DOCUMENT RESUME

ED 110 502 TM 004 775

AUTHOR White, D. R.
TITLE The Stirling Lesson-Sampling Instruments.
NOTE 13p.

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Classroom Observation Techniques; *Evaluation Methods; Lesson Observation Criteria; *Measurement Techniques; Microteaching; *Student Teachers; Teacher Behavior; Teacher Education; *Teacher Interns; Teaching Skills; Test Reliability; Tests
IDENTIFIERS Scotland; Stirling University

ABSTRACT

A long-term Leverhulme Research Project was established at Stirling University in 1970 to investigate the potential of microteaching as a major ingredient in the preparation of graduate high school teachers in Scotland. Members of the research team developed systematic observation schedules for each of the skills, in order to sharpen the focus of interns and their supervisors on the behaviors involved. This paper reports an attempt to extend the use of systematic observation schedules to the supervisors' classroom visits during conventional practice teaching. A conceptual framework of the professional skills was first devised, under the headings of questioning, structuring the discourse, varying the stimulus and reinforcing. A battery of lesson-sampling instruments was then designed, using a mixture of rating scales, categories and sign systems. Studio trials with teams of three trained observers recording simultaneously from video-taped episodes of lessons taught by experienced Scottish high school teachers gave encouraging results. After revision, the instruments were submitted to a field test. Seven pairs of supervisors from a college of education observed three lessons given by each of the interns allocated to them. Instruments were used by mutual agreement within the pairs, allowing roughly half the available time. Acceptable levels of agreement were obtained, suggesting the feasibility of the proposed approach. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED110502

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM 004 275

THE STIRLING LESSON-SAMPLING INSTRUMENTS

Microteaching courses play a part in an increasing number of teacher education programs throughout the Western world. During such courses, tutors and intern teachers build up a shared conceptual understanding of the professional skills of teaching and develop a common language for talking about them. Yet at Stirling, where microteaching has formed part of the course in education since the University opened in 1967, tutors did not take particular note of these professional skills when observing conventional classroom practice. Some used an entirely unstructured approach, others used the Stanford Teacher Comparison & Appraisal Guide. This calls for ratings on seventeen dimensions of teacher behaviour. It is a high inference system, leaving the individual tutor a considerable measure of responsibility for determining the interpretation of each heading and of levels of performance meriting the different ratings. The technique is heavily dependent upon subjective impressions drawn from a mass of feedback to which the tutor is subjected during his visit. No record of what actually happened is available for post-lesson discussion with the intern.

The Stirling Lesson-Sampling Instruments were designed in an attempt to make use of the shared understanding developed during the microteaching skills program. The writer designed a battery of instruments to be used selectively by tutors when observing interns on conventional teaching practice. Using low inference approaches whenever possible, recording systems were planned to show sequential patterns of behaviour, providing

---

Footnote The writer spent a sabbatical year with the Leverhulme Research Project Team under the general direction of Prof. Elizabeth Perrott, and the more specific guidance of Donald McIntyre, Senior Research Fellow. He wishes to acknowledge his indebtedness to all members of the team for assistance with the work reported here.

some common ground for post-lesson discussion.

RATIONALE

An observer cannot 'see' everything that happens in the classroom. What he attends to and his evaluation of what he attends to will both be affected by his own frames of reference. Interns are aware of this, and tend to view the criteria for good grades as being more directly related to a correct 'reading' of their tutor's foibles than to successful exercise of the profession skills covered in the course! The procedures specified for the use of the Stirling Lesson-Sampling Instruments leave the tutor free to follow his own intuitive directions for approximately half the time, but prescribe patterns of observation for the remainder.

The instruments provide for a sampling of the teacher's behaviors in four broad groups selected to cover the skills being taught at the time in Stirling University's microteaching program. The records obtained merely indicate that certain behaviors were observed (or not observed), and retain some aspects of the sequence in which they occurred. Judgement as to the appropriateness of the behaviors can only be made in terms of the intern's stated objectives and a number of other factors. This calls for sensitivity, understanding, skill and experience on the part of the tutor. The instruments provide the date upon which the tutor will base his judgements and guidance, and help the intern to recall the specific incidents referred to.

The first requirement for a lesson-sampling approach is a clear plan, produced by the intern, indicating the objectives of the lesson and the sequence of activities. On arrival, the tutor scans the lesson plan, and decides which instruments he will use and at what stage. The instruments covering motivating and reinforcing can be used for a very wide range of teaching interactions but the other two can, of course, only be used when the teacher is asking questions or explaining something. Each instrument

requires a 5 to 10 minute period of recording. Assuming that a tutor will visit each intern at least five or six times, it should be possible to sample performances several times with each instrument, still leaving ample time for attending to aspects of teaching not covered by the instruments. Two sets of rating scales were also produced, one for use immediately after a single lesson, the other for completion after observing not less than three lessons.

THE INSTRUMENTS

INSTRUMENT 1: QUESTIONING A fairly detailed description of this instrument will be given here, as illustrative of the general approach. The other instruments will be described more briefly.

Fig. 1 provides an example of a record of a five-minute episode, taken from a mathematics lesson.

Fig. 1. Instrument 1 : Questioning

Five-minute episode from math lesson

Higher order		q x q	q
Application	q r r r q r	qq	qq
Lower order	q		
probability one die		Why 1/36 ?	Use?

The three categories are defined as questions calling for:

Lower order	simple recall, recognition, definition, sensory description, unqualified opinion
Application	direct application or exemplification of some previously stated or implied rule or relationship, measuring in a given context (particulary in language lessons)
Higher order	explanation showing functional, causative, purposive or similar relationship, or justification

intention, using all the available clues. A problem experienced on several occasions arose from pseudo-open questions. The wording of the teacher's question led the observers and the children to infer that he intended to ask an open question, but his reaction to the children's responses showed quite clearly that this was not so. Under such circumstances, the observer simply cancels the ringing and makes a mental note to comment on this during the post-lesson discussion.

There is a large number of decisions to be made, sometimes at fairly high speed. The following order of decision making is recommended:

probe	higher order			
		open	affective	answer accepted
redirect	application			
		closed	non-affective	no acceptable answer
other	lower order			

It is often helpful to write one or two key words below the recording to assist the observer to recall specific lines of questioning.

In the episode recorded in Fig. 1, the teacher opened by asking for a statement of the rule for calculating the probability of a given number being obtained with one throw of a die. He followed this with a series of questions, some of them redirected, calling for applications of this rule. He then stated how the rule could be extended to calculate the probability of obtaining a given score from one throw of two dice, and asked why the probability of throwing 2 was $\frac{1}{36}$. Failing to obtain an answer, he gave further information, asked two application questions, then repeated his demand for an explanation. The answer was partially acceptable, so he probed for elaboration. Two further application questions tested the students' ability to use the new rule, and he ended with an open question: "In what way might you make use of your ability to calculate probabilities?"

These three categories are intended to indicate the logical structure underlying the pattern of questioning.

A further classification is made by the use of different symbols which can be placed in any of the three categories.

r — redirected, a question which is essentially the same as the previous one, put to a different pupil

probe, a question asked of the pupil who answered the previous question, asking him to carry his first answer a little further, including prompts and requests for clarification or justification

q — all other questions

Two further classifications are included - open/closed and affective/non-affective. Open questions are defined as those leading to a wide range of acceptable answers, not likely to have been preconceived by the teacher. Affective questions are those likely to call into play the student's own emotions, value-systems or preferences, or require the student to interpret those of another person. Open questions are indicated by ringing the appropriate symbol, as ' q ' at the end of the top line in Fig. 1. Affective questions are shown by underlining the symbol (there were no affective questions in the episode recorded in Fig. 1).

The unit of a question was defined as follows: a question is marked off as a unit by: a reply; teacher calling a student's name; teacher pointing to a student; teacher looking expectantly at a student or pausing without saying anything for at least two seconds. Procedural and rhetorical questions are ignored. The symbol for a question to which no answer is received, or the answer to which is rejected, is followed by an 'x'.

In using the instrument, the observer often has to infer the teacher's

INSTRUMENT 2 : STRUCTURE OF THE DISCOURSE This instrument was designed to provide an analysis of the logical structure of the verbal interchange, with an indication as to whether a given contribution was made by the teacher or by a student. It incorporates a record of interruptions and of statements lacking clarity. Instrument 2 may be used whenever the teacher's main objective is to tell or explain something to the class, or to an individual. There is some overlap with the questioning instrument.

Only two basic symbols are used: a 't' representing a contribution by the teacher alone; and an 's' representing a student contribution. Where a teacher's question leads to a contribution from a student, the complete unit is recorded 's'.

The three categories - stating a rule, applying or exemplifying a rule, and explaining a rule - correspond with the lower order, application and higher order categories respectively on Instrument 1. A rule is defined as 'a statement connecting one object, set of objects, event, action, state of affairs, principle or pattern with another or others (e.g. 'women tend to be shorter than men', 'an acid reacts with a base to give a salt and water'). The concept of 'rules' required here is that they form the building bricks from which the discourse is constructed. 'Prunes are made from plums' satisfies the requirements for a rule but in a given lesson it may or may not be of importance in the structure of the discourse. This problem may be minimized by requiring the interns to indicate clearly in their lesson notes the rules (as defined above) which they consider to be the key building bricks for any part of the lesson in which their main objectives are to convey information or develop understanding. The observer's task is then limited to recording the sequences in which these rules are stated, applied or explained. As with the questioning instrument, a scribbled word or two indicating the informational content of some of the entries facilitiates subsequent discussion of the analysis.

Self-interruptions by the teacher are indicated by an oblique stroke - 't/'. Some discontinuity is caused when a student's contribution has to be rejected, or when a student fails to respond to a request for a contribution. Such instances are indicated by the combined symbols 'sx'. Lack of clarity may result from muddled thinking, or poor choice of language, or the teacher may use words or phrases not likely to be understood by the children. Whenever the observer judges that a contribution lacks clarity, a ring is drawn around the appropriate symbol.

INSTRUMENT 3 : VARYING THE STIMULUS In the use of this instrument, observers attempt to record teacher behaviors likely to result in useful variation of attention, and certain student behaviors indicating shifts of attention. Attempts to define units of behavior were unsuccessful, so a sign system, instead of a category system, was adopted. The instrument is marked off vertically at thirty-second intervals, as shown in Fig. 2.

Figure 2. Instrument 3 : Varying the stimulus

	1	2	3	4	5	6	7	8	9	10
Teacher movement										
Teacher gesture										
Change in t's speech pattern										
Change in sense-mode or direction										
Students participate verbally										
Students participate physically										

Behaviors are recorded with ticks against the six headings as they occur, but observers only record once on a given line in each thirty-second interval. Timing is approximate, using an ordinary wrist-watch, and observers soon get into a rhythm of recording. The sequential pattern, an important feature of Instruments 1, 2 and 4, is lost by this procedure.

INSTRUMENT 4 : REINFORCEMENT

The unit of behavior is: 'a reaction by the teacher to any one contribution by a student'. Contributions may be oral, written, a piece of practical work, a musical performance, a physical activity, a map, etc.

Three symbols are used. A '+' indicates that, in the opinion of the observer, the teacher's reaction was likely to encourage the student to make further contributions. Reactions likely to discourage further contributions are recorded '-', and a 'o' indicates that the observer judged the teacher's reaction to be neutral in this respect. Fig. 3 shows a recording made on this instrument.

Figure 3. Instrument 4 : Reinforcing

Five-minute episode from history lesson

Accepted verbally	+		+	+				-		+
Rejected verbally		+			o	o	-			
No verbal reaction									-	
Used	+									+

The fourth line, 'Used', corresponds roughly with Flancers' Category 3 behavior (1). It is defined as: Using a student's contribution as the premise for an argument or question or as a link with the next step in the discourse; relating a student's contribution to ideas expressed earlier; or summarizing ideas expressed by students. The entry on the fourth row is linked (as shown) with the symbol recording the contribution which was 'used'.

In the episode recorded in Fig 3 the teacher accepted a student's definition of the word 'tryst' and went on to explain how a tryst developed at Falkirk. He gently rejected the answer to his next question ('no, but

that was a good try'), accepted the next two with a smile and verbal commendation, and simply said 'no' to the next two. The next answer was rejected strongly ('rubbish, laddie'), the next accepted with a sarcastic comment ('nice to see you're still with us, Willie') and he jerked away in angry silence from the next. Eventually he obtained a satisfactory answer which he accepted with praise and used as a premise for the next question.

INSTRUMENT 5 : LECTURE PERFORMANCE

The approach with this instrument is quite different, in that it calls for simple post-observation ratings. After having observed a lesson or part of a lesson during which the teacher has attempted to convey some information, the observer awards a rating on a five-point scale on each of the following dimensions:

Affective set	the extent to which the teacher developed interest in the topic and a desire to succeed in mastering it
Cognitive set	the extent to which the teacher made clear what was to be studied and, where applicable, linked it with previous work
Thread of discourse	the extent to which the teacher showed clearly the connections among the key points and developed the theme clearly and logically
Appropriateness of language	undefined
Fluency	undefined
Closure	the extent to which the teacher consolidated and made explicit what had been learned

INSTRUMENT 6 : CLASSROOM PERSONALITY & SUBJECT COMPETENCE

This instrument may not be used until the rater has observed the teacher on at least three separate occasions. It calls for a rating on a five-point scale on the following dimensions:

warm, kindly sympathetic	v	cold, unkind, aloof
businesslike, systematic	v	slipshod, unsystematic
stimulating, imaginative	v	dull, stereotyped
having clear master of subject	v	having inadequate knowledge

RELIABILITY OF INSTRUMENTS

Tests of observer agreement were made at Stirling University with three observers simultaneously viewing episodes from video-recorded films of Scottish high school teachers, covering English, history, mathematics, Latin, geography, biology, and chemistry. After some modifications, field tests were carried out in schools. Seven pairs of tutors recorded episodes live from lessons taught by 27 interns during a normal teaching practice. The two sets of data cannot, of course, be directly compared.

Table 1. Results of tests of observer-agreement

Percentage observer-agreements, except where stated

	Studio test	Field test	Comment
Questioning	58%	57%	
Structure of discourse	71%	58%	Radical changes between tests
Varying the stimulus	W = 0.80	63%	Kendall's coeff. of concordance (W) Radical changes
Reinforcement	80%	64%	
Lecture performance	r = - 0.06 to + 0.89	r = + 0.41 to + 0.82	Product-moment correlations (+)
Classroom personality & subject competence	r = + 0.13 to + 0.84	r = + 0.62 to + 0.81	Produce-moment correlations (r)

Considering the complexity of the records (for example, a question can be recorded in 36 different ways on Instrument 1), these results are encouraging. Instruments 5 and 6 were not given a fair test during the studio trials, as the observers only saw two short episodes from each lesson. With the exception of the result for appropriateness of language, all product-moment correlations obtained on the field test were significant at p 0.01. An investigation of the stability of the behaviors - the extent to which a given intern produced similar profiles on different occasions - suggested that this tended to be high unless the activities were markedly dissimilar.

It would appear reasonable to conclude that the procedures for lesson sampling developed in this research pass the test of feasibility.

IMPLICATIONS

There is no suggestion arising from this research that copies of the Stirling Lesson-Sampling Instruments be printed and offered for sale. They are only appropriate for institutions following patterns of micro-teaching similar to that on which they were based. Any institution considering using an approach of this kind would need first to examine its own program of professional skills and then to devise suitable instruments for sampling its interns' performances of these skills in the classroom. However, the Stirling Lesson-Sampling Instruments might provide a suitable starting point, and the writer will gladly enter into correspondence with any reader interested in pursuing this line.

Self-evaluation is perhaps the most desirable form of guidance in teaching. If pairs of interns, trained to use the instruments, can work together, one might record while the other teaches. Then, with the evidence of the record as a basis for discussion, the interns might hopefully be led to evaluate their own and one another's performances. This would remove

the potential threat caused by the presence of a tutor, who is known to have an assessing role to play as well as an advisory one.

Instruments of this nature may well have a part to play in teacher-effectiveness research, providing a means of reliably measuring selected variables. The Stirling Lesson-Sampling Instruments were used successfully by Millar (2) to differentiate among teaching performances which subjects in his research were required to evaluate.

References

(1) Flanders, N.A. (1970), Analyzing Teaching Behavior, Addison - Wesley, Reading, Massachusetts.

(2) Millar, C. (1972), "A procedure for analyzing evaluations of observed teaching and its application in measuring outcomes of professional education', unpublished M.Sc. thesis, Stirling University.

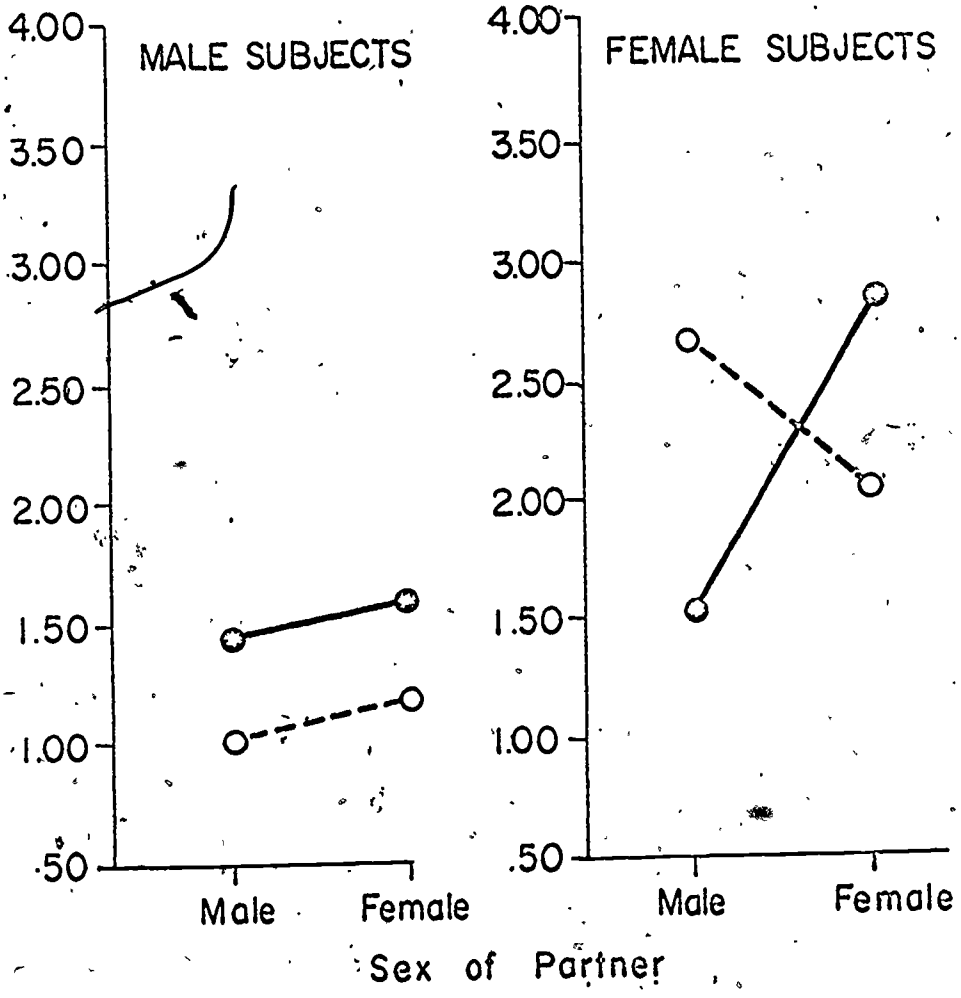

Fig. 1 - Number of altruistic responses given by experimental and control group males and females, with male or female partners.

TABLE 2 (Male Subjects Only)

Card Type

	Relative Gain		Relative Loss		Rel. Gain + Abs. Gain	
Age Group	Exp.	Con.	Exp.	Con.	Exp.	Con.
Youngest	1.688	0.875	0.844	0.969	0.875	0.938
Middle	2.125	1.156	0.656	0.594	1.438	0.500
Oldest	2.093	2.000	1.656	1.000	1.875	2.250

TABLE 3 (Female Subjects Only)

Card Type X Sex of Partner

	Relative Gain		Relative Loss		Rel. Gain + Abs. Gain	
Age Group	Male	Female	Male	Female	Male	Female
Youngest	3.531	3.406	0.688	1.812	3.000	1.625
Middle	3.906	2.688	1.438	2.125	0.500	1.125
Oldest	3.031	3.719	1.844	1.594	3.062	3.625