

Resolving Old Dilemmas in Diagnostic Evaluation

Supervisors can help teachers identify their strengths and weaknesses by systematically comparing data from three sources: classroom observation, teacher self-analysis, and student descriptions of teaching practices.

The ineffectiveness of teacher evaluation in public schools has been noted repeatedly over the years by scholars and practitioners alike. Despite these criticisms, however, emphasis on teacher evaluation has grown, and advocates continue to claim it improves teaching. Stipnieks' 1981 study of recent changes in evaluation systems revealed almost universal commitment to improving teaching through evaluation. State mandated programs in California (Popham 1975), Georgia (1983), Florida (Fry et al. 1985), and Texas (1984) have all been politically supported as efforts toward "excellence" in teaching.

Supervisor's Dilemma

The combination of abusive practices of the past, crudely fashioned current practice, and restrictive state mandates gives little hope that these approaches will improve teaching. This makes the supervisor's dilemma substantial. To neglect evaluation as a tool for improving teaching and learning is like flying a plane without a compass or radio.

Faced with few alternatives, supervisors have turned to management by objective (Redfern 1978), clinical supervision (Cogan 1973), connoisseurship (Eisner 1975), and Hunter's (1980, 1983) methodology as ways to be ana-

lytical in mode, yet developmental in effect. To some extent these are useful strategies for improving teaching. None of them, however, addresses the most serious defects of traditional teacher evaluation ritual, and they simply do not adequately guide the teacher.

Diagnostic Evaluation

On the other hand, diagnostic evaluation can promote good teaching for most teachers because it has the following features.

1. Teaching behaviors are described in detail.
2. Descriptive data are analyzed by predetermined criteria of desired behavior.

Classroom observations from several sources provide information for use in teacher evaluations



3. Teacher strengths and weaknesses are clearly identified via data manipulation with minimum opinion.

4. Diagnoses lead to action alternatives.

5. Diagnosed needs for improvement are so explicit that individual teachers can initiate changes.

Multiple Data Sources

The diagnostic methodology that has been field-tested with some promising results is called *congruence analysis*. This term means making diagnoses from two or more sets of data based on common criteria. When several reasonably valid data sources report congruent or similar descriptions of a specific practice, then practitioners can make decisions for action based on this congruence.

The several sources of reasonably valid data regarding teaching practices are the teacher as *introspector*, the observer as systematic *describer*, and students as thoughtful *interactors* (Harris 1986). Other data sources are, of course, available and potentially

useful (McGreal 1984). However, I shall consider only these three because they can all be low inference, are potentially valid regarding teaching practices in the classroom,¹ and are readily available with existing methodology.

When a teacher-introspector uses a set of performance criteria to analyze his or her teaching, a descriptive profile of perceived practices can be produced. Such a profile has a chance of being reasonably valid if provided thoughtfully by a teacher using discriminating instruments. Such instruments are provided by Pearson (1980) and Harris and Hill (1982) for teacher evaluation, and by Bailey (1985) for administrator and supervisor evaluation.

These criteria are not the simple checklists or self-rating scales that have been fully discredited in the past.

They use forced choice or Q-sort formats so that discriminating descriptions of teaching practices can result.

The observer-describer can be a supervisor or other professionally experienced individual. Trained observers using explicit criteria with descriptive or coding procedures, such as those advocated by Flanders (1970), Harris (1985), Hunter (1983), and Medley and others (1984), can produce highly reliable, nonjudgmental records.

Student-interactors as data source are most controversial. Teachers have doubts and reservations, however, a growing body of evidence from both research and practice indicates that student source data, when restricted to simple description of teaching practices, are quite valid. In the Harris/Hill



“The diagnostic methodology that [shows] some promising results [in teacher evaluation] is called congruence analysis [or] making diagnoses from two or more sets of data. . . .”

DeTEK system (1982), we used student inventories designed for easy reading at the third-grade level. These simple checklists call for no opinion or value judgments. They do, of course, presume a minimum level of reading comprehension, experience in the classroom of the designated teacher, and criteria restricted to practices readily observable to and understood by third-grade students.

Diagnostic Methodology

When these three data sources are used with a teacher as the focus, congruence analysis takes the following form.

1. Data are scaled from high to low on each specific criterion separately for each source. A three-level scale is generally used to estimate the extent of performance as defined below.

2. Scaled estimates of data are displayed in a diagram such as Table 1.

3. Diagnoses are determined on the basis of the following three categories of congruence:

a. High congruence + high scaled estimates = probable “accomplishment” or clear, unequivocal evidence that the specific criterion performance is a part of the teacher’s repertoire.

Table 1
Hypothetical Congruence Analysis Process for Diagnosis of Specific Teaching Performances

Criteria of Teaching Performance	Data Sources/Scaled Estimates			Congruence Level	Diagnosis
	Self Introspector	Observer Describer	Student Interactor		
W	3	3	3	High/High	Accomplishment
X	1	1	1	High/Low	Need for improvement
Y	2	2	2	High/Moderate	Uncertainty or need to upgrade
Z	2	1	3	Low/Mixed	Uncertainty

b. High congruence + low scaled estimates = probable “need for improvement” or clear, unequivocal evidence that the specific criterion performance is not a part of the teacher’s repertoire.

c. Low congruence + mixed scaled estimates = uncertain diagnosis or no clear knowledge about the criterion performance.

The analysis of each set of three scaled estimates for each specific criterion produces a set of diagnoses as defined above.

Refining Diagnoses

In a reading teacher’s classroom, the focus for diagnostic analysis produced the information summarized in Table 2. The scaled estimates of performance for each of these four criteria reflect different diagnoses. The estimates of all three data sources strongly support criterion 4C-(1). *H* refers to a high level of congruence between observed practices and the criterion specifications by the teacher (self) and another observer. The number of students (14) out of the total in the

classroom (15) also supports evidence of the practice. Uncertainty is reflected in data for 4C-(2) with little congruence among data sources. Needs for improvement are suggested by the consistently low levels of estimates by all three sources for Criterion 4C-(3) and 4C-(4).

In Table 3, guidelines for diagnosis and action are suggested for various data patterns for any given criterion. Notice that the three basic diagnoses—*need*, *accomplishment*, and *uncertainty*—can be extended substantially using different kinds of congruence patterns. Three highly congruent data sources reporting moderate levels of evidence can be diagnosed as a need for upgrading or refining the practices related to the criterion.

Similarly, incongruence patterns offer diagnoses of several kinds. When data sources are all at odds with each other, uncertainty is clear and further data gathering seems logical. However, when two of three data sources are in agreement, the deviant source may need further attention.

Table 2
Diagnostic Analysis in Sixth-Grade Reading Class

Criteria	Self Introspector	Observer Describer	Student Interactor	Diagnosis
4C-(1) Uses current events	H	H	14/15	Accomplishment
4C-(2) Uses multiple sources	H	M	4/15	Uncertainty
4C-(3) Gives topic choices	N	N	1/15	Need for improvement
4C-(4) Encourages individual projects	N	N	9/15	Need for improvement

Table 3
Guidelines for Diagnosis and Action for a Variety of Data Patterns

Sources and Levels			Diagnosis	Suggested Action
(1)	(2)	(3)		
High	High	High	Accomplishment	Reinforce-Reward
Low	Low	Low	Need for improvement	Develop growth plan
Moderate	Moderate	Moderate	Need for upgrading or refinement	Encourage independent efforts
High	Low	Moderate	Uncertainty a	Gather further data
High	Low	Low	Uncertainty b Possible leniency error	Restudy data source (1)
High	High	Low	Uncertainty c Possible biased sample	Restudy data source (3)
Moderate	Low	Low	Uncertainty d Mild leniency error	Probable need for improvement
High	High	Moderate	Uncertainty e Mild estimation error	Possible accomplishment with further refinement

Action Alternatives

Diagnoses with suggested action are most useful in improving instruction. Ideally, every diagnosis should have action alternatives distinctly different from all other diagnoses.

In Table 3, direct links are proposed between each of eight data patterns, their diagnoses, and actions to follow. Each diagnosis employs a consistent pattern of logic based on previously stated assumptions.

A congruent pattern of high estimates leads to a diagnosis of "accomplishment" with action alternatives including reinforcement or reward. Conversely, a pattern of congruent low estimates is diagnosed as "need for improvement." At least one logical action alternative is to develop a growth plan for promoting and guiding such improvement. A pattern of highly congruent moderate estimates of performance is logically diagnosed as neither accomplishment nor need for urgent improvement. Instead, it is diagnosed as "need for upgrading or refinement with an action alternative possibly less structured than the more urgent need diagnosis."

Uncertainties shown in Table 3 result from five different patterns of data. Accordingly, each diagnosis and suggested action is slightly different. When a lone data source is sharply divergent from two others, error in

data gathering or estimation is a reasonable explanation, and those data should be studied more carefully. However, when a single data source deviates only to a mild degree, known tendencies toward leniency or error estimation are reasonable considerations, and useful actions can be recommended.

A Practical Solution

The use of teacher evaluations for improving teaching practices is widely intended but rarely applied. However, substantial progress toward applying evaluation technology and research on teaching effectiveness to supervisory support can be made by using diagnostic evaluation methods. They are relatively simple and fairly powerful, and they offer opportunities for teachers and supervisors to work in a collaborative, mutually supportive fashion to improve teaching. □

1. For a detailed discussion of data sources and their advantages and limitations, see Ben M. Harris, *Developmental Teacher Evaluation* (Boston: Allyn and Bacon, Inc., 1986), Chapter 6.

References

Bailey, Danita "The Relationships Among Supervisor Competencies, Job Expectations and Position Types." Doctoral diss., University of Texas at Austin, 1985.

Cogan, Morris *Clinical Supervision*. Boston: Houghton Mifflin Co., 1973.

Eisner, Elliot W. "Some Alternatives to Quantitative Forms of Educational Evaluation." *Thrust for Educational Leadership* 5 (November 1975): 13-15.

Flanders, Ned A. *Analyzing Teaching Behavior*. Reading, Mass.: Addison-Wesley, 1970.

Fry, Betty, et al. "Case Study: Using Research on Effective Teaching to Improve Teacher Education in Florida." *R & D Center Review* 111 (Spring-Summer 1985): 11.

Georgia Department of Education "Guidelines for the Implementation of Performance-Based Certification." Mimeographed. Atlanta: Division of Staff Development, 1983.

Harris, Ben M. *Developmental Teacher Evaluation*. Boston: Allyn and Bacon, Inc., 1986.

Harris, Ben M., and Jane Hill. *Developmental Teacher Evaluation Kit*. Austin, Tex.: Southwest Educational Development Laboratory, 1982.

Harris, Ben M., Kenneth E. McIntyre, Vance Littleton, and Dan Long. *Personnel Administration in Education*, 2nd ed. Boston: Allyn and Bacon, 1985.

Hunter, Madeline "Script Taping: An Essential Supervisory Tool." *Educational Leadership* 41, 3 (November 1983): 43.

Hunter, Madeline. "Six Types of Supervisory Conferences." *Educational Leadership* 37 (February 1980): 408-412.

McGreal, Thomas L., et al. "Artifact Collection." *Educational Leadership* 41 (April 1984): 20-21.

Medley, Donald, Homer Coker, and Robert Soar. "Measurement-Based Evaluation of Teacher Performance." New York: Longman, 1984.

Pearson, Harry Delmer. "Development of a Forced-Choice Teacher Behavior Scale." Doctoral diss., University of Texas at Austin, 1980.

Popham, W. James. *Educational Evaluation*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1975.

Redfern, George. "Evaluating Teachers and Administrators: A Performance Objective Approach." Boulder, Colo.: Westview Press, 1978.

Stipnieks, Anita T. "A Study of Teacher Evaluation Practices in Selected School Districts in the U.S." Doctoral diss., University of Texas at Austin, 1981.

Texas Education Reform Bill. 68th Texas Legislature, 1984.

Ben M. Harris is Professor of Educational Administration at the University of Texas at Austin, Education Bldg. 310, Austin, TX 78712.

Copyright © 1987 by the Association for Supervision and Curriculum Development. All rights reserved.