

Two Approaches to Evaluation

Keith Jones and Ann Sherman

Eisner's Educational Criticism model produces a richer description of an English classroom than does Flanders' Interaction Analysis.

Educators are turning to ethnography, jurisprudence, and art criticism for models of evaluation to replace or complement conventional approaches. We have spent three years investigating and articulating one of these models, Educational Criticism (Eisner, 1976). In an attempt to clarify any advantages this approach might have over more quantified methods, we conducted a study comparing Educational Criticism with one of the more popular observation systems, Flanders' Interaction Analysis¹ (Flanders, 1970). By employing both approaches in similar circumstances, we believed we could identify their relative merits.

With this goal in mind, we observed a high school literature class for three weeks. For the first two weeks we used Educational Criticism as our evaluation model; the third week we used Flanders' Interaction Analysis.

Most readers are probably familiar with the Flanders system. An observer records verbal behavior in a classroom, classifying teacher and student talk into categories (Figure 1). At three second intervals the observer determines the type of event taking place and tallies it. Approximately 1,000 responses are coded in a 45 minute period.

Once the coding is completed, tallies are summed for each category. A grand total is figured and percentages are calculated showing the degree to which certain verbal behaviors were observed. To further analyze the data, a matrix of paired observations can be constructed leading to additional ratios and a more complex interpretation of teacher-student interactions. From this matrix patterns of teacher-student interactions are ascertained. These, in turn, can be diagrammed into flow charts that clearly depict the most

frequent behavior sequences and their relative incidence.

Educational Criticism, by contrast, emphasizes understanding the specific conditions under which educational practice proceeds. Education is viewed not

Figure 1. Categories Used in Flanders' Interaction Analysis

Teacher Talk	Indirect Influence	1. Accepts Feeling: accepts and clarifies the feeling tone of the students in a nonthreatening manner. Feelings may be positive or negative. Predicting or recalling feelings is included.
	Direct Influence	2. Praises or Encourages: praises or encourages student action or behavior. Jokes that release tension, but not at the expense of another individual; nodding head, or saying "um hm?" or "go on" are included. 3. Accepts or Uses Ideas of Students: clarifying, building, or developing ideas suggested by a student. As teacher brings more of his/her own ideas into play, shift to Category 5. 4. Asks Questions: asking a question about content or procedure with the intent that a student answer.
Student Talk		5. Lecturing: giving facts or opinions about content or procedures; expressing own ideas, asking rhetorical questions. 6. Giving Directions: directions, commands, or orders with which a student is expected to comply. 7. Criticizing or Justifying Authority: statements intended to change student behavior from nonacceptable to acceptable pattern; bawling someone out; stating why the teacher is doing what is being done; extreme self-reference.
		8. Student Talk—Response: talk by students in response to teacher. Teacher initiates the contact or solicits student statement. 9. Student Talk—Initiation: talk by students, which they initiate. If "calling on" student is only to indicate who may talk next, observer must decide whether student wanted to talk. If student did, use this category.
		10. Silence or Confusion: pauses, short periods of silence, and periods of confusion in which communication cannot be understood by the observer.

¹ Flanders' Interaction Analysis was chosen as an exemplar of the precategorized approach on the basis of its wide usage, its similarity to other counting systems, and the opportunity we had to discuss our work with its author.

as a nomothetic activity (law abiding and predictable) but as ideographic (guided by the unique characteristics of the situation) (Eisner, 1976, p. 136). Because the emphasis is on context, predefined instruments are not used. Instead, the observer immerses herself/himself in the culture of the classroom before choosing the most significant aspects on which to base an evaluation. For this reason the observer must be a seasoned educator. Instead of relying on a pre-categorized instrument to focus his/her attention, he/she relies on perceptive skills sharpened by knowledge and experience.

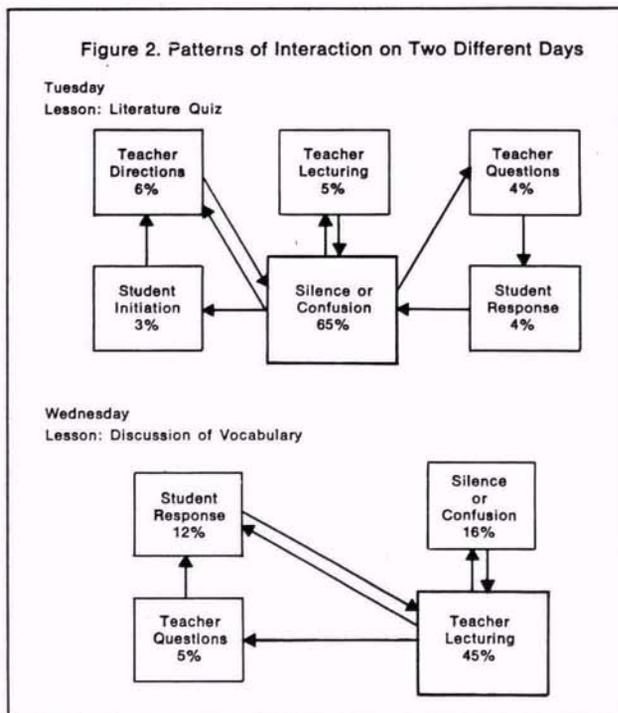
After an extensive observation period (several days to several weeks) an Educational Criticism is written having three aspects: description, interpretation, and evaluation. The descriptive aspect provides a vivid rendering of the qualities of the classroom life. It is ordinarily a narrative that allows the reader an opportunity to experience vicariously what the critic has experienced. The intent is to have the reader know, in a very real sense, what it is like to be in that classroom. Figurative prose may be employed to help the reader empathically understand the critic's perceptions.

The interpretative aspect of an Educational Criticism provides a theoretical analysis of what has been described. Drawing from theory and research, the educational critic analyzes and explicates what has been portrayed.

The evaluative aspect of an Educational Criticism appraises or places educational value on what has been perceived. Here the critic reaches conclusions. He/she judges the program and offers the reader an intelligent opinion as to its educational significance.

Results Using Interaction Analysis

Because there was a high degree of correlation in the interaction data collected by both authors, a Flanders' Interaction Analysis was performed on only one of our two codings. The day-by-day analysis may be summarized as shown.



Each of the five days had different frequencies of teacher talk, student talk, and silence or confusion. As might be expected, on those days when the majority of the period was devoted to class discussion, the most teacher and student talk was found. On those days when there were quizzes the most silence and confusion was found. In addition, on discussion days the sustained teacher talk was greater than on the quiz days when interchanges were shorter. Typical of the verbal interaction patterns are the relationships seen in the diagrams in Figure 2.²

When the direction of the interchange is combined with the percentages of each category, summary statements such as on the next page may be made:

² These summary diagrams were constructed with the aid and advice of Ned Flanders.

Writing for Educational Leadership

We welcome manuscripts about any aspect of curriculum, instruction, supervision, or leadership in education. Papers should be written in direct, readable style and be as brief as possible (eight to ten pages typed double-spaced). We reserve the right to edit for brevity, clarity, and consistency of style.

We usually use footnotes but do not object if authors list all references at the end. For form of

footnotes and references, follow the current edition of Kate L. Turabian, *A Manual for Writers* (University of Chicago Press), or see a recent issue. Double-space everything, including quotations and footnotes.

Please send two copies. Rejected manuscripts are not returned unless the author provides a stamped, self-addressed envelope.

On Tuesday, a major portion of the period was spent in either silence or confusion. This was occasionally interrupted by teacher directions, teacher lecturing, teacher questions, and student initiation. Questions coming from students (student initiation) were most often followed by teacher directions which, in turn, brought about either silence or confusion. There were few, if any, questions when the teacher lectured.

On Wednesday, teacher lecturing accounted for almost one half of the period. Silence or confusion, student response, and teacher questions followed consecutively in frequency of occurrence. Most teacher questions followed teacher lecturing which was then followed by student response. However, there was a good deal of student response (talk by students in response to the teacher) that did not follow teacher questioning but rather teacher lecturing.

Another way of looking at the data is to compare the weekly average frequencies with averages obtained in other high school classrooms (Figure 3). Within this classroom there appears to be a comparatively low amount of teacher praise, teacher questioning, and teacher use of student ideas while at the same time there is a high degree of giving directions and silence or confusion.³ The high degree of silence or confusion and the apparent need to give a considerable number of directions suggest that the students are not paying close attention to the instructor. Another interpretation would be that the high degree of silence or confusion may be related to the small amount of praise and the infrequent use of student ideas.

Figure 3. Daily Average Percent of Time in Each Category Compared With National Averages*

Category	Observed Classroom	National Averages
1. Accepts feelings	0.0%	1%
2. Praises	0.2%	5%
3. Uses ideas of students	1.0%	8%
4. Asks questions	4.5%	14%
5. Lecturing	22.5%	40-60%
6. Giving directions	10.7%	4%
7. Criticizing	3.0%	1%
8. Student response	14.2%	21%
9. Student initiation		
10. Silence or confusion	43.6%	12%

* National averages were given to us by Ned Flanders.

Results Using Educational Criticism

After observing in the classroom for two weeks, each author wrote an Educational Criticism. Each piece was approximately 20 pages long and included description, interpretation, and evaluation. Excerpts from one of these are presented below. The headings indicate the three types of analysis.

Title: "Mrs. G. vs. the Novices"

Description:

The ping-pong match starts with the bell. Mrs. G.,

the professional teacher with 12 years experience, takes on the class of 34 novices. The setting is a typical high school classroom with the U.S. flag in one corner, a bulletin board to one side, and a dominating blackboard. The lighting is good with natural illumination coming from the east and west windows aided by ceiling fixtures. The traditional dark green linoleum playing field has been coordinated with pear-green walls. Everything is familiar to both spectators and players

Interpretation:

Interspersed between plays are occasional interferences of jokes and unrelated conversations, but as usual Mrs. G. is able to gain control. She deftly coordinates activities so the novices can put as much English on the ball as they want and feel that they are winning. In reality, however, she is in command

Evaluation:

What this means educationally is that learning can take place horizontally as well as vertically (Elkind, 1974). Students are able to accumulate knowledge (vertical learning) and also relate it to knowledge they have already acquired (horizontal learning). Elkind points out that in most classrooms horizontal learning is neglected because of the emphasis on vertical acceleration. This he regards as educationally wasteful, for only when there is elaboration of what is learned does it have an impact on one's memory

Here in Mrs. G.'s class students are not apt to forget the vocabulary words they are served when such striking horizontal associations are made as "pissed" with the word "peeved," "Jesse James" with the word "nefarious," and "wearing loud Hawaiian shirts" with the word "gauche." When students are able to improvise as they learn, they find hangers on which to place their newly acquired learnings, which leads to greater retention of the material than if it were force fed

The Two Approaches Compared

In comparing the method and results of the Flanders' Interaction Analysis with that of the Educational Criticism, we believe there are three salient issues: thick versus thin description; objectivity versus subjectivity; and quantity versus quality.

Clifford Geertz (1973) introduces the distinction between "thick" and "thin" descriptions in his book, *The Interpretation of Cultures*. The notion is taken from Gilbert Ryle, who distinguishes between descriptions of phenomena that provide rich interpretation and those that provide meager interpretation:

Consider, he says, two boys rapidly contracting the eyelids of their eyes. In one this is an involuntary twitch; in the other, a conspiratorial signal to a friend. The two movements are, as movements, identical; from an I-am-a-camera "phenomenalistic" observation of them alone, one could not tell which was twitch and which was wink, or

³ It should be noted that on two of the five days of observation a portion of the hour long periods was devoted to quizzes, which added to the composite average of silence or confusion. Nevertheless, there was no day on which the percentage of silence or confusion did not exceed the national average.

indeed whether both or either was twitch or wink. Yet the difference, however unphotographable, between a twitch and a wink is vast; as anyone unfortunate enough to have had the first taken for the second knows (p. 9).

A thin description would be that each of the two boys was rapidly contracting his eyelids. A thick description would be that the one boy had an involuntary twitch due to a nervous disorder, while the other boy attempted to signal his friend that a conspiracy was in motion. Geertz holds that in order to accurately describe movements one must decipher the whole network of structured and unstructured communication patterns that give movements the status of gestures. "Analysis, then, is sorting out the structures of signification . . . and determining their social import" (p. 9). Geertz's point is that "thick" description involves understanding the cultural significance of movements—it involves interpretation. This is not to imply that "thin" description does not also involve interpretation; even to say that something is an "eye" or is "rapidly contracting" involves selection and interpretation of the world. The difference is that "thick" description involves interpretation of the particular social significance of behavior.

In comparing the two types of evaluations we found that our Flanders' Interaction Analysis was thin while our Educational Criticisms were thick. Note these examples:

Interaction Analysis

Teacher lecturing
35%

Silence or Confusion
16%

Student Initiation
5%

Educational Criticism

vs. Absently sucking on the frames of her glasses and gazing out the window, Mrs. G. wistfully replies that it is "almost not worth living if you're just going to exist."

vs. The novices are whooping it up, paying little attention to their opponent who is busily searching for a quiz.

vs. A large blond male, who playfully swats at many of Mrs. G.'s remarks, boldly finesses by asking, "If someone walked into a dinner party wearing a loud Hawaiian shirt, would that be gauche?"

We believe there is an advantage to thick descriptions in educational evaluation. By answering such questions as "What is the context in which the behaviors take place?" "What is the importance of a particular action to the participants in a program?" "How does a particular behavior relate to other behaviors?" and "Why do the participants behave as they do?", thick descriptions can disclose information critical to the understanding of educational programs. By dealing with the particulars of a classroom culture and using these as examples to be interpreted and appraised, Educational Criticism sheds light on substantive concerns often difficult to observe and explicate.

Can Evaluation Be Objective?

Some might argue that standardized procedures ensure objectivity; that methods like Educational Criticism are subjective because they express only one individual's point of view. They might claim that without pre-defined categories there is no way of controlling what the observer will attend to, so interpretation may be biased.

We maintain that every method of evaluation is subjective in the sense of imposing a perspective which biases. One biased view held by a number of people is no more objective (in the sense of conveying the entire meaning of a situation) than are a number of different "biased" viewpoints. That is, all evaluation instruments influence what is seen as "objective" reality. The Flanders' system is not neutral; it imposes a very specific framework on "reality." It reflects certain assumptions about what is educationally significant—verbal behavior. In this sense it is as subjective as Educational Criticism. Instead of having the observer judge things according to his or her own individual point of view, each observer adopts the perspective of the system. This in no way assures a

A framework for planning curricula

Curriculum Building for Adult Learning

By John R. Verduin, Jr.

The author's goal is to provide a general model for educators to use in effective curriculum building. The principles derived here are broad enough to apply to any adult-learning program. \$9.00

Also of interest...

Censors in the Classroom

The Mind Benders

By Edward B. Jenkinson

Documenting the expanding role of parent censors in the public schools, Jenkinson offers informed discussion of the issues at hand and suggestions for effectively countering the efforts of pressure groups.

\$12.50

SOUTHERN ILLINOIS UNIVERSITY PRESS
P.O. Box 3697, Carbondale, Illinois 62901

Duplicating

Educators are authorized to reproduce a single article from this publication without making a written request provided that 1) duplication is for an educational purpose in a not-for-profit institution; 2) copies are made available without charge beyond the cost of reproduction; and 3) each copy includes full citation of the source. Permission to reproduce more than one article will be granted under the same conditions to those who make a reasonable request in writing.

This authorization does not apply to materials, if any, for which ASCD doesn't hold the copyright. Any such materials are so identified.

more complete view of "objective" reality; it simply assures that people will see the same slice.

Closely related to the issue of objectivity vs. subjectivity is the issue of quantity vs. quality. Educators often place greater faith in numerical data than in analyses that do not quantify their observations. Indeed, there are certain advantages to numerical data. It allows interpretations to be made based on statistical differences, ratios, and correlations. In Flanders' Interaction Analysis, by becoming aware of how often certain behaviors occur and comparing these to a standard, one can make judgments as to the desirability of increasing or decreasing certain behaviors. What is overlooked is that how often a behavior occurs may be less important than the way it occurs. For example, corporal punishment occurring a small percentage of the time may have a major impact on the demeanor of teachers and students. Likewise, an event that happens frequently, such as teacher lecturing, may have little impact on students.

In our own criticisms of Mrs. G.'s classroom, we found it relevant to our evaluation that students felt free to make playful word associations. It was not the number of times that students made such associations that was revealing, but rather the permissiveness that allowed students to respond with unrestrained word associations like "pissed" with the word "peevd" and "the wearing of a loud Hawaiian shirt" with the word "gauche."

In the final analysis, the most striking difference between these two approaches to the description of educational life is that one attends mainly to the incidence of behavior while the other attends to the social meaning of action. Flanders' Interaction Analysis is like a net cast over a teacher-student discourse. It permits calculation of frequencies of certain behaviors and sequences of events. Using it one can determine amounts of overt behavior and the order in which they occur. Such an analysis may be useful when the particular context and character of the events are unimportant.

By contrast, Educational Criticism emphasizes the unique social meaning of classroom events. It is

like a portrait of the life teachers and students spend together, with vivid images and probing interpretations. From it one can learn something of the character of a classroom and the educational import of what goes on there.

Quality and Meaning

For the past 25 years classroom research has concerned itself primarily with the quantification of behavior. We believe the scope of attention should be broadened to include the quality and meaning of events shaping the lives of students. As LaPiere (1943) cogently states, "Quantitative measurements are quantitatively accurate; qualitative evaluations are always subject to errors of human judgment. Yet it would seem far more worthwhile to make a shrewd guess regarding that which is essential than to accurately measure that which is likely to prove quite irrelevant" (p. 237). *ET*

References

- Elkind, David. *Children and Adolescents: Interpretive Essays on Jean Piaget*. New York: Oxford University Press, 1974.
- Eisner, Elliot. "Educational Connoisseurship and Educational Criticism: Their Form and Functions in Educational Evaluation." *The Journal of Aesthetic Education* 10 (1976).
- Flanders, Ned. *Analyzing Teaching Behavior*. Menlo Park, Calif.: Addison-Wesley, 1970.
- Geertz, Clifford. *The Interpretation of Cultures*. New York: Basic Books, 1973.
- LaPiere, Richard. "Attitudes vs. Actions." *Social Forces* 13 (March 1943): 237.
- Levine, Murray. "Scientific Method and the Adversary Model: Some Preliminary Thoughts." *American Psychologist* 29 (September 1974): 661-77.
- Magoon, A. Jon. "Constructivist Approaches in Educational Research." *Review of Educational Research* 47 (Fall 1977): 651-93.
- Rist, R. C. "Ethnographic Techniques and the Urban School." *Urban Education* 10 (1975): 86-108.



Keith Jones and Ann Sherman are doctoral candidates at Stanford University, Stanford, California.

Copyright © 1980 by the Association for Supervision and Curriculum Development. All rights reserved.