

Evaluation and Professional Improvement Aspects of the Florida Performance Measurement System

Thorough research and step-by-step testing produced formative and summative instruments that accurately identify teaching behaviors.

What is pedagogical knowledge, and how can it be organized to facilitate the professional development of the teacher? These are the questions we have been attempting to answer with the Florida Performance Measurement System (FPMS) during the last six years.

Our Assumptions

Improvement of pedagogical practice requires a body of objective knowledge about teaching procedures and their effects. Only if teachers have such tested knowledge can they make proper diagnoses and select procedures that satisfy professional expectations and requirements. In the evaluation of teachers, we attempt to determine the degree to which teachers have attained such knowledge and the degree to which they apply it in their instruction.

Learning is the product of teaching. It is the dependent variable against which we attempt to measure teacher effectiveness. However, learning is conditioned by factors beyond those under the teacher's control. This in turn conditions the manner in which we evaluate teachers. For example, we can all agree that no one should be held responsible for outcomes beyond one's influence, but the teacher can and should be held responsible for following valid methods of diagnosis and pedagogical procedures. No less than this is expected of any professional whether a physician, an agriculturalist, or whatever.

The Knowledge Base

In the first phase of developmental work, we assembled the research on generic modes of teaching. We obtained this research largely from two sources: process-product and experimental studies, each of which yields knowledge of effective practice.

In the second phase, we organized the research results into six domains of teaching behavior: planning; management of student conduct; organization and development of instruction; presentation of subject matter; verbal and nonverbal communication; and testing; student preparation, administration, feedback. These domains consist of 34 concepts, 124 indicators with examples, and supporting research. Here is an example of a concept and one of its indicators.¹

Concept: Presentation of interpretive (conceptual) knowledge.

Definition: Teacher performance involved in analyzing and presenting information to facilitate the acquisition of concepts.

Indicator: Identifying attributes.

Definition: Teacher utterances or actions that give a category and identify the attributes for telling whether or not a given object or event belongs to the category, or utterances or actions that induce students to do so.

Example: In a sixth-grade class the teacher says, "A triangle is a plane figure with three angles and three sides." She continues, "This is a plane figure (draws triangle on board), because it has length and no depth. Note that this

plane figure has three sides, which makes it different from all other plane figures."

Development and Uses of Instruments

The third phase of our research consisted of developing six observation instruments, five formative and one summative, for coding teacher performance in the classroom. We derived items on the instruments from the research-based concepts and indicators.

In the tasks of identifying the performance on which teachers need assistance, each type of instrument serves a special function. The summative instrument (fig. 1) first serves a screening function, enabling a trained observer to identify the problem areas such as management of student conduct or organization and conduct of instruction in which the teacher is experiencing difficulties.

The formative instruments (fig. 2) help in pinpointing the specific teacher behaviors that require remediation. For example, the screening observation may indicate that a teacher is experiencing trouble with student misbehavior, but not the specific obstacle. The specific difficulty, say fragmenting group movement, can be identified by observing the teacher's performance more closely with an appropriate formative instrument.

The summative instrument can also be used as a means of comparative scoring of teachers as part of the final

evaluation. When this is done, scores are based on normative data of a comparable group (e.g., beginning teachers or experienced teachers).

Tests of the Summative Instrument

An instrument designed to measure and evaluate the effectiveness of teaching performance, like similar instruments in any field, must be reliable, valid, and normed. The fourth phase of the Florida System's development consisted of testing these properties of the summative instrument. Even a

brief digest of the studies of these properties over the last four years is too extensive for presentation here. Besides, they are available in ERIC reports.² However, a brief description of the results follows.

The summative instrument was submitted to tests of reliability, using trained observers—teachers and school administrators—in testing intercoder agreement, stability, and discriminant reliability. The coefficients from these tests ranged from $r = .79$ to $r = .98$.

Norming of the summative instru-

ment was based on 1,223 observations conducted by trained and certified observers in grades K-12. Results of the norming study showed no differences in teacher scores on the FPMS attributable to any of the frame factors—sex, race, experience, class enrollment, SES, and so on. However, there was a significant difference in scores between elementary and post-elementary grades. We adjusted this difference by creating two norm groups. We corrected a difference in instruction-format variables through the scoring procedure.

Domain	Effective Indicators	Frequency	Frequency	Ineffective Indicators	
3.0 Instructional Organization and Development	1. Begins instruction promptly			Delays	
	2. Handles materials in an orderly manner			Does not organize or handle materials systematically	
	3. Orients students to classwork/ maintains academic focus			Allows talk/activity unrelated to subject	
	4. Conducts beginning/ending review				
	5. Questions: academic comprehension/ lesson development	asks single factual (Domain 5.0)			Poses multiple questions asked as one, unison response
		requires analysis/ reasons			Poses nonacademic questions/ nonacademic procedural questions
	6. Recognizes response/amplifies/gives corrective feedback			Ignores student or response/expresses sarcasm, disgust, harshness	
	7. Gives specific academic praise			Uses general, nonspecific praise	
	8. Provides for practice			Extends discourse, changes topic with no practice	
	9. Gives directions/assigns/checks comprehension of homework, seatwork assignment/ gives feedback			Gives inadequate directions/no homework/ no feedback	
10. Circulates and assists students			Remains at desk/circulates inadequately		
4.0 Presentation of Subject Matter	11. Treats concept—definition/attributes/ examples/nonexamples			Gives definition or examples only	
	12. Discusses cause-effect/uses linking words/applies law or principle			Discusses either cause or effect only/uses no linking word(s)	
	13. States and applies academic rule			Does not state or does not apply academic rule	
	14. Develops criteria and evidence for value judgment			States value judgment with no criteria or evidence	
5.0 Communication: Verbal and Nonverbal	15. Emphasizes important points				
	16. Expresses enthusiasm verbally/ challenges students				
	17.			Uses vague/scrambled discourse	
	18.			Uses loud-grating, high-pitched, monotone, inaudible talk	
2.0 Mgt. of Std. Conduct	19. Uses body behavior that shows interest—smiles, gestures			Frowns, deadpan or lethargic	
	20. Stops misconduct			Delays desist/doesn't stop misconduct/ desists punitively	
	21. Maintains instructional momentum			Loses momentum—fragments nonacademic directions, overdwells	

Fig. 1. Summative Observation Instrument

Four studies have been completed to estimate the predictive validity of the summative instrument. In these studies, scores of the summative instrument were the independent variable, student task engagement was the intervening variable, and measures of student achievement the dependent variable. These studies have consistently shown positive and significant relationships between instrument scores and student task engagement, and between student task engagement and student achievement. In all cases, instrument scores have shown a positive direct relationship to student achievement, and in all but one case this relationship was significant.

Performance Measurement in Florida

Florida school districts have used the Florida System extensively as part of their annual evaluation program and in the Beginning Teacher Program. Its use in Florida's Master Teacher Program was challenged by a teacher's union (*Sweeney v. Turlington* 1986) on the grounds of validity and reliability, which included questions of observer training and accuracy, and instrument norming and scoring. The hearing officer upheld the System's use in the Master Teacher Program on all counts.

Factor analysis conducted on the data collected from the Master Teacher Program ($n = 36,000$ observations of 18,000 candidates) supports the domain structure, as currently defined, and the equal weighting of items in scoring the summative instrument.

To sum up, the Florida Performance Measurement System appears to meet the following legal and professional requirements: validity (content and predictive), reliability, and job relatedness.

Observer Training Program

As developmental work proceeded, we initiated a training program. Accurate observation of teacher performance with FPMS instruments requires trained observers. This requirement is common to all objective observation systems. Additionally, unless the observer, even though trained, has nothing at stake regardless of the outcome of the measurement, the observation may be biased. If a teacher is being

2.0 Management of Student Conduct				
Category	Effective Indicators	Frequency	Frequency	Ineffective Indicators
2.1 Rule Explication and Monitoring	Specifies a rule			Does not specify when rule needed
	Clarifies a rule			Does not clarify rule
	Practices rule			
	Reprimands rule infraction			Does not correct rule infraction
2.2-2.4 With-itness: Desist, Quality, Overlapping	Stops deviant behavior			Does not stop deviancy/deviancy spreads
	Corrects worse deviancy			Corrects lesser deviancy
	Desists student causing disruption			Desists onlooker or wrong student
	Suggests alternative behavior			Uses rough, angry, punitive desists
				Uses approval-focused desist
	Attends task and deviancy simultaneously			Ignores deviancy, continues task/ ignores task, engrosses in deviancy
	Attends to two instructional tasks simultaneously			Ignores other students needing help/drops task, engrosses in intrusion
2.5 Group Alert	Poses question—selects reciter			Selects reciter—poses question
	Alerts class/calls on one reciter			Alerts group—unison response
	Alerts nonperformers			Ignores nonperformer
2.6-2.7 Movement Smoothness/Slowdown	Ignores irrelevancies/ continues on task			Reacts to or interjects irrelevancies/flip-flops/dangles
	Gives short, clear nonacademic directions			Overdwells or fragments nonacademic directions
	Moves whole/subgroup			Fragments group movement
2.8 Praise	Praises specific conduct			Uses general conduct praise
	Praises non-deviant, on-task behavior			
	Gives low-key, quiet praise			Uses loud praise
	Uses contingency praise			
	Uses authentic, varied, warm praise			
	Controls class reaction to misconduct			Allows class to reinforce misconduct

Fig. 2. Formative Classroom Observation Instrument

observed to determine retention or promotion in position or salary, the teacher's principal may be a poor choice as the observer. If the purpose, on the other hand, is to determine where a teacher needs assistance, the principal's observations might be less questionable. However, even disinterested observers require training and verification of competency, if their observations are to be considered dependable.

Florida provides an extensive training program involving primarily principals and supervisors. The training program consists of a three-day period of observation training followed by a criterion test of observation competence, an examination on teacher effectiveness research, a coding quiz, and periodic update sessions. Approximately 7,000 principals and supervisors and 3,000 teachers have qualified as observers.

Special Features and Advantages

We must emphasize that the Florida System is not designed for rating teachers, but rather for observing and coding the behavior of teachers and for the subsequent comparative measurement and evaluation of teacher performance in the classroom. Unlike rating scales that require the observer to code and evaluate simultaneously, the FPMS requires the observer only to code teacher behavior (independently of his or her opinion of its value). This separation of observation and evaluation removes a primary source of subjectivity and error.

The fact that the system does not require a particular mode or style of teaching is one of its most important characteristics. The indicators can be incorporated in the performance of the teacher regardless of whether the instruction is inductive or deductive, highly structured or loosely structured, teacher-centered or student-centered. The indicators are simply elements of effective teaching, and teachers can build them into their classroom performance at will. The same can be said about the teacher's instructional style. Whether the teacher smiles one way or another, as long as it is genuine, makes no difference. Or again, the teacher's particular way of asking questions, as long as they are clear, is of no particular significance.

Moreover, the System's research base establishes its professionalism. When teachers understand the research underlying their performance and realize that what they are doing is not based on opinion or mere personal experience, they feel much more secure in their behavior and are likely to act with more enthusiasm and confidence than if what they do has no research support.

The research underlying the Florida System also provides a subject matter for pre- and inservice programs of teacher preparation.³ This is especially true for programs leading to the preparation of beginning teachers for those areas covered by the domains of the system. Teachers who are prepared in the skills of planning, management of student conduct, organization and development of lessons, instruction in the forms of knowledge, in communication, and in the preparation of students to take examinations are well on the way to becoming effective classroom teachers. It is precisely in such areas of teacher performance that our studies of beginning teachers, as well as experienced teachers, show that they need assistance. It is not an unusual occurrence to hear experienced teachers say that they have profited from a study of the research base of the Florida System.

The Florida System also provides a means whereby districts and colleges can receive information on the success of their respective programs of preparation. The first set of observations made at the very beginning of a teacher's classroom experience, when properly interpreted, supplies colleges and departments of education with information about the quality of their products. This same information, given to the districts and followed by observations of the teacher after a year's preparation and experience, provides the basis for assessing the Beginning Teacher Program in the various districts. It is now possible for a district to receive information on the effectiveness of its inservice program for beginners and to pinpoint the weaknesses in that program by using data gleaned from portfolios kept by the various districts. Furthermore, the Florida System can supply each beginning teacher with information indicating where he or she is either strong or weak in teaching performance.

The Florida Performance Measurement System is among the first, if not the first, system to break completely with the rating-scale mode of observation and evaluation of teacher performance. It is based on research knowledge of teacher effectiveness rather than consensus of opinion. While pedagogical research is still in its early stages of development, it is what we have, and it is folly to ignore it. Furthermore, the Florida System does not cover every aspect of a teacher's work, thus leaving important teacher behaviors to be considered by other means. However, it does embrace important generic classroom behaviors. It has been extensively tested and therefore provides a valid and reliable means of diagnosing and evaluating classroom performance. □

1. Panhandle Area Educational Cooperative, *Domains Knowledge Base of the Florida Performance Measurement System* (Chipley, Fla.: Panhandle Area Educational Cooperative, 1983).

2. *Teacher Evaluation Project Final Report for 1982-1983* (Document No. SP 027190), Tampa: University of South Florida (ERIC Document Reproduction Service No. ED 266 121); *Teacher Evaluation Study: Report for 1983-1984* (Document No. SP 027191), Tampa: University of South Florida (ERIC Document Reproduction Service No. ED 266 122); *Teacher Evaluation Project, 1984-1985* (Document No. SP 027189), Tampa: University of South Florida (ERIC Document Reproduction Service No. ED 266 120); *Teacher Evaluation and Assessment Center: Report for 1984-1985* (Document No. 026680), Tampa: University of South Florida (ERIC Document Reproduction Service No. ED 266 089); *Teacher Evaluation Study: Report for 1985-1986* Tampa: University of South Florida, in press.

3. Nineteen *Teacher Learning Packages* are planned. The following are available from the Panhandle Area Educational Cooperative, 411 West Blvd., Chipley, Fla.: *Planning for Instruction, Using Questioning Techniques, Conducting Effective Lessons, Teaching Concepts, Using Time Effectively, Specifying Student Behavior, and Constructing Tests*.

B. Othanel Smith is Professor of Education Emeritus, living at 854 Island Way, Clearwater, FL 33515. **Donovan Peterson** is Professor of Education, and **Theodore Micceri** is Research Associate, both at the University of South Florida, 4202 E. Fowler Ave., Tampa, FL 33620.

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