DOCUMENT RESUME

ED 360 325 TM 020 122

AUTHOR Jaeger, Richard M.
TITLE Live vs. Memorex: Psychometric and Practical Issues in the Collection of Data on Teachers' Performances in the Classroom.
PUB DATE Apr 93
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; Classroom Research; Comparative Analysis; *Data Collection; Documentation; Educational Technology; Elementary School Teachers; Elementary Secondary Education; *Evaluation Methods; Naturalistic Observation; Professional Development; *Psychometrics; Reliability; Scoring; Secondary School Teachers; *Teacher Evaluation; Validity; *Videotape Recordings
IDENTIFIERS *Performance Based Evaluation

ABSTRACT

Issues in collecting data on the classroom performance of teachers are discussed, with a focus on written descriptions of teacher behavior in contrast with videotaped observations. The central issue in this comparison is the comparability of the information when it is observed live and when it is simply recorded on videotape. The fundamental psychometric questions associated with any assessment technique can also be posed in terms of validity, reliability, and bias. In any data collection effort, the purposes must be clear, and the inferences to be drawn must be recognized. The situational generalization of observations is limited in different ways for live observers or videodocumentation. Live observers restrict their attention, and videorecorders may restrict the observers' focus. However, the videodocumentation of classroom teaching produces a record that can be viewed, interpreted, and evaluated by many observers. Limited research from J. R. Fredericksen and colleagues suggests that live observation and videotape scoring produce very similar evaluations of teachers, with adequate interscorer reliability. Various practical issues in videotape use, and opportunities it provides for professional development are discussed. Research needs are reviewed. (SLD)

Live vs. Memorex: Psychometric and Practical Issues in the Collection of Data on Teachers' Performances in the Classroom

Richard M. Jaeger
Center for Educational Research and Evaluation
University of North Carolina, Greensboro

Presented at the annual meeting of the American Educational Research Association in a symposium titled *Using Videotapes for Assessing Teacher Instructional Practices: Experiences from the National Board for Professional Teaching Standards Field Trials*, Atlanta, Georgia, April 12-16, 1993.

From a measurement perspective, teacher assessment requires three distinguishable steps: The first is data collection, either through observation of teaching practice in an approximation of a natural setting or through elicitation of teacher performance on tasks that are thought to be precursors of effective teaching. Sending an observer or using a video camera to record activities in a teacher's classroom are examples of the former type of data collection. This kind of observation only approximates a natural setting because it is intrusive and therefore alters the natural flow of classroom activity and interaction in unknown ways. It invokes the Heisenberg Principle, which states that phenomena are inevitably altered by the act of measurement. Using a written examination of a teacher's pedagogical knowledge provides an example of the latter type of data collection.

In the second step, the data collected must be organized into subsets of evidence that provide coherent representations of valued teaching behaviors or knowledge. Many systems of live observation in classrooms have combined the first two steps by using observation scales that require coding of specific incidents of teacher behavior or teacher-student interaction.

Third, the evidence in each subset must be evaluated and aligned with a scale of judged teaching quality, and the resulting judgments must be integrated across subsets. An overall judgment of teaching quality will then result. When live observation of teachers' classroom practice has been used in the past, this step often has been a mechanical task. Observers recorded the incidence or frequency of specified teacher actions or interactions with students, and a formulaic transformation of these data resulted in a score on a purported teaching quality scale (cf., Flanders, N., 1970).

However, live observation of teachers' classrooms could result in a written ethnographic record of observed actions and activities, leaving organization of observations and evaluation of teaching quality to the second and third steps. This procedure would parallel that of video tape recording a teacher's classroom. The distinction would be in the written record, filtered through the eyes of the observer, versus the auditory and visual record, filtered through the preferences and skills of the video camera installer or videographer. In either case, only a selected portion of actual classroom activity will be observed and recorded. And the recorded activity will be modified to some degree by the presence in the classroom of the observer or the video camera.

The Live Versus Video Debate

The principal question surrounding the debate on video recording versus live in-class observation of teaching performance concerns the first step in the measurement process, and its implications for the second and third steps. The question can be posed in a variety of ways, but the central issue is the comparability of information observed and recorded when the observation is taken through a video camera rather than through a live observer. Fredericksen, Sipusic, Gamoran, and Wolfe (1992) characterized the difference between these two modes of observation in terms of "... control of attentional focus, access to conversations, ability to directly assess classroom context, and the ability to review events while adopting different perspectives or for documenting findings." (p. 112). Control of attentional focus is concerned with the impossibility of simultaneously attending to all events and activities that are taking place in a classroom, and the need to select a focus of attention at each moment of the observation period. Access to some conversations taking place in a classroom is limited by the placement of live observers or a video camera in a classroom, and some conversations are therefore

lost to observation. Assessment of classroom context includes the assessment of artifacts that are placed in bulletin boards or walls in the room, and other physical classroom elements that are instructionally relevant. The ability to review events from different perspectives is concerned with the impossibility of simultaneously observing classroom activities from multiple observational perspectives, such as, for example, the level of student engagement, the cognitive complexity of student-teacher interactions, the subject-matter focus and depth of the instruction, etc. These issues are discussed in some detail in a latter section of this paper. It will be shown that each method of observation enjoys some advantages with respect to these issues.

A fundamentally important question arises when teaching performance is to be judged against a fixed criterion (as in the case of determining whether a teacher will receive certification from the National Board for Professional Teaching Standards). Of concern is whether the same decisions will be reached on the basis of videodocumentary evidence as on the basis of live observational evidence. That is, will the same teachers be found worthy of receiving National Board Certification? If the objective of observation is mere description of teaching performance, the fundamental question is similar, but the implications of the answer are, very likely, less immediately consequential: Will the same conclusions be drawn on the basis of videodocumentary evidence as on the basis of live observational evidence?

One can also pose the standard psychometric questions associated with a comparison of any measurement processes: What are the comparative validities of scores derived from live observations versus video documentation of classroom teaching? What are the comparative reliabilities of scores based on the information? How do the bias properties of the scores compare? Answering these questions requires strict specification of the objectives of measurement and the inferences to be drawn from the data collected.

We will examine these questions in the balance of this paper. Unfortunately, we will often be forced to speculate on their answers due to a paucity of compelling research evidence. In the final section of the paper we will describe some experimental research that might inform these questions and thus help to resolve the live observation versus videodocumentation debate, at least as it applies to the evaluation of classroom teaching.

Types of Inferences About Teaching Quality

The purposes of data collection must be made clear in any comparison of modes of data collection. What inferences are to be drawn from the performances observed? Are observations, whether collected live or through a video camera, to be used in inferring the quality of a teacher's typical classroom performance, or in inferring the quality of a teacher's teaching *capability*, either without regard to, or with only weak inference to, typical classroom practice?

Assessment for purposes of licensure or certification of professionals generally focuses on capability, not on typical practice. Licensure is based on a demonstration of knowledge sufficient to avoid practices that are dangerous or damaging to the grouped served. Thus a state-awarded license to teach in the public schools provides no guarantee of effective classroom practice, either in the short or long term. It merely indicates that the licensee has demonstrated a minimally sufficient accumulation of professional knowledge. Dangerous or damaging practice in the classroom cannot therefore be attributed to ignorance. Similarly, professional certification is no guarantor of outstanding professional practice. It too, indicates that the certificate holder has demonstrated sufficient accumulated knowledge and skill to *enable* outstanding practice. Without the knowledge and skill connoted by certification, outstanding practice is impossible or unlikely. With that knowledge,

outstanding practice might occur. However, certification provides no guarantee that it will occur with any regularity, or even at all, in ongoing professional practice.

When comparing the validity of teacher assessment results based on live observation with those based on videodocumentation, one must therefore compare the validity of inferences concerning teachers' capabilities, not inferences concerning their typical practice. This distinction is fundamental to an appropriate comparison of the validity of the two modes of observation.

It might well be the case that teachers' beliefs concerning the comparative validity of live observation and videodocumentation of their teaching, and their often-voiced perception that live observation is essential to accurate assessment of the quality of their teaching, stem from an assumption that observation will be used to gather information concerning their typical teaching performance. It might also be the case that the two observation methods are thought by teachers to be used for different purposes. Personal conversations on this point with a number of experienced teachers (unfortunately, a selected, non-representative sample) suggest that some teachers regard live classroom observation as a method of gathering information about their typical teaching practice and video recording of their classes as a method of gathering information about their teaching capability. Frequent discussions concerning the fairness and importance of giving teachers "prior notice" about the scheduling of classroom observations arise whenever classroom observation is considered. Attention to this issue indicates the widespread misperception that classroom observation is used to collect information on teachers' typical practice. When teachers' capabilities are to be assessed, they should always be informed in advance about the scheduling of classroom observations.

The Nature of the Validity Issue

Any observation of classroom teaching involves sampling from a domain of performance. When observation is conducted for purposes of teacher licensure or certification, the domain of interest is the teacher's performance capability. At a minimum, sampling of the domain takes place along three dimensions: time sampling, situational sampling, and observer sampling.

Were samples of teaching performance to be generalized to a teacher's typical practice, the domain of generalization would be readily definable along the time dimension. The domain would consist of all classes to be taught during the period in which the inference was claimed to be valid — a given school year, a five-year period, etc. The time period might be defined by the renewal period of the license or certificate. However, if as is typical of professional certification, assessment results are regarded as indicators of capability rather than typical performance, the time domain of generalization is less obvious. Two issues become important. First, the time domain cannot be so great that candidates are likely to diminish in capability over the period of generalization. Many professions have recognized this issue by requiring periodic renewal of certification. Examples include the National Board for Professional Teaching Standards and the American Board of Family Practice Physicians. Second, the sampling period and conditions must be such that candidates for certification have the opportunity to exhibit their capabilities. Teachers must be observed on a sufficient number of occasions, and for a sufficient length of time on each occasion, that they are afforded adequate opportunity to exhibit their capabilities. Otherwise, generalization is invalid.

Situational generalization involves an inference from the performance recorded or observed to a teacher's actual performances during the time period sampled. As Fredericksen, et al. (1992) have noted, this type of generalization is limited by several factors regardless of the mode of data collection. Neither live

observers nor a video camera can focus simultaneously on all classroom activity. When live observers pay attention to a specific interaction between a teacher and a student, for example, some portion of of the activity of other students will be lost to them. Since live observers are typically relegated to the back of a classroom in an attempt to reduce their intrusiveness, they will not be able to determine the content of many conversations; their range of hearing within the confines of a noisy classroom will be limited. Although live observers can more readily assess such elements of classroom context as the content of bulletin boards, students' overall attentional stance, and students' entering and leaving behaviors than would be possible using a video camera, their ability to reconsider classroom activities from alternative perspectives is severely limited. It depends on powers of [1]redintigrative memory that are, for most observers, limited and unreliable.

Situational generalization of the results of video recording of teachers in their classrooms (i.e., generalization from what the camera records to what actually took place in the classroom during the recording period) is also limited in predictable ways. Although comparative experiments involving variations in use of the technology have not been reported, several alternatives are obvious. The video record includes a coordinated audio record as well. If a single video camera with an attached microphone is used by an untrained videographer, costs will be minimized but substantial portions of the video and audio record of classroom activity will be lost. The record will be screened through the eye of the videographer, who may fail to anticipate correct placement of the camera. In addition, the audio record may be of such poor quality that it is difficult to comprehend important statements by the

[1]Redintigrative memory is the ability to reconstruct a totally accurate, vividly detailed picture of an event in one's mind, including, for example, exact recall of a conversation and a richly detailed description of the context in which the conversation took place. Many persons have this ability for a few highly significant events in their lives (such as the moment when they learned that President John Kennedy had been killed), but cannot redintigrate typical events in their lives.

teacher and interactions between the teacher and students or between students. Fredericksen, et al. (1992) found that use of two microphones, one worn by the teacher and the other placed in a location that enables students' voices to be heard, produces a clear audio record when controlled by a trained videographer. These researchers also recommend the use of two video cameras, with one trained on the teacher or the chalkboard and the other trained on the class. This form of video recording is far more costly than is use of a single-camera alternative that leaves to the teacher, the task of finding a school-based camera operator (perhaps a student in the school). Experiments involving simultaneous use of these alternatives are needed to determine how much the less-expensive alternative limits valid situational generalization.

Observation of classroom activity by a live observer produces a verbal record that can take on many forms. At one extreme, an observer can complete a closed-option observation instrument that produces a record of the incidence or frequency of pre-specified events during the observation time. Such a record is subject to omission of events from the prespecified incidence or frequency record because the observer failed to note them, omission of important classroom events or activities because the observation instrument failed to provide a place to record them, recording errors due to imperfect translation or interpretation of observed events that are accommodated by the observation instrument, and false recording of the incidence or frequency of events because the observer was ill-trained, biased, or careless. Such errors cannot be detected, much less corrected, if a single observer is used. Through the use of multiple observers of the same classroom activities, the effects of some of these errors can be reduced. However, the larger the number of simultaneous observers in a classroom, the greater the risk of changing the processes and activities that are being observed. The Heisenberg principle states that

the very act of measurement alters that which is being measured. Two observers in a classroom are likely to cause greater change than would a single observer.

Live observers could attempt to produce a written record of what they observe in a classroom, much as an ethnographer attempts to capture the essence of a culture through extensive observation and recording of events as they unfold. Live observers who are allowed to produce written records of their observations with no pre-specified restrictions on what can be observed or recorded have the potential of capturing events and activities that would be lost through use of closed-option recording instruments. However, the advantage of inclusiveness carries with it the risk of greater attentional bias, selective recording in ways that are difficult to detect, and interpretations that might reflect the predispositions of the observer as much as the capabilities of the teacher observed. The risk of selectively focusing on a particularly memorable classroom event, to the exclusion of activities and interactions that are more central to a valid assessment of teaching capability, is also present when observers are allowed to produce a narrative record of their observations. Again, if a single observer is used, these risks are substantial and statistical interaction effects between observers and teachers are virtually undetectable. The magnitude of observer idiosyncrasies can be estimated if multiple observers are assigned to the same classrooms at the same time and observers are required to record, interpret, and score their observations independently. But, as noted earlier, the use of simultaneous, multiple observers risks altering the classroom activity observed.

It is thus clear that situational generalization is limited in different ways through use of live observers or videodocumentation of classroom teaching. Videographers edit the observational record through their imperfect use of the recording device, failing to include some classroom events and activities either through incompetence or idiosyncratic selective attention. However, they do not

alter their reporting of that which the camera sees. Live observers not only restrict their attention to events and activities they regard as worthy of observation, but they edit the observational record through their further selection of features they deem worthy of recording and through their subsequent interpretation of those features. The process of selection and editing by live observers thus incorporates many more opportunities for distortion of the situation observed than does video recording of classroom teaching. Perhaps of greater importance, videodocumentation of classroom teaching produces a record that can be viewed, interpreted, and evaluated by many observers, including the teacher whose classroom is video taped. The importance of this advantage is widely recognized in the literature on professional development of teachers and other professionals (cf., Ajayi-Dopemu & Talabi, 1986; Fuller, & Manning, 1973; Fredericksen, et al., 1992).

Research on the Comparative Results of Live and Video Observation

Research on the comparative psychometric properties of assessments based on live observation and videodocumentation of classroom teaching is surprisingly sparse. A very small experiment involving two observers and four teachers was conducted by Fredericksen, et al., 1992. Although not conclusive, their results are important.

Fredericksen and his colleagues had two observers assess the effectiveness of four high school mathematics teachers on four evaluative dimensions labeled "Pedagogy," "Climate," "Mathematical Thinking," and "Management." Observers' judgments were based on a single class period of observation with each teacher. A videographer was sent to each classroom to record the events and activities taking place during the same period that the live observers were present in the classroom.

Several weeks after she had observed and scored the four teachers' classes, one of the live observers watched and scored the videotapes of the same classes she

had seen through live observation. According to the authors (p. 68), "To the best of her ability, MG watched and scored the video tapes on their own merit without being influenced by her previous scores." An additional assessor who had not been present in the classes also scored the videotapes of the classes.

Both the live observers and the video scorers used a six-point scale with score values labeled "Counterproductive(1)," Missed Opportunities(2)," "Acceptable(3)," "Good(4)," "Masterful(5)," and "World Class(6)." A detailed description of scorer training is provided in the Fredericksen, et al. report. One of the live observers had been trained to score video tapes of teachers' classroom instruction but had not received additional training in classroom observation. The other observer had viewed numerous video tapes of classes and was familiar with the scoring framework used to assess video tapes, but had not been trained as a scorer or a live observer.

Fredericksen, et al. examined several research questions. The first concerned the comparability of the scores awarded the teachers by the two live observers. Of sixteen scores awarded (to four teachers on each of four scoring dimensions), the two observers agreed perfectly on seven, disagreed by one point on seven, and disagreed by two points on two. There was no clear pattern of disagreement in these scores. The mean score awarded by Observer 1 was 3.9 and the mean score awarded by Observer 2 was 4.2. Observer 2 awarded a higher score than did Observer 1 in seven cases; the reverse was true in four cases. The overall correlation between the two sets of scores was 0.49; a reasonable level of agreement given the absence of observer training.

The inter-scorer agreement between the two scorers of video tapes of the four classes was similar to that between the two live observers. The two scorers awarded identical scores five times and differed by one point eleven times. Scorer 2 awarded a higher score than did Scorer 1 seven times and the reverse was true four times.

The overall correlation between the two sets of scores was 0.54, a value that was just slightly larger than the correlation between scores assigned by the live observers. In sum, the inter-scorer agreement of video tape scorers appeared to be quite similar to that between live observers.

The rate of agreement between scores assigned by the single observer who used both live observation and scoring of video tapes was found by Fredericksen and his colleagues to be quite high. Of the sixteen scores awarded, eight were identical across methods and eight differed by one point on the six-point scale. An equal number of the differing scores differed in each direction. The overall correlation between scores based on live observation and video scoring by the single observer who used both methods was 0.74, a value that was probably inflated by the scorer's memory of her earlier scoring of live observations at the time she scored the videotapes.

The limited evidence amassed by Fredericksen and his colleagues suggests that live observation and video tape scoring produce very similar evaluations of teachers, and that inter-scorer agreement is similar for both methods of assessment. However, the generalizability of these conclusions is limited by the sizes of the teacher and observer samples used in this research. Additional research is clearly essential.

Practical Issues

The rate of inter-scorer agreement reported by Fredericksen, et al. suggests that more than one observer will be needed to assess teachers' classroom effectiveness if the observational record is restricted to a single class period for each teacher. The cost of an assessment system that required on-site observation of teachers' classes by two observers would be substantial. For example, consider the following cost assumptions. Suppose that teachers were paid at the rate of $100 per

eight-hour day for the time they spent learning to observe teachers and record their judgments reliably, and for the time they spent completing observations. Suppose further, that observers could complete an average of three observations per day, including travel time, observation time, and scoring time because each observer restricted her/his observations to teachers in a single school system, and travel time between observations was no more than 30 minutes. Finally, suppose that each observer would be a classroom teacher who would require two weeks of part-time training at the rate of two hours per day (a conservative estimate in light of the experience of Fredericksen, et al., 1992), and would assess a total of 15 teachers during a one-week period of observation. A system that required two observers per teacher would then cost about $100 per teacher observed. This $100 figure just reflects the cost of training observers and paying them for completing observations. It does not include reimbursement of travel expenses or the costs associated with development of the observation system or interpreting and scoring observational records.

The use of classroom teachers as live observers would require that the observers be absent from their own classrooms during the regular school day. The preceding cost analysis assumes that each teacher-observer could not miss more than one week of her/his own classes during a school year, and would require re-training in successive years.

If video observation of teachers' classes were to be used instead of live observation, video tapes could be made by a non-teaching videographer or by personnel available within the teacher's school. The effect of using video camera operators selected by teachers rather than trained, professional videographers is unknown. Whether use of local video camera operators jeopardizes teachers' opportunities to be certified as highly accomplished teachers should be examined experimentally.

Video tapes of teachers' classroom instruction need not be scored and evaluated by other classroom teachers during regular school hours. Thus teacher-assessors could be trained and could complete their scoring activities after regular school hours or on weekends, without disruption of their own teaching schedules. The cost of video assessment of teachers' classroom instruction would be lower than that of live observation for several reasons. First, transportation costs and personnel costs would be reduced. At worst, a single videographer would have to be sent to each teacher's classroom instead of multiple observers. Second, since scoring need not take place during regular school hours, trained scorers could be used to assess a larger number of teachers, thus amortizing training expenses over a larger number of assessments.

Finally, if it is found that additional scoring of a teacher's classroom instruction is needed, or if the accuracy of an observational assessment is challenged, the use of video-taped observation has clear and compelling advantages compared to live observation. A more complete record of the original "data" is provided by a video tape than by a written transcript produced by one or two observers. Although the video record is filtered through the selective attention of the videographer, it is not distorted by the evaluative rendering of an observer. If necessary, additional scorers can be shown the videotape, or an impartial board of judges could be assembled to review the basis of a challenged score.

The Professional Development Advantage of Videodocumentation

A signal advantage of the use of video assessment of teachers' classroom instruction rather than live observation is the opportunity teachers have to view and analyze tapes of their own classes. The professional development advantages of video feedback have been documented over several decades by researchers in a number of professional fields, including teacher education (Ajayi-Dopemu & Talabi,

1986; Deasy & Heitzenroder, 1991; Eley & Hess, 1992; Foley, 1977; Fuller, & Manning, 1973; Gardner, Clements, & Rodriguez, 1982; Hays, R. B., 1990; Love, 1978; Krajewski, 1976; Rabozzi, 1977; Rogers, 1987; Salomon & McDonald, 1970; Star, 1977; Star, 1979; Taylor-Way, 1981).

Most recently, Eley and Hess studied video feedback effects on undergraduate teacher education students and concluded (p. 8): "Overall lesson videotaping appears to be a motivating factor for aspiring pre-service teachers in addition to an effective behavioral modifier." They also found that the video observation experience induced facilitative anxiety in students and thus improved their teaching performances.

Deasy and Heitzenroder (1991) investigated the perceived effectiveness and usefulness of self-evaluation via video tape review and external written evaluation by supervisors as perceived by in-service teachers who were participating in specialized training for instruction of behaviorally disordered students. The teachers were found to prefer video-based self-evaluation by a two-to-one margin. Those who preferred video-based self-evaluation noted its contribution to their own awareness of their teaching behaviors, its perceived objectivity, and its provision of "concrete" information that enabled them to identify their deficiencies and remediate them (p. 7).

The implications for the National Board for Professional Teaching Standards of the large body of research on the facilitative effects of video feedback on teacher development are obvious. Since an important objective of the assessment work of the National Board is improvement of the skills and capabilities of the nation's teachers, methods that have been shown consistently to contribute to that objective should surely be used.

Fredericksen and his colleagues (1992) found that their program of video review and discussion among teachers contributed to the achievement of several

valued outcomes. First, teachers learned to analyze the quality of their teaching and to plan modifications that would remediate deficiencies and enhance personal effectiveness. Second, shared review of video tapes of their teaching provided teachers with unprecedented opportunities to engage in professional dialogue on effective teaching methods and strategies for their own development. These self-evaluative activities are the hallmark of a profession.

Experimental Research That Would Inform the Debate

Clearly, there has been too little research on the relative effectiveness and costs of live observation and videodocumentation of classroom teaching. Although the effectiveness research conducted by Fredericksen and his colleagues provides an important beginning, many critical issues have yet to be examined. Most fundamental is the comparative validity of inferences drawn from observations recorded live and those derived from videodocumentation of teachers' classroom performances. The use of known-groups studies involving teachers who have been classified on the basis of external indicators as outstanding teachers, journeymen teachers, and novice teachers would provide information on the comparative "hit" rates of live observation and video-based classifications. In contrast to the Fredericksen study, numbers of teachers and observers sufficient to detect important differential effects with reasonable statistical power must be used.

The cost of videodocumentation of classroom teaching will vary widely, depending on the use of professional videographers outside the school and school system, the use of media specialists who work in schools, or the use of students for video recording. Likewise, single-camera and microphone recording will be less expensive than will multi-camera and multiple microphone setups with appropriate sound mixing equipment. The need for professionals and complex video technology to secure records of teaching that can be assessed reliably and

validly has not been substantiated. Again, comparative research on the effectiveness of alternatives is sorely needed.

When records of teaching performance contribute to important selection decisions, such as selecting teachers for National Board Certification, the comparative decision consistency of scores derived from live observation and video recording must be investigated. Since the assessment packages to be used by the National Board will contain a wealth of evidence beyond that secured through direct observation of classroom teaching, decision consistency comparisons must be made in the context of all additional evidence that will be used to determine the certifiability of National Board candidates.

As noted earlier, an important objective of the National Board's certification program is its contribution to the professional development of the nation's teachers. Beyond the measurement issues that must guide a choice among classroom observation methods, the National Board must obtain information on teachers' perceptions of the comparative utility of written assessments of their classroom performances and video documentation (perhaps coupled with written evaluations) in pinpointing needs for development.

Finally, the comparative costs of live observation and video recording of classroom teaching must be assessed through careful study that varies such cost-intensive factors as observation frequency, number of observers, complexity of the video equipment used, and the professional status of the videographer.

Only with the kinds of solid research needed to inform important psychometric and practical measurement and professional development questions can a reasoned choice among observation procedures be made. The assessment and certification program of the National Board for Professional Teaching Standards provides an ideal opportunity to resolve these critical issues.

References

Ajayi-Dopemu, Y. & Talabi, J. K. (1986)(. The effect of videotape recording on microteaching training techniques for education students. *Journal of Educational Television, 12,* 39-44.

Deasy, G. & Heitzenroder, W. (1991). A comparison of methods for providing evaluative feedback to practicum students on segments of teaching during special education practicum in rural areas. In *Reaching our Potential: Rural Education in the 90's.* Conference Proceedings, Rural Education Symposium, Nashville, TN, March 17-20, 1991.

Eley, R. K. & Hess, C. A. (1992). Effects of videotaping on pre-service teacher performance. Unpublished research report. Department of Secondary Education, the University of Akron, Akron, OH.

Flanders, N. (1970). *Analyzing Teacher Behavior.* Reading, MA: Addison Wesley.

Foley, R. (1977). A departmental approach for improving lecture skills of medical teachers. *Medical Education, 10,* 369-373.

Fredericksen, J. R., Sipusic, M., Gamoran, M. & Wolfe, E. (1992). *Video Portfolio Assessment: A Study for the National Board for Professional Teaching Standards.* Berkeley, CA: Educational Testing Service.

Fuller, F. F. & Manning, B. A. (1973). Self-confrontation reviewed: A conceptualization for video feedback in teacher education. *Review of Educational Research, 43,* 469-528.

Gardner, C. H., Clements, R. O. & Rodriguez, M. (1982, March). Comparing videotapes and written narrative records of second-grade reading classes: selected methods for particular observational goals. Paper presented at the annual meeting of the American Educational Research Association, New York.

Hays, R. B. (1990). Self-evaluation of videotaped consultations. *Teaching and Learning in Medicine,* 2, 232-236.

Krajewski, R. (1976). Clinical supervision: To facilitate teacher self-improvement. *Journal of Research and Development in Education, 9,* 58-66.

Love, J. (1977). Improving college teaching through videotape/feedback. *Journal of Teacher Education,* 4, 105-118.

Rabozzi, M. O. (1977). Microtesting/videotaping experiences in the methods course. *Social Education, 41,* 229-231.

Rogers, S. (1987). If I can see myself, I can change. *Educational Leadership, 45,* 64-67.

Salomon, G. & McDonald, F. J. (1970). Pretest and postest reactions of self-viewing of one's teaching performance on video tape, *Journal of Educational Psychology, 61,* 280-286.

Star, B. (1977). The effects of videotape self-image confrontation on helping perceptions. *Journal of Education for Social Work, 13,* 114-119.

Star, B. (1979). Exploring the boundaries of videotape self-confrontation. *Journal of Education for Social Work, 15,* 87-94.

Taylor-Way, D. G. (1981, April). Adaptation of interpersonal process recall and a theory of educating for the improvement of college instruction. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, CA.