes the suggestion and explains any precautions to be taken
4. When telling the mother how she can obtain information about her child's condition	tells her nothing	when the mother enquires, refers her to the nurse	when the mother enquires, tells her the visiting hours and the persons she should ask	spontaneously informs the mother how to obtain information about the child's condition outside official hours	encourages the mother to come and obtain information about her child's condition outside official hours
5. When suggesting to the mother how a rupture of the mother/child relationship can be avoided	refuses to say how when the mother enquires	makes no suggestion	puts off the question until later	spontaneously suggests that the mother should stay at the hospital with her child	attempts to overcome the difficulties in the way of the mother staying with her child

Criterion. The student should score 7 out of 10 on the above rating scale

After Professor J -P. Grangaud (Test and measurement workshop, Timimoun, 1977)

The above example uses one of the most frequently employed rating scales. Others may be found in *The assessment of attitudes* by Agnes G. Rezler, WHO Public Health Paper No. 52, Geneva, 1973, and in J. P. Guilford (1954) *Psychometric methods*, McGraw Hill.

EXERCISE

Select a professional task (educational objective) whose main component is an attitude and construct a descriptive rating scale for its evaluation

Task/Objective: The student should be able to: _____

Attitudes	- 2	- 1	0	+ 1	+ 2

Criterion/acceptable level of performance: _____

EXERCISE

Try once more — it is not easy. Show your first attempt to several of your colleagues separately, discuss it in a small group including users of health services

Task/Objective: The student should be able to: _____

Attitudes	- 2	- 1	0	+ 1	+ 2

Criterion/acceptable level of performance: _____

a few words about the traditional oral examination

Definition. An examination consisting of a *dialogue* with the examiner who asks questions to which the candidate must reply.

In its standard form, the oral examination is a *closed-book test*. In *that form* it can evaluate only level 1 educational objectives (see page 1.19) in the domain of intellectual skills. Like traditional written examinations using short, open answer questions or MCQ, it provides a check on whether the student can express, more or less clearly, his knowledge of isolated facts or groups of facts that he ought to remember. Most often, it takes the form of a series of not necessarily interrelated questions.

There is a dialogue only if the examiner so wishes

It should be pointed out that, apart from its advantages and limitations which were described on page 2.30, this type of examination suffers from a scarcity of examiners who are really capable of making the best use of it in practice.

A better way of assessing a student's ability to communicate orally with another person is to use *simulation* methods, such as role playing or a telephone conversation, which are much nearer to actual professional tasks.

Personal Notes

long and short written questions

The use of the "essay question"

- Use "essay questions" only to evaluate a type of performance which cannot be measured just as efficiently by other methods (synthesis of a group of complex concepts, summary of a document, comparison of two phenomena, cause analysis, finding relationships, criticising the relevance of a concept, formulating a plan of action).
- Limit the problem posed so that it is clearly apparent to the candidate and define the structure of the answer.
- Employ terms that are as explicit as possible, such as "summarise", or "compare", "evaluate", "define", "arrange in order", etc. rather than "discuss" or "state everything you know", so that all the candidates immediately know what they have to do.
- Choose problems which call for careful consideration but whose solution can be briefly set out in the time allowed
- Allow no choice among the questions set.
- *For every question, set out yourself the elements which, according to you, should appear in the answer* (scoring procedure).
- Mark papers anonymously.
- When two or more teachers correct the same test, they should *agree on the scoring procedure before the test and correct the results separately*.
- Use a point system of scoring based upon those elements that are expected to appear in the answers.
- Try out the scoring procedure on a few papers. Preferably, have *all* the teachers then read *all* the answers to a given question; or, if need be, have one of them read all the answers to a particular question in all the papers, have another teacher do the same for another question, and so on
- Score the answers of all the students to one question before going on to the

scoring of another question.

- Do not form a judgment of a candidate on the basis of only one question but calculate for each candidate a summative score based on the reading of several different essays. Since such scores are more reliable than the score for any one essay taken separately.

Use of short, open answer questions

This involves series of questions drafted in such a way that the answer calls for a predetermined and precise concept. As their name indicates, the answer expected is short and can be expressed in different forms (open). Ideally, only one answer is acceptable no matter in what terms it is expressed.

The author of the question must define in advance (and in cooperation with colleagues) the answer called for by the wording of the question. If it appears that conceptually different answers will do for a given question then it should be reworded until that drawback disappears.

The pagination should allow the necessary space for the answer below each question. Marking is theoretically simple since the answer has been predetermined. In practice the frequency of difficulties arising from the conflicting views of two correctors is found to be of the order of 2%. There remain the problems of illegible writing,* absence of answers, copying mistakes, etc.

Moreover, all the rules concerning essays apply.

Short, open answer questions are also called "restricted response tests".**

A set of short, open answer questions preceded by a case history is sometimes called: Modified Essay Questions ***

*The same problem arises in the use of MCQ with automatic marking, but its frequency does not exceed 0.4%, and is very often less

**Rezler, A. G. The restricted response test: one alternative to multiple choice and essay tests. In NLN Publ. (15 - 1761), National League for Nursing, New York, 1979.

***Knox, J. D. E. The modified essay question, Dundee Association for the Study of Medical Education 1976 (Medical Education Booklet No. 5)

EXERCISE

Following some of the principles set out on the preceding page, prepare a written, "open-book" examination question of the essay type, and three short, open answer questions; indicate the *standard of performance* for each so that marking can be objective, i.e., a *rating scale* comprising *all the information that another person will need to be able to mark in the same way as you*.

Written question of the essay type, including the text of the expected answer.

Scoring instructions: Acceptable level of performance

*An "open-book" question means that the candidate may refer to any book, document, handout or personal notes. Such a question is not, therefore, intended to evaluate ability to memorise.

EXERCISE

Now go on to draft some short, open answer questions.

Text of the question	Acceptable answers
Scoring Instructions: Acceptable level of performance	

directions for writing multiple-choice questions

1. Make certain that the stem consists of a complete statement, not just a single word.

2. Place all common elements in the stem of the item. This adds simplicity and compactness to the item.

3. Make each item completely independent of answers to other items (for instance, the stem of one should not suggest the answer to another).

4. Eliminate all unrelated details from an item.

5. In general, avoid negative statements, but if a negative expression does appear in the stem of the question, underline it to draw the student's attention to it.

6. Use *plausible* or logical *distractors*. Each distractor should, by its content or nature, be such that it appears to have something to do with the question. Unrelated distractors appear silly to a thoughtful examinee. Since the
- number of possible answers is thereby cut down, the item loses some of its value.

7. Avoid the use of clues that may suggest the correct answer

8. Be sure that the distractors and the correct response possess homogeneity, that is, they should be fairly similar in content or in the total number of words.

9. Be cautious of the use of "none of the above" as a distractor or as a correct answer.

10. If it is impossible to obtain more than three plausible responses, do not waste time trying to invent some others.

11. When dealing with items that have numerical answers, arrange the answers in order from large to small or vice-versa.

12. Arrange the place for the correct answer in such a way that, for the test as a whole, no letter corresponding to a given answer appears more frequently than some other letter.

examples of multiple-choice questions

One "best response" type

Question 1

In differentiating cirrhosis of the liver from chronic constrictive pericarditis, a useful physical sign is.

- (a) hepatomegaly;
- (b) ascites;
- (c) distension of the neck veins,
- (d) pitting oedema of the ankles and legs,
- (e) splenomegaly

Question 2

Active immunisation is available against all of the following diseases *except*

- (a) tuberculosis,
- (b) smallpox,
- (c) poliomyelitis;
- (d) malaria
- (e) yellow fever

The multiple-true-false type (also called "multiple-response item").

This type consists of a stem followed by several true or false statements. The candidate is to determine whether or not *each of the four* statements which follows is true or false. He then responds according to a code which permits one out of five possible combinations or responses whereby one, two, three, or all four statements may be true

- when properly written, the multiple-true-false item type tests the student's knowledge or understanding of several related aspects of a substance, a disease, or a process;
- each of the statements or completions offered as possibilities must be clearly true or false. This is in contrast to the type "a" format in which alternatives which are "partially correct" may be used as distractors,
- this type of item should be written so that no two of the alternatives are mutually exclusively, i.e. the answer "all are correct" *must be a possible response*.

The directions for this item type are as follows:

For each of the incomplete statements below, *one or more* of the completions is correct. On the answer sheet blacken space under

- (a) if only 1, 2 and 3 are correct,
- (b) if only 1 and 3 are correct,
- (c) if only 2 and 4 are correct;
- (d) if only 4 is correct,
- (e) if all are correct

Question 3

A child suffering from an acute exacerbation of rheumatic fever usually has:

- 1. an elevated sedimentation rate;
- 2. a prolonged P - R interval;
- 3. an elevated antistreptolysin O titre;
- 4. subcutaneous nodules

The matching type

Directions for constructing matching items¹

- Limit the number of entries to about 10. If situations arise where 20 or 30 entries must be considered, construct two or three matching items. When long lists have to be matched, the student wastes too much time in trying to find the correct response
- Do not break items by the bottom of the page. The complete item should be on the same page.
- Have a longer list of questions than of possible answers and state in the directions that these may be used more than once. When there are an equal number of questions and answers, it is possible for the student, after responding to some of them, to complete his task by elimination and guessing
- Strive for homogeneity.

¹ Downie, N. M. Fundamentals of Measurement Techniques and Practices, New York, Oxford University Press, 1967.

The directions given to examinees for this type of item are as follows

"Each group of questions below consists of lettered headings followed by a list of numbered words or statements. For *each* numbered word or statement, select the *one* heading that is most closely associated with it and blacken the corresponding space on the answer sheet. Each lettered heading may be selected once, more than once, or not at all

Examples:

Questions 4 to 9

- (a) increased metabolic activity,
- (b) hyperinsulinism,
- (c) lack of storage of glycogen in the liver,
- (d) storage of an abnormal glycogen in the liver,
- (e) decreased secretion by pituitary or adrenal glands,
- 4. adenoma of islets of Langerhans,
- 5. violent exercise;
- 6. hyperthyroidism;
- 7. Simond's disease;
- 8. Von Gierke's disease,
- 9. epidemic hepatitis;

Questions 10 to 14

- (a) sodium bicarbonate;
- (b) sodium carboxymethylcellulose,
- (c) aluminium hydroxide gel;
- (d) none of the above,
- 10. a gastric antacid which is also used in the therapy of hypoparathyroidism because of its property of reducing the absorption of phosphorus;
- 11. because it is absorbed, it may cause alkalosis, particularly in infants and elderly patients;
- 12. a gastric antacid which has the disadvantage of causing "acid-rebound";
- 13. a gastric antacid which precipitates and inactivates gastric pepsin,
- 14. a gastric antacid and demulcent which can be converted to liver glycogen.

The comparison type

The "comparison" type permits one to compare and contrast two diseases, signs, symptoms, laboratory findings, etc

When using this type of item, one must be careful to:

- avoid the trivial,
- avoid selecting as one of the pair something that is rare or unusual. For example, if the item asks about the relation of a certain symptom to disease "x" or "y", and the frequency of the symptom in the two diseases is 90% and less than 1% respectively, then the examinee is in a dilemma. If he follows the principle of the "general rule", he may select answer A ("x" only); but if he is aware that the symptom does occur in the exceptional case of disease "y", then he may select answer C ("both"). Which response is correct?

The instructions for this type of item are as follows

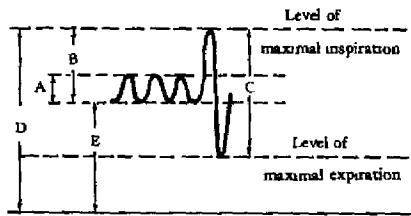
"Each set of lettered headings below is followed by a list of numbered words or phrases. For each numbered word or phrase, blacken the space on the answer sheet under:

- (a) if the item is associated with (a) *only*,
- (b) if the item is associated with (b) *only*;
- (c) if the item is associated with both (a) *and* (b);
- (d) if the item is associated with neither (a) *nor* (b)

Examples:

Questions 15 to 17

- (a) hookworm disease;
- (b) ascariasis;
- (c) both,
- (d) neither;
- 15. eosinophilia,
- 16. hypochromic anaemia,
- 17. infection through the skin



18. Total lung capacity
19. Inspiratory capacity
20. Vital capacity

[illegible]

Better information would help to disarm the defence mechanisms displayed by teachers when the problem is tackled scientifically, and would thus reduce heated reactions.

[illegible][illegible]

If no reference is made to questions of the true-false type it is not a chance omission! They are really very bad and should not be used

[illegible]

examples of items constructed in accordance with the level of intellectual processes*

Level 1 – Recall of facts

Item 1:

Which of the following corresponds to the result of your calculations?

- | | |
|---|-----|
| A | 140 |
| B | 178 |
| C | 200 |
| D | 225 |
| E | 250 |

Clinical case (for questions 3, 4 and 5)

A 25 year old male patient comes to your office complaining of cramps in the muscles of his arms or legs whenever he engages in strenuous muscular exercise. This complaint has become more prominent over the past three to four years. The general physical examination is normal. The patient has previously been seen by another doctor who performed a muscle biopsy. When you call the other doctor's office, you learn that the only abnormality noted on the biopsy was excess accumulation of glycogen in the muscle fibres. Electromyography was normal. The blood chemistry determinations related to carbohydrate metabolism which are performed in the hospital where you work are glucose and lactate.

You decide that you want to find out why the patient has muscle cramps and excess glycogen. To obtain further data, you apply the Fick principle to the right leg and insert small polyethylene catheters in the femoral artery and femoral vein at the level of the inguinal ligament. You make a blood flow estimation for the right leg and obtain a value of 200 ml/min at rest. You also take two resting blood samples from the arterial and venous catheters.

You then have your patient pump a bicycle ergometer for five minutes and at the end of exercise you take a series of five arterial and venous samples at two minute intervals. You have glucose and lactate determinations performed on each of the 14 blood samples.

$$A = \frac{A_{O_2} - V_{O_2}}{O_2 \text{ Consumption (ml/min)}}$$

$$B \quad \frac{\text{Cardiac Output (litres/min)}}{\text{Heart Rate (beats/min)}}$$

$$C = \frac{\text{Heart Rate (beats/min)}}{\text{Cardiac Output (litres/min)}}$$

$$D = \frac{O_2 \text{ Consumption (ml/min)}}{A_{O_2} - V_{O_2}}$$

$$E = \frac{\text{Cardiac Output (litres/min)}}{\text{Body Surface Area (sq. metres)}}$$

Level 2 – Interpretation of specific data

Item 2:

Apply the Fick principle to the following data and calculate the amount of glucose (in mg/min) extracted by the spleen of a dog

Cardiac output = 2 litres/min

Stroke volume ~ 200 ml/min

Cardiac index = 2.8 litres/sq. metre

Arterial glucose -- 100 mg/ml

Splenic vein glucose — 96 mg/ml

Splenic blood flow = 50 ml/min

* Re-read pages

EXERCISE

Compose one or two multiple choice questions of each of the types described above (one best response, multiple true-false, matching, and comparison types). At least half of your questions should measure an intellectual process above level 1, recall of facts (either level 2, interpretation of data, or level 3, problem-solving). Use the objectives that you yourself drew up on pages 1.54 and 1.55