1976

The Assessment of Teaching Practice : What criteria should we choose?

C. P. Hodgson
La Trobe University

The Assessment of Teaching Practice; What Criteria Should We Choose?

by C.P. Hodgson

Centre for the Study of Cirriculum and Teacher Education,
La Trobe University

For a number of years now, the practical element of pre-service teacher education has been taken in primary and secondary schools under the guidance of members of the school staff, and tutors from the college or university department of education.

The role of the college tutor in this situation has been admirably described by Catherine Fletcher (1958). It is essentially one of super-vising, helping and encouraging the work of the student teacher, but at some stage during the practice, the tutor may also be called upon to give an assessment of the student's ability to teach. The assessment required has varied from grades on a fifteen-point scale, to one of simply 'satisfactory' or 'unsatisfactory' sometimes accompanied by a written report (Anders-Richards, 1969).

Some of the difficulties inherent in assessment by grading have been outlined by Morris:

1.   The teaching mark lacks validity, i.e. it does not assess what it purports to assess . . . it reflects only a strictly limited number of teaching skills rather than the whole range of the student's teaching ability, and is inextricably linked with the discipline and organisation of the practice school.

2.   Assessment by grading is not reliable, i.e. not reproducible.

3.   Assessment by grading is inappropriate. A student undergoes complex and subtle changes during teaching practice.

4.   Assessment by grading has little practical value.

5.   Assessment by grading hinders the student's realisation of certain of the objectives of teaching practice and can impair the student-tutor relationship (Morris, 1970, p. 65).

In view of these and other criticisms, Morris suggests that assessment by grading should be rejected, even if accurate methods were to be found, and replaced by continuous evaluation without quantified assessment.

What is suggested is a searching and detailed evaluation of the progress of the student based not on generalised categories of teaching ability but on the student himself at a given moment in a given educational environment. (Morris, 1970, p. 70.)

One interesting departure from assessment by grading has been reported by Caspari and Eggleston (1965). In this experiment, the tutor did not enter the classroom in which the student was teaching. Instead, the tutor

supervised by consideration and discussion of "case histories" submitted by the student in the manner of social workers.

Despite the objections raised by Morris and various attempts to break away from assessments by grading, the fact remains that the majority of student-teachers practise and are assessed, in the school system, by tutor and supervising teachers.

The recurrent problem, then, is on what criteria should the assessment be based?

Teacher Effectiveness Criteria

Mitzel (1960) used Brownell's (1948) concepts when classifying teacher effectiveness criteria as product criteria, process criteria, and presage criteria.

Product criteria are described in terms of changes in student behaviour. Rabinowitz and Travers (1953), Ryans (1949, 1953) and Remmers (1952) have all argued in favour of assessing teacher competence by student gain. The use of such criteria presents many problems even if student-gain measurements are confined to the classroom. The problems would be even greater if such criteria were used in the student-teacher context, because of the very short contact time involved. It cannot be denied that effective teaching should produce some student gain — but gain in what? Even if some acceptable definition of "gain" could be found, it would be exceedingly difficult to construct the necessary measuring instruments. It is, perhaps, not surprising that, although theoretically important, product criteria have not received very much attention in studies on teacher effectiveness.

Of 138 studies, summarised by Barr (1948), only 19 used student gain as a criterion of teacher competence, and Mitzel and Gross (1956) reported only 20 studies which had a student growth criterion to measure teacher effectiveness.

Class Room Climate Criteria

Process criteria are defined in terms of those aspects of pupil and teacher behaviour which are deemed to be worthwhile in their own right. They are often described and measured in terms of climate or situations involving the social interaction of pupils and teacher. Two distinct categories of process criteria have been identified. One is obtained from observations of teacher behaviour, the other from observations made of pupil behaviour. Examples of this type of criteria would include the behaviour associated with the teacher's "presence" when taking a class, on the one hand, and "pupil co-operation and response" on the other.

The identification of process criteria through the measurement of classroom climate demands direct, sympathetic and detailed observation of the classroom situation (Medley & Mitzel, 1963). It is an area which has attracted a good deal of attention from workers dedicated to direct observation techniques and in which there has been some measure of success.

The foundations of this line of research seems to have been laid by Dorothy Thomas (1929) and her associates who, breaking with the rating methods then in use, conducted objective studies of nursery school behaviour by compiling descriptive accounts. She recognised that the problem was:

> . . . to find means of recording the particular stimuli in the uncontrolled environment to which a given individual, at a given moment, reacts overtly — what consistency is observable in his selective responses over a period of time and what variability is shown among different individuals. (Thomas, et al., 1929, p.5.)

As a result of her initial findings, Thomas decided to concentrate on interactions between individuals and thus opened a new dimension of research on teaching.

Anderson (1945, 1946a, 1946b), also working with nursery-school children, first identified and measured dominative and integrative behaviour and later extended the work by subdividing the two categories to embrace evidence of conflict and working together. Thus, from the original pair of behaviour criteria a list of sixty categories emerged by which teacher contact and pupil behaviour were recorded.

The most detailed and comprehensive study of this kind to date, however, was undertaken by Flanders (1965) who, in addition to identifying ten categories of "teacher and pupil talk" was also able to preserve the sequence of events taking place. The technique requires the observer to log the numbered categories of communication behaviour taking place at three-second intervals during a specified period of time, changes in activity being indicated by a double line to denote the end of an "episode". A tape recording of the proceedings enabled checking of any doubtful or missed categorisation, and the addition of other material, to take place more leisurely outside the classroom. The information collected in this way, can then be entered into a ten-by-ten matrix, thus giving a diagrammatic representation of the proportion of each category of behaviour used in a particular episode.

This sophisticated technique and its adaptations is limiting in its general application by its utter dependence on highly trained and experienced observers. Nevertheless, it represents a major step forward in the observation and evaluation of classroom behaviour.

In spite of considerable research activity into classroom climate, no universally acceptable list of process criteria has so far emerged, and Smith (1970) goes so far as to suggest:

> that more direct and primitive analyses of teaching behaviour are needed as a preface to experimental and correlational studies. (Smith, 1970, p.3.)

His study is based on transcribed recordings of high school discourses. These larger units of verbal behaviour are termed episodes and are analyzed "logically" rather than "psychologically" or "linguistically". By using much larger units of verbal discourse, it is hoped to isolate logical patterns of behaviour but no measurement is attempted.

One of the most comprehensive studies of teacher and pupil behaviour is that of Ryans (1960). After experimenting with checklists, frequency factors, estimate procedures and intuitive procedures, his final instrument contained twenty-two teacher and pupil behaviours, which are rated on a bipolar seven-point scale, together with a glossary in which the behaviours are define operationally.

Ryans emphasises the fact that no matter how good a rating schedule may be, the success of the measuring instrument depends, ultimately, on the skill of the observer. The dimensions or criteria, used in the rating instrument, can be classified under the heading of classroom climate and so come into the process criteria category.

Other parts of the work deal with personal characteristics and attitudes to the community which can be described as presage criteria.

Presage criteria are so called because of their origin in guessed predictions. Their inclusion is a matter of precedent since a large proportion of the research on teacher competence has employed dependent variables which can be included in this category. Examples of these variables include intelligence, character and personal characteristics which have been accepted on the basis of "common sense" appeal but which seem to have little to offer in the context of operationally definable criteria. Product criteria, however, are given little prominence in the study which is perhaps surprising in view of the plea, previously made by Ryans, for the use of pupil gain in assessment of teaching, but at the same time their absence supports the thesis that the construction of instruments to measure product criteria is a very difficult task.

Pupils as Observers

It has been argued that the subjects who observe most of a teacher's behaviours are the pupils themselves, and a number of rating scales, employed to assess teacher behaviour and pupil attitudes, use the pupils as observers. Three such scales include:

1. The Purdue rating scale for instruction, which has been constructed so that it can be scored on an IBM graphic item counter; (Remmers, 1960);

2. A diagnostic rating of teacher performance developed by Cosgrove (1959), who used an ingenious forced choice method to obtain diagnostic measures;

3. The Purdue instruction performance indicator, which also uses a forced choice technique. (Snedeker and Remmers, 1960.)

While such methods may be of some limited help to individual teachers,

they have been severely criticised by Brogan (1968), who sees little value in ". . . the opinion of immature and relatively uninformed adolescents whose experience does not extend beyond the local campus. (Brogan, 1968, p. 191.)" A less emotive but perhaps more pertinent comment is made by Margaret Mead (1958) when she observes that:

> . . . the role of the teacher — as reflected in the comments of a whole class — is an exceedingly interesting one. The disliked teacher is personalised and vivid, the teacher, who has obviously been very successful and has caught the imagination and enthusiasm of the whole class does not emerge as a person at all, but, instead, sinks into the background of good classroom conditions, together with good laboratory equipment. (Mead & Metraux, 1958, p. 461).

Certainly, even though they do observe a substantial portion of the teacher's behaviour, pupils cannot be regarded as skilled enough observers to satisfy the exacting standards of Flanders (1965) or Ryans (1960).

Leeds and Cook (1947), however, in devising a scale for teacher-pupil attitudes known as the Minnesota Teacher Attitude Inventory (MTAI), used pupil reaction as basic validating criteria. Later, they correlated inventory scores with ratings obtained by their own observations, from principals and from pupils. The final form of the MTAI has been used as the basis of a large number of studies concerning teacher personality and characteristics and attempts have been made to relate attitudes measured by the MTAI to different personality factors. (Getzels & Jackson, 1963).

It would seem from this work that pupil ratings of teachers can be of value if considered in conjunction with assessments made by professional observers.

The Stanford Appraisal Guide

One method of teacher assessment which would seem most closely to fit the needs of the student-teacher situation is the Stanford Appraisal Guide of Teacher Competence, developed over ten years by the Stanford University School of Education. This Appraisal Guide divides teaching competence into five major categories namely:

1. Choice and formulation of aims;

2. Planning how to fulfil the aims;

3. Fulfilling the aims;

4. Assessing how successfully the aims have been fulfilled; and

5. Maintaining professional standards in school, parent-teacher relationships and in the community as a whole.

Each category is divided into sub-divisions and each sub-division is defined by a statement. The teacher's performance is appraised in detail against each statement on a seven-point scale by the teacher himself, the university supervisor, the school supervisor and by the pupils. The results

are used in discussion with the students, who thus receive constant feedback. The amount of computation, however, needed to produce the appraisal would make it impractical to use in its present form for routine school practice evaluation. The evidence available from studies of teacher effectiveness supports the conclusion, reached by Evans (1951) that:

> probably, the best criterion of any teachers' work would be a composite measure based on pupil gains in information, ratings by competent observers, and a rating based upon the opinions of pupils. (Evans, 1951, p. 94.)

It is doubtful, however, that such a composite measure of a teacher's work could be satisfactorily obtained, even if the necessary measuring instruments were available, because of the limitations of the practice situation. The length of time which a student spends in any given teaching situation would make measures of student gain and student opinion dubious to say the least. It is, therefore, not surprising that studies, specifically concerned with the assessment of practice teaching, have mainly concentrated on attributes, thought to be associated with good teaching, which can be demonstrated by student teachers and observed by their supervisors.

U.K. Studies

Catell (1931) produced a rating scale of twenty-two qualities which "good" teachers should possess, based on an opinion poll of professional groups. Panton (1934) found that standards of assessment of practical teaching varied between colleges and also devised a rating scale.

Robertson (1957) asked eighteen supervisors of post-graduate students to rank fifty student attributes associated with successful teaching and found that ". . . the degree of general agreement about the attributes which contribute to success in practical teaching was not high." (Robertson, 1957, p. 122.)

Poppleton (1968) developed an assessment form for use in the University of Sheffield, Department of Education. Ratings of 249 students by schools and department supervisors employing this form yielded product moment correlations of 0.60 with the supervisors emphasising academic qualities, whereas the schools were equally concerned with the affective aspects of teaching.

An evaluative study by Wiseman and Start (1965) followed up 248 teachers who qualified at seven colleges and one university department, within one Area Training Organisation in 1955. Although handicapped by a high non-response rate (64%), the findings included the fact that ". . . little correlation was found between college assessment and the various criteria of success in the profession." (Wiseman & Start, 1965, p. 358.) More recently, Stones and Morris (1970) collected information on six areas connected with the assessment of practical teaching from 122 (representing a response of 65%) colleges and departments of Education in the United Kingdom and concluded that:

> individual institutions and Area Training Organisations are looking for, and assessing different behaviours and qualities in their students. (Stones & Morris, 1970, p. 19.)

Although mainly concerned with modes of assessment in this study, rather than specifically seeking information on criteria used, Stones and Morris (1970) list 123 different criteria which were mentioned in connection with assessment procedures.

Other Factors

It seems clear that one of the major reasons for the discrepancies noted in practice-teaching assessments is the lack of agreed criteria. An additional factor, however, could be the different weighting given to the same criterion by different assessors.

Hodgson (1972) conducted a questionnaire enquiry to examine what qualities (criteria) are looked for in students by their tutors. In this study, fifty different criteria were listed and the opinion of tutors sought about (a) the importance, or otherwise, of these criteria in describing students' teaching performance, and (b) the relative weighting given to the criteria, agreed as important, in arriving at an overall student assessment.

At the same time, a similar enquiry among students sought their opinions on these and other questions concerning the discussion of criteria and the influence of a particular school environment on a teaching performance. From the results of the enquiry, it was possible to relate the agreed criteria to five main categories which were:

1. Pre-teaching activities — i.e. those performed before the practice properly begins

2. Pre-lesson activities — i.e. those performed before a particular lesson is taken

3. Lesson activities — i.e. those involving pupil contact in the classroom

4. Post-lesson activities — i.e. those performed after pupil contact in the classroom

5. Activities in the school environment — i.e. those involving personal contact or behaviour in the school, but outside the classroom

All these categories were deemed "important" although to different degrees, as indicated by the suggested weighting figures given below.

Category		1	2	3	4	5
Average percentage weightings suggested	by Tutors —	16	18	46	9	11
	by Students —	11	18	46	13	12

There was a similar agreement between tutors and students as to the specific activities which they thought should be used in making an assessment of practical teaching. The tutors tended to favour those activities most readily discernible/observable, i.e. planning and classroom, while the students tended to favour those which involve direct contact with pupils both in and outside the classroom.

In spite of the fact that the students claimed not to have been told what criteria were taken into account in the assessment or had limited feedback from previous practices, the measure of agreement suggests that such information is acquired indirectly if not specifically. 80% of the students thought that their teaching had been affected by the type and situation of the school and the whole sample felt that some account should be taken of this when the assessment is made.

Evaluation Relevant to the Specific Situation

All the available evidence seems to suggest that no concensus is likely to be reached on valid criteria to be used in the assessment of teaching practice.

It may be that a re-examination of the broad aims and more specific objectives of teaching practice is required. In such a re-examination the issue of assessment needs careful consideration for, although teaching practice occupies a central place in the education of teachers, and as such is assessed as part of the course, the major concern need only be to detect those students who are unsuited to teaching.

It would also be appropriate, in any re-examination of broad aims, to acknowledge the artificiality and limitations of the teaching practice situation and to recognise the vital role of the supervising teacher in particular and the school in general, when deciding what the student teacher can realistically be expected to achieve. In this way, it would be possible to establish suitable criteria by discussion between the tutor, supervising teacher and student, to cater for individual situations. Such criteria should include activities which are a necessary pre-requisite for good classroom practice, taking into account the location and climate of the school, the background of the children, the teaching-learning methods employed and the theoretical perspective of the total teacher education programme. They should also include the activities which a teacher is expected to perform, outside the classroom, in the school environment. In this way, the teaching practice could be viewed as a co-operative, on-going process of personal development and professional growth of the student-teacher and one for which the supervising teacher/school, tutor/education department and student are mutually accountable.

References

ANDERS-RICHARDS, D., Teaching Practice: Assessment Procedures. *Education for Teaching,* 1969, 80, 71-72.

ANDERSON, H.H. and BREWER, Helen M., Studies of Teachers' Classroom Personalities. I. Dominative and Socially Integrative Behaviour Of Kindergarten Teachers, *Applied Psychology Monograph,* 1945, No. 6.

ANDERSON, H.H. and BREWER, J.E., Studies of Teachers' Classroom Personalities. II. Effects of Teachers' Dominative and Integrative Contacts on Childrens' Classroom Behaviour. *Applied Psychology Monograph,* 1946(a), No. 8.

ANDERSON, H.H. BREWER, J.E. and REED, Mary F., Studies of Teachers' Classroom Personalities. III. Follow-up Studies of the Effects of Dominative and Integrative Contacts on Childrens' Behaviour. *Applied Psychology Monograph,* 1946(b), No. 11.

BARR, A.S., The Measurement and Prediction of Teacher Efficiency: A Summary of Investigations. *J. Exp. Ed.* 1948, 16, 203-283.

BROGAN, H.O., Evaluation of Teaching. In *Improving College and University Teaching,* 1968, 16, 3, 191-192.

BROWNELL, W.A., Criteria of Learning in Educational Research. In *Improving Educational Research,* A.E.R.A., 1948, 106.

CASPARI, I.E. and EAGLESTON, J., A New Approach to the Supervision of Teaching Practice. *Education for Teaching,* 1965, 68, 42-52.

CATTELL, R.B., The Assessment of Teaching Ability. *Brit. J. Ed. Psychology,* 1931, 1, 1, 48-71.

COSGROVE, D.J., Diagnostic Rating of Teacher Performance. *J. Ed. Psychology,* 1959, 50, 200-204.

EVANS, Kathleen M., An Annotated Bibliography of British Research on Teaching and Teaching Ability. *Ed Res.,* 1961, IV, 1, 69-80.

FLANDERS, N.A., Teacher Influences, Pupil Attitudes and Achievements. *COOP Res., Monograph,* 1965, 12, USOE, 371, 102, FS.

FLETCHER, Catherine, Supervision and Assessment of Practical Teaching. *Education for Teaching,* 1958, 47, 17-23.

GETZELS, J.W. and JACKSON, P.W., The Teacher's Personality and Characteristics. In *GAGE, Handbook of Research in Teaching,* 1963, Ch. 11, 508-522, Rand, McNally and Co., Chicago.

HODGSON, C.P., An Investigation into the Assessment of Practical Teaching. Unpublished M.Sc. thesis, University of East Anglia, 1972.

LEEDS, C.H. and COOK, W.W., The Construction and Differential Value of a Scale for Determining Teacher-Pupil Attitudes. *J. of Exp. Ed.,* 1947, 16, 149,159.

MEAD, Margaret and METRAUX, Rhoda, Image of the Scientist Among High School Students. In *BRANDWEIN, et al, A Book of Methods,* New York. Harcourt, Brace and Co., 1958.

MEDLEY, D.M. and MITZEL, H.E., Measuring Classroom Behaviour by Systematic Observation. In GAGE, *Handbook of Research on Teaching,* Ch. 6, Rand, McNally and Co., Chicago, 1963.

MITZEL, H.E., Teacher Effectiveness. *In Encyclopedia of Educational Research,* (3rd edition), 1960, 1482.

MITZEL, H.E. and GROSS, Cecily F., A Critical Review of the Development of Pupil-Growth Criteria. *In Studies of Teacher Effectiveness,* C.C.N.Y., 1956, 28.

MORRIS, S., The Assessment and Evaluation of Teaching Practice. In

Towards Evaluation: Some thoughts on Tests and Teacher Education. *Educational Review*, occasional publication, University of Birmingham, School of Education, 1970, 4, 64-70.

PANTON, J.H., The Assessment of Teaching Ability with Special Reference to Men Students in Training, Unpublished M.A. Thesis, University of London, 1934.

POPPLETON, Pamela K., The Assessment of Teaching Practice: What Criteria Do We Use? *Education for Teaching*, 1968, 75, 59-64.

RABINOWITZ, W. and TRAVERS, R.M.W., Problems of Defining and Assessing Teacher Effectiveness. *Educational Theory*, 1953, 3, 212-219.

REMMERS, H.H., Report of the Committee on the Criteria of Teacher Effectiveness. *Ed. Res.*, 1952, 22, 238-263.

REMMERS, H.H., Second Report of the Committee on Criteria of Teacher Effectiveness. *J. Ed. Res.*, 1953, 46, 641-648.

ROBERTSON, J.D.C., An Analysis of the Views of Supervisors on the Attributes of Successful Graduate Student Teachers. *Brit J. Ed. Psychology*, 1957, xxvii, 2, 115-126.

RYANS, D.G., The Criteria of Teaching Effectiveness. *J. Ed. Res.*, 1949, 42, 690-699.

RYANS, D.G., Teacher Personnel Research, *J. Ed. Res.*, 1953(a), 4, 19-27, 73,83.

RYANS, D.G. Investigation of Teacher Characteristics *J. Ed. Res.*, 1953(b), 34, 370-396.

RYANS, D.G., *Characteristics of Teachers.* American Council on Education, Washington, D.C., 1960.

SMITH, B.O. (et al), *A Study of the Logic of Teaching.* University of Illinois Press, Urbana, Chicago, 1970.

SNEDEKER, J.M. and REMMERS, H.H., *The Purdue Instruction Performance Indicator.* West Lafayette, Ind. University, Book Store, 1960.

STONES, E. and MORRIS, R., *The Assessment of Practical Teaching.* mimeo., 1970, School of Education, University of Birmingham.

THOMAS, Dorothy, S. et al, Some New Techniques for Studying Social Behaviour. *Child Development.* Monograph, 1929, 1.

WISEMAN, S. and START, K.B., A Follow-Up of Teachers Five Years After Completing their Training. *Brit J. Ed., Psychology*, 1965, xxxv, 3, 342-361.