Dimensions of Classroom Behavior
Measured by Two Systems
of Interaction Analysis

Donald M. Medley
Russell A. Hill

Perhaps the most important and exciting recent development in the study of teaching has been the appearance of several new instruments for analyzing teacher behavior (Medley and Mitzel, 1963; Simon and Boyer, 1967). This development is particularly important to the supervisor because he is now able to record observed teacher behavior more accurately and objectively than ever before. With more accurate records, better feedback and diagnosis are possible, and changes in teacher behavior can be assessed with greatly increased precision.

Effective use of these powerful tools depends to a considerable extent on familiarity with the strengths and weaknesses of each, and specifically on knowing which technique can be used most efficiently and accurately for which objectives. There is presently more need for research into the dimensionalities of the various analytical systems already in existence, the overlap among them, and the unique aspects of classroom behavior tapped by each, than there is for the development of still more new instruments.

This article describes some findings of a study which compared two systems for analyzing classroom interactions. The two methods studied were Flanders' Interaction Analysis Technique (Amidon and Flanders, 1963) and OscAR 4V (Medley, Impellitteri, and Smith, 1966). The subjects of the study were 70 first-year teachers of junior or senior high school English, mathematics, science, or social studies, enrolled in an internship training program for liberal arts college graduates.

During February each teacher was visited by a team of two observers on two different occasions, and during late May or early June each teacher was visited two more times by the same observer team. For 20 minutes on each visit, one member of the team recorded verbal behavior using Flanders' Interaction Analysis Technique, and the other observer recorded the same behavior using OscAR 4V. Twelve teachers were assigned to each team, making a total of 72, but two teachers resigned during the year, reducing this total to 70. The data of the study, therefore, consist of four pairs of records made in the classrooms of each of 70 teachers, or 280 pairs of records in all.

Flanders' system is too well known to need any detailed description here. Suffice it to say that it consists of ten categories of
verbal behavior, and that the observer listens to all the verbal behavior in a classroom and records the dominant category he hears during each three-second interval. The proportion of time during which each of the ten categories is recorded during the period of observations is the primary information used in describing the observed behavior. In addition, a two-way classification called a matrix, which indicates the relative frequency of each of the 100 possible combinations of pairs of categories recorded in successive three-second intervals, is used to produce a more detailed description.

OScAR 4V (Observation Schedule and Record 4, Verbal) is one of a series of "OScARs," each a revision of the last. The observer using it listens to the teacher and tallies each non-interactive statement into one of six categories, and each interaction between teacher and student, or interchange, in a two-way matrix according to the type of entry (question) and the type of exit (teacher reaction to pupil comment). Forty-two different kinds of events may be recorded on OScAR 4V (see Table 1).

Each of the records was scored according to procedures devised for that system. Forty-four "measures" were scored on each Flanders' record; 42 on each OScAR. Analysis of variance revealed that a number of scores did not detect stable differences among teachers; these were discarded. The 75 remaining scores were intercorrelated and the first 10 principal components of the matrix of intercorrelations were extracted and rotated to a varimax solution.

Each of the rotated factors identified a dimension of teacher behavior whose nature could be inferred by inspection of the "measures" which contributed to it. Simplified scoring keys were devised for these ten dimensions, so that a score on each dimension could be obtained on each behavior record. The ten "factor" scores on a record contained most of the information in the record (65 percent), and presented it in terms of fewer and more meaningful variables than the original 75. These ten scores may be said to describe the most important differences detected in the behaviors of these 70 teachers by these two instruments.

Four of the dimensions or factors had to do with questions asked (or answered) by the teacher in developing the substantive content of the lesson; three described student talk and the teacher's reaction to it; and three contained information about how the teacher managed classroom procedures.

Questioning behavior was described in terms of four dimensions: (a) amount of questioning relative to the amount of lecturing by the teacher; (b) quality of questioning—relative numbers of simple questions in comparison with questions calling for more thoughtful answers; (c) difficulty of questions, as reflected in how often the teacher evaluated pupil responses as correct; and (d) source of questions, that is, relative number of questions asked by teacher and student.

Student talk was described on three dimensions: (a) duration—that is, how long the teacher allowed a student to continue to talk without interrupting him; (b) originality—that is, how much student talk not directly responsive to teacher questions was observed; and (c) acceptance—that is, the amount of extended acceptance of pupil talk exhibited by the teacher.

Management behavior was described in terms of three dimensions: (a) amount—that is, number of teacher statements devoted to procedural matters; (b) rebuking—that is, number of times the teacher reprimanded a student; and (c) permissiveness—the number of times the teacher either permitted students to initiate procedure or invited them to do so.

Seven of these ten dimensions are scorable on Flanders' records (accounting for 75 percent of the variance); three are not. Only one of the three procedural dimensions (rebuking) appears in a Flanders record; and no score on question difficulty is available.

Eight of the ten dimensions are scorable on OScAR 4V (accounting for 80 percent of the variance). Both of the scores not available on OScAR records relate to student behaviors (originality and acceptance).

Flanders' instrument appears more sensitive to student behaviors and less able to...
discriminate teacher behaviors related to substantive content from behaviors related to procedure or management. OSCAR is less useful in examining student behavior, but provides more information about how a teacher divides his time between management and instruction, and the quality of both. Both systems seem to measure a number of important stylistic variables of a type with which supervisors and teachers in training are likely to be concerned. Which would be more useful in a given instance would seem to depend on the type of problems which concerned the teacher in question.

Table 1. Summary of Categories of Verbal Behavior on OSCAR 4V

I. STATEMENTS
   A. Teacher Statements. Utterances which neither respond to nor solicit a response from a pupil are classified as follows:
      1. AFFECTIVE. A statement revealing sensitivity to pupil feelings is classified as CONSIDERING. A statement criticizing pupil conduct is classified as REBUKING.
      2. SUBSTANTIVE. A statement containing no affect but referring directly to content to be learned by pupils is classified as INFORMING if it conveys a fact, generalization, or the like, or PROBLEM STRUCTURING if it sets up a question or issue to be solved.
      3. PROCEDURAL. A statement which contains neither affect nor substance is classified as DIRECTIVE if it contains a command or instruction with the force of a command. A statement which does not clearly fall into one of the above categories is classified as DESCRIBING.
   B. Pupil Statements. Utterances by pupils addressed to other pupils are classified as PUPIL STATEMENTS.
   C. Sequence. If a teacher makes two or more successive statements which may be classified in the same category, all except the first are classified as CONTINUING. The first statement in a series of the same kind is classified as INITIATING.

II. INTERCHANGES
   An interchange is an episode in which a pupil says something to the teacher and the teacher reacts.
   A. Substantive Interchanges are those in which the pupil's utterance refers to content to be learned. Such interchanges contain two parts: entry and exit.
      1. Entries. A substantive interchange begins with one of four types of entries:
         a. PUPIL INITIATED. The pupil addresses a statement or question to the teacher.
         b. ELABORATING. The teacher addresses a question to a pupil which refers directly to a previous pupil comment.
         c. DIVERGENT. The teacher addresses a question to a pupil which does not refer directly to a previous pupil comment, and which offers him a choice of two or more acceptable or "correct" answers.
         d. CONVERGENT. The teacher addresses a question to a pupil which does not refer directly to a previous pupil comment and to which there is only one acceptable answer.
      2. Exits from Completed Substantive Interchanges. After the pupil has asked his question or made his answer, the teacher disposes of the answer in one of six ways, called Exits. Exits are first classified according to the information they contain about the correctness or acceptability of what the pupil has said.
         If the teacher clearly indicates that what the pupil has said is correct or acceptable, the interchange is classified as SUPPORTED if praise or enthusiasm is shown, as APPROVED if praise is not given.
         If the teacher clearly indicates that what the pupil has said is incorrect or unacceptable, the interchange is classified as CRITICIZED if disapproval of either the pupil or what he has said is expressed, or as NEUTRALLY REJECTED if no disapproval is expressed.
         If the teacher makes some response to what the pupil says which does not clearly indicate whether it is correct (acceptable) or incorrect (unacceptable), the interchange is classified as ACCEPTED; if the teacher makes no response, it is classified as NOT EVALUATED.
   B. Non-Substantive Interchanges are those in which the pupil's contribution does not refer to content to be learned.
      1. TEACHER-INITIATED non-substantive interchanges are classified as POSITIVE OR NEGATIVE according to the affective content of the teacher's question.
      2. PUPIL-INITIATED non-substantive interchanges are classified as POSITIVE if the teacher supports, approves, or accepts the pupil's suggestion, and as NEGATIVE if he criticizes, neutrally rejects, or ignores it.

May 1969
One other difference between the two instruments was noted. Flanders' system appears to be the simpler of the two, involving only ten categories in comparison to the 21 on which OSCAR 4V is based. The nature of the scoring keys for the two systems reflects this difference. Keys for the Interaction Analysis record tend also to be much simpler than those for OSCAR. In some cases a single category (such as criticizing) is the only one scored on a factor dimension. In one instance, a single cell in the matrix (Teacher Lectures followed by Student Talk-Initiating) stands alone as the measure of one factor dimension (question source). This tendency for single categories to correspond to dimensions on which teachers differ is impressive evidence of the psychological insight of the author of the system.

Keys for OSCAR tend, with few exceptions, to involve larger numbers of categories of behavior. Question source, for instance, which is scored on a single cell in a Flanders matrix, is based on six categories on an OSCAR record.

To the supervisor, the more complex key has its advantages. Inspection of the categories contributing to a dimension and the weights assigned to them gives some useful insights into the nature of the dimension, and an examination of a teacher’s performance on each item often points to a diagnosis of his difficulty and suggests a strategy for change on the dimension.

The complexity of a key also provides some protection against side effects and treating symptoms instead of causes. Suppose, for example, that a teacher wants to change his score on question source—to elicit more questions from students. Diagnosis indicates the best way to do so is by working on one item, that is, by accepting more pupil questions. If he unintentionally changes his behavior on any other item relevant to the dimension in question, he may find he has achieved his objective in terms of the specific item, but has not succeeded in achieving his more important goal of eliciting more pupil questions. The OSCAR records of his behavior will reflect exactly what has happened. Diagnostic information of this type cannot be obtained from the simpler categories of the Flanders system.

These differences have been cited, not to indicate that one system is generally superior to the other so much as to illustrate the points that different instruments should be used for different purposes, and that a skilled practitioner must be able to select and use a variety of tools according to what he is trying to accomplish.

References


—DONALD M. MEDLEY, Senior Research Psychologist, Educational Testing Service, Princeton, New Jersey; and RUSSELL A. HILL, Associate Professor of Curriculum and Instruction, Temple University, Philadelphia, Pennsylvania.