A review of studies on effective principals reveals that a number of methodological problems have prevented the accumulation of findings adequate to producing well developed models of instructional leadership and its effects on school outcomes. Problems of measurement and research design in the existing literature are reviewed and suggestions for future studies made.

These suggestions point to future research meeting three standards: (1) measures of principals' leadership behavior must be better grounded in the behavioral processes found in schools; (2) measures of school effectiveness must be made less unidimensional and more valid and reliable; and (3) research designs connecting leadership and effectiveness should become more sensitive to issues surrounding the demonstration of causality and the potential for interactions.

(Author/GK)
METHODOLOGICAL CONSIDERATIONS IN STUDIES OF EFFECTIVE PRINCIPALS*

Brian Rowan
David C. Dwyer
Steven T. Bossert

Far West Laboratory for Educational Research and Development


* Work on this paper was supported in part by a contract from the National Institute of Education, Department of Education, under Contract No. 400-80-0103. The contents of this paper do not necessarily reflect the views or policies of the Department of Education or the National Institute of Education.
ABSTRACT

A review of studies on effective principals reveals that a number of methodological problems have prevented the accumulation of findings adequate to producing well developed models of instructional leadership and its effects on school outcomes. Problems of measurement and research design in the existing literature are reviewed and suggestions for future studies made.
METHODOLOGICAL CONSIDERATIONS IN STUDIES OF EFFECTIVE PRINCIPALS

Although it is common to assume that instructional leadership by principals is crucial to school effectiveness (see, e.g. Edmonds, 1979; Blumberg and Greenfield, 1980; Lipham, 1981), a careful examination of the literature on principal leadership (for a review, see, Bossert et al., 1982) reveals a number of design and measurement problems that make such a conclusion problematic. In this paper, we discuss these problems and suggest some standards for conducting future research on principal effectiveness.

Our discussion centers around three specific aspects of research on principals. First, measures of leadership used in the research are often disconnected from the practical activities involved in school management, especially instructional management. As a result, little is known about specific leadership behaviors that increase instructional effectiveness. An additional problem is that measures of school effectiveness, especially those purporting to measure instructional effectiveness, are narrow and unreliable. That is, studies frequently use unstable measures of instructional effectiveness and ignore the relationship between instructional and other types of student outcomes. A final problem is that research designs in studies of effective principals have not resolved issues of causal ordering and have neglected the role of contextual factors in shaping the relation between leadership and school effectiveness. As a consequence, current research is divorced from contingency theories of organizational effectiveness and presents an extremely optimistic view of leadership effects.

Despite these shortcomings, the idea that principal leadership is
related to school effectiveness cannot be dismissed. Certainly the consistency of findings across different research traditions suggests that the correlation between principal leadership and school effectiveness should be investigated further. At the same time, the methodological problems that plague existing research need to be corrected in future studies.

In order to encourage more work on principal effectiveness and to insure that this work does not repeat the mistakes of previous studies, we propose three standards for conducting future research in this area. In our view, the standards proposed here will insure that information useful both to the academic research community and to local education agencies will be generated. The standards, briefly stated, are as follows:

Standard 1: Descriptions of principals' leadership behavior should refer to concrete, school-based activities principals actually engage in.

Standard 2: Measures of school effectiveness should be valid and reliable and reflect the diversity of goals under which schools operate.

Standard 3: Research connecting leadership with effectiveness should be both longitudinal (to disentangle issues of causal ordering) and comparative (to uncover potential interactions between context and leadership).

BACKGROUND

The critique and suggestions offered here are based on our reading of the literature on effective principals, a literature we have reviewed in depth elsewhere (Bossert et al., 1982). For the purposes of this paper, we merely note that our critique is of three distinct research traditions. The first can be called the leadership behavior tradition (see, e.g., Hemphill et al., 1962; Gross and Herriott, 1965; and Stogdill,
1974: 130-133), a literature that has attempted to quantify the abstract notion of leadership by developing empirical indices of leader behavior. A second approach, which we call the effective schools tradition, has attempted to identify particularly effective schools and to compare these with ineffective schools across a number of dimensions, including principal leadership (see, e.g., Weber, 1971; Michigan State Department of Education, 1974; New York State, 1974; Brookover and Lezotte, 1977; California State Department of Education, 1977; Wellisch et al., 1978; Brookover et al., 1979; Brundage, 1980; Lihman, 1981). A third approach, the fieldwork tradition, has devoted less attention to issues of principal effectiveness than the previous two (an exception is Blumberg and Greenfield, 1981), but one branch has used Mintzberg's (1973) method of tracking administrators' time allocations to make inferences about school-level instructional leadership (see, e.g., Goodman, 1976; O'Dempsey, 1976; Peterson, 1978; Martin and Willower, 1981).

In the remaining sections of this paper, we illustrate how various studies in these three traditions have attempted to deal with the problems of methodology discussed above, and we suggest means for improving research on effective principals.

THE MEASUREMENT OF PRINCIPAL LEADERSHIP

Principal leadership, especially instructional leadership, has been measured in a variety of ways in previous research. Although some scholars have criticized these measures from the point of view of conventional measurement theory (Charters, 1964; Erickson, 1964), recent research indicates that leadership measures attain acceptable levels of validity and reliability when used with samples of principals (e.g., Lucietto,
Despite evidence of validity and reliability, there are reasons to be cautious about the practical implications of results based on leadership measures. As Erickson (1977:455-467) noted, many of the measures used in previous studies do not ask for reports of behavior, but rather for summary reports about respondents' evaluations of behavior. To the extent that these evaluations fail to describe the actual activities one must engage in to increase school effectiveness, they provide practicing administrators with few concrete guidelines for school improvement.

Consider, for example, studies that have employed reputational measures of leadership (e.g., Blumberg and Greenfield, 1980). Clearly, reputational measures suffer from problems of validity and reliability, as the large sociological literature on community leadership demonstrates (Wolfinger, 1960). More importantly, however, the reputational approach provides practicing administrators with few insights into how to become more effective within their own schools. Even if it is granted that a reputation for leadership is central to actual effective leadership, little is said in studies employing reputational measures about how to establish such a reputation.

More common than reputational measures are those derived from "in-basket" studies (e.g., Hemphill et al., 1962). These measures have in fact gained widespread use in assessment centers (Hersey, n.d.). Although in-basket measures have reasonable reliability and validity (Earles and Winn, 1977), a major objection to their use in studies of principal effectiveness is that the measures are not based on procedures reflecting the routine demands and activities of principals. The in-basket procedure...
measures performance solely on written work, while numerous field studies have shown that principals engage in very little written work (Wolcott, 1973; Martin and Willower, 1981). Thus, as Erickson (1964) pointed out, the in-basket technique may not accurately reflect principals' effectiveness in carrying out activities that are more typically associated with the principal's management role.

Moreover, in-basket measures, like reputational measures, represent inadequate guidelines for administrative behavior. The categories of leadership behavior generated by the in-basket procedure tend to be highly abstract, and connections between the factors as defined and the daily world of administrative behavior are ambiguous. For example, Hemphill et al. (1962) rated behaviors on such factors as "maintaining relationships" and "analyzing the situation." To the principal attempting to improve his or her performance, the empirical referents for such complex factors in the context of day-to-day school management remain unclear.

More closely grounded in the day-to-day activities of school managers are behavioral indices of leadership based on questionnaires (e.g., Gross and Herriott, 1965; Halpin, 1966; California Department of Education, 1975). To the extent that answers to questionnaires describe routine administrative behavior by leaders rather than affective responses by other school personnel, the former measures appear superior to in-basket measures of leadership. For example, questionnaire studies show that routine behaviors such as making clear standards and rules, criticizing work, acting with the consultation of the staff, offering constructive suggestions to help teachers with problems, or making teachers' meetings valuable increase instructional effectiveness.
Nevertheless, some practical issues remain unresolved, even when these more explicit indices of leadership are employed. For example, although the measures tell us that offering constructive criticism increases school effectiveness or that acting with the consultation of the staff is important, the operationalization of questionnaire items by using some form of Likert scaling tells practitioners very little about exactly what sorts of criticism are constructive or how much consultation is effective. For example, a principal may know that “very much” consultation is better than “very little,” but he or she will still face considerable ambiguity about how much consultation is “very much” and how much is “very little.”

One way around the problem of descriptive ambiguity is to employ qualitative methodologies that generate thick descriptions of principals’ leadership behaviors. Such descriptions would at least indicate some concrete ways principals can make constructive criticisms or increase the decision participation of their staffs. A problem with qualitative research on principals, however, has been that thick descriptions of concrete instructional management activities have generally been edited into “thin” summaries or transformed into frequency counts of acts that are then classified under some larger category such as “instructional leadership” (Martin and Willower, 1981; Morris et al., 1981). A result is that school administrators have been deprived of valid and concrete descriptions of successful instructional management across different school contexts.

Towards a resolution. We suggest that a first step in improving measurements of instructional leadership involves the assembly of thick descriptions of concrete behaviors used by principals to manage instruction.
Moreover, we propose that these descriptions be generated, in part, by using existing questionnaire items as starting points. While a number of correlates with school effectiveness have already been located in questionnaire studies, the specific behavioral referents of these questionnaire items need further probing. Finally, we urge a more complete reporting of qualitative descriptions of instructional management behaviors. The goal of such a strategy would be to develop more concrete, phenomenologically valid indicators of behaviors actually used by principals to manage instructional processes. Such research would more closely resemble the types of information principals currently use (discussions with peers) and perhaps end the underutilization of research by principals (Fillos et al., 1980). It would also offer principals a more concrete referent for judging their own leadership behavior.

THE MEASUREMENT OF SCHOOL EFFECTIVENESS

While measures of leadership have tended to meet the standards of traditional measurement theory, measures of school effectiveness have not. Problems of both validity and reliability plague attempts to assess the instructional effectiveness of schools. These problems not only raise doubts about the empirical relationship between leadership and effectiveness, but also have practical implications for the assessment of schools.

At the outset it is easy to dismiss a number of measures of effectiveness that have been employed in past research, especially those that employ the perceptions of school personnel about effectiveness (e.g., Hemphill et al., 1962; Gross and Herriott, 1968). Instead, we concern ourselves with indicators that are derived from student achievement scores.
The use of achievement scores as the sole criteria for judging school effectiveness is common. For example, virtually all of the studies in the effective schools tradition employ this unidimensional criterion. Yet as Steers (1975) pointed out in his general discussion of measures of organizational effectiveness, most theorists and participants in organizations view effectiveness as a multidimensional construct.

By viewing school effectiveness as a unidimensional phenomenon, current research neglects a number of interesting and important issues. For example, numerous constituencies view the purpose of schooling as broader than simple academic training. Citizenship training, development of self esteem, independence training, and the development of self-discipline exist as important alternative goals. By focusing exclusively on academic achievement, much of the literature on school effectiveness has ignored the relationship between achieving effectiveness in academic outcomes and achieving effectiveness along these other dimensions. We urge more attention to the relationship between these various criteria, a process that would require the development of a multidimensional view of school effectiveness.

Despite our call for a multidimensional approach to measuring school effectiveness, the issue of instructional effectiveness remains important in its own right. Virtually all attempts at measuring instructional effectiveness at the school level recognize the high correlation between the SES composition of a school and its aggregated achievement score. As a result, it is common to use a particular adaptation of regression analysis first suggested by Dyer (1972) to eliminate SES composition from instructional effectiveness measures. This measure
regresses school-level achievement on school SES variables to arrive at a level of achievement that would be expected for any school given its student body composition. This predicted score is then compared to the actual score to obtain a measure of effectiveness. Those schools with actual scores falling far above predicted levels are judged as effective while those falling far below are judged as ineffective.

A number of studies have uncovered severe problems with the reliability of Dyer's (1972) technique. For example, Forsythe (1973) found that schools did not consistently appear as effective or ineffective when judged by Dyer's measure at two points in time. In fact, the correlations of residuals measured one year apart varied from .11 to .50 depending on the cohort of students being measured, with the average correlation being .28. Forsythe concluded that the measure's stability was unsatisfactory and, thus, that it lacked reliability.

The unreliability of Dyer's (1972) measure has important consequences both for research on effective principals and for attempts to judge the effectiveness of local schools. For example, since the measure is known to be unreliable, its use in evaluating local schools is obviously unfair. Yet numerous school systems have begun using the measure. Moreover, we are acquainted with at least one state-level assessment program that aggregates its data in a way that apparently maximizes the unreliability of the measure (cf. Dyer et al., 1969; Law, 1977).

Other problems with the measure call into question the assumption that instructional leadership by principals is related to school effectiveness. For example, one study (Law, 1977) found that measures of effectiveness based on Dyer's (1972) procedure tended to be correlated to school size, with smaller schools tending to have a greater chance of
being outliers than larger schools. It should be noted that this correlation arises as a statistical artifact (errors in school prediction equations show heteroskedasticity) and does not reflect an inherent tendency for small schools to be effective. Because of this problem, it is possible that studies of effective schools have labeled a form of administration quite suited to smaller organizations (high degrees of staff decision participation, much personal contact between teachers and principal) as effective merely because small schools are more likely than large ones to be scored as outliers using Dyer's (1972) criterion.

Towards a resolution. Clearly, when school effectiveness is judged on the unidimensional criterion of instructional performance, measures with higher degrees of reliability are needed. One obvious way to avoid the instability and error inherent in Dyer's (1972) approach is to employ a measure that requires schools to make consistent gains from year to year in aggregate achievement (see, e.g., Wellisch et al., 1978; Gregory and Herlihy, 1980). When using this measure, however, it is important to guard against the possibility that gains arise from changes in student body composition. Moreover, measures of instructional effectiveness should be analyzed in conjunction with indicators of non-academic outcomes. A multidimensional view of school effectiveness, in our view, would strengthen the literature on effective schooling.

PROBLEMS IN RESEARCH DESIGN

More serious than the measurement problems discussed above are two problems in research design: (1) the cross-sectional nature of research on effective principals, and (2) the failure of research studies to address the possibility of interaction effects in the data.
Cross-sectional research. The cross-sectional nature of research on effective principals is particularly disturbing. Given the correlational nature of research connecting principal leadership with school effectiveness, a number of interpretations of the data are equally plausible. At least some of these alternative interpretations call into question the prevailing view that principal leadership precedes school effectiveness by arguing that attributions about leadership follow from evaluations of organizational effectiveness (e.g., Bossert, 1981). For example, in effective schools research, it is quite possible that effectiveness is attributed to principals precisely because they head effective organizations. A similar criticism can be made of the findings that principals in effective schools work more frequently with teachers and have stronger evaluation systems. A plausible explanation of these findings is that effective schools have effective teachers, and that effective teachers allow principals to more frequently observe their teaching and evaluate their outputs.

In fact, some quantitative studies in the literature (e.g., California Department of Education, 1975; Wellisch et al., 1978) have arrayed data in such a way as to test these latter interpretations rather than the argument that leadership "causes" effectiveness. For example, a common data reporting technique in the effective schools literature is the display of tables that use t-tests to compare leadership behaviors in effective versus ineffective schools. Although the intention in these studies is to test the idea that leadership brings about effectiveness, the tables are the equivalent of regression analyses where the independent variable is effectiveness and the dependent variable is principal effectiveness. Thus, prevailing data analyses actually support the
argument that instructional performance affects leadership, while theories about the "effects" of leadership on organizational effectiveness remain largely untested.

Towards a resolution. It is precisely this flaw that makes the prevailing interpretation of effective principals research problematic and the design of in-service training packages premature. Before training programs based on the effective schools literature become widespread, it seems reasonable to demand longitudinal evidence on the relationship between leadership and effectiveness. Qualitative studies, for example, might search for naturally occurring experiments, perhaps by finding principals who are assuming jobs in different schools and by carefully observing their leadership activities and their effects through time. Quantitative studies should begin to use some form of the "cross-lagged correlation" design or at least estimate parameters in cross-sectional designs by using estimation techniques that test the plausibility of alternative causal orderings (e.g. 2SLS or LISREL).

Interaction effects. In addition to the problems of interpretation raised by the cross-sectional research on effective principals, a second shortcoming can be found in the research designs of prevailing studies. Little effort has been made to investigate the possibility that different styles of leadership lead to effectiveness in different settings. This is unfortunate since there are both theoretical and empirical reasons for believing that this may be the case. For example, at least two studies (Brookover et al., 1973; California Department of Education, 1975) have found interaction effects suggesting that the relationship between leadership behavior and effectiveness is contingent upon school context. Moreover, such interaction effects are highly consistent with
current contingency theories of organizational effectiveness (e.g., Fiedler, 1967).

In light of both current theories and empirical findings, the failure of effective principal studies to consider potential contingent relationships is surprising. For example, many leadership training programs of the 1970's endorsed a prominent contingency theory of leadership dealing with staff maturity (Sergiovanni, 1979). Given the popularity of this theory in educational staff development, it is surprising to find that studies of principal leadership have not included measures of staff maturity in their designs. Such oversights demonstrate the need for a closer connection between prevailing theories of leadership and research designs employed in studies of effective principals.

Another prominent omission in the literature concerns differences between urban and other types of schools. Studies by Brookover et al. (1973) and the California Department of Education (1975) show consistent interactions between variables measuring principal leadership and the SES composition of schools or their urban location. Since numerous effective schools studies have been done in urban schools with high proportions of low SES students (for a review, see Clark et al., 1979), the generality of the model of effective leadership found in this tradition can be questioned. Will the model prove as valid in suburban schools or those with lower proportions of low-SES students? The interaction effects found in previous studies suggest that no one best style of leadership exists and that leadership styles successful in one setting may not be effective in others.

Towards a resolution. The above discussion suggests that future research on principal effectiveness can be more sensitive to school
context. This involves a more explicit linking of current contingency theories of leadership to prevailing research designs. For quantitative studies, we recommend a search for interaction effects. For qualitative studies, we urge a more careful reporting of the context of the study or designs that explicitly compare leadership strategies in different (and theoretically relevant) settings. Only then can researchers come to understand leadership contingencies and use this understanding to develop context-specific leadership training programs.

CONCLUSION

We conclude by reiterating our suggestions about how to improve research on effective principals and how to make it more useful to practitioners. We suggest that future research meet at least three standards: (1) measures of principals' leadership behavior must be better grounded in the behavioral processes found in schools; (2) measures of school effectiveness must be made less unidimensional and more valid and reliable; and (3) research designs connecting leadership and effectiveness should become more sensitive to issues surrounding the demonstration of causality and the potential for interactions.

While current research on effective principals has located an important relationship between leadership and school effectiveness, future research meeting the standards suggested here could substantially improve both the practice of school leadership and its theoretical underpinnings.
REFERENCES


Edmonds, R. Some schools work and more can. Social Policy, 1979.


Sergiovanni, T. J. Is leadership the next great training robbery? Educational Leadership, 1979, 36, 388-394.


