

Monitoring School Performance: A Guide for Educators

by

J. Douglas Willms

Centre for Policy Studies in Education
University of British Columbia
and
Centre for Educational Sociology
University of Edinburgh

P. 1-11



The Falmer Press

(A member of the Taylor & Francis Group)

Washington, D.C. • London

Introduction

Educators and administrators have dramatically increased their efforts to collect data describing the performance of their educational systems. Many countries are establishing programs to collect 'indicators' of school quality for monitoring performance at national, regional, and local levels. The quest for more and better data is widespread. Nearly every country in Europe is developing a monitoring system based on performance indicators. The UK Government established a national curriculum of core and foundation subjects, and mounted an ambitious testing program to assess pupils' attainment of the curricular objectives. The US Department of Education, through its Center for Education Statistics, collects a variety of indicators describing the 'health' of the elementary and secondary schooling systems. From these data, it publishes the Secretary of Education's 'wall chart', which includes state-by-state comparisons for a number of performance indicators (Smith, 1988). Currently the National Governors' Association Panel to Monitor the National Education Goals is developing a 'national report card' for monitoring progress towards the national educational goals recently established by President Bush and the Governors (Lewis, 1991; Piphon, 1991). Most states have established monitoring systems based on performance indicators (Selden, 1988), and many school districts are following their lead.

An 'indicator' is simply a statistic describing some feature of the schooling system associated with its performance, such as the average test score of a school, the proportion of drop-outs, or the pupil-teacher ratio. Like most statistics, an indicator derives its meaning from its trend over time, from its variation within a sample, or from comparison to some standard. The standard can be the average value for a set of schools, a predefined goal, or a socially determined standard (Oakes, 1986; Shavelson, McDonnell, Oakes, Carey and Picus, 1987).

The collection of performance indicators is not a new phenomenon. There is a long history of using national data to monitor long-term trends in educational performance; to examine inequalities in attainment between racial, ethnic, and social-class groups; to make inter-regional comparisons; and to assess the impact of major educational reforms (Chartered Institute of Public Finance Accounts, 1986; Koretz, 1986; McPherson and Willms, 1987; Murnane, 1987; Powell and Steelman, 1984; Stern, 1986; Welch, 1987).

Schools and school districts¹ have collected data for planning and decision-making, for assessing strengths and weaknesses in various curricular areas, and for selecting and allocating pupils to different types of programs.

What distinguishes the current work on indicators from earlier evaluation efforts is the amount and kind of data that are being collected, and the way data collection is being institutionalized at all levels. Previously, programs to monitor performance included easy-to-measure indicators, such as graduation rates, pupil-teacher ratios, and achievement test scores. Now administrators are acquiring information on a much wider range of outcomes, including not only cognitive measures, but also affective measures such as self-concept and attitudes to school and work. Many programs include measures describing pupils' family background, and various indicators of school 'processes' believed to be related to schooling outcomes. Before the recent interest in performance indicators, data of this kind were collected only in evaluations or research studies addressing particular questions. Many school districts now collect and analyze these data as part of an institutionalized routine. The routine includes the production of annual reports to school boards, parents, teachers, and administrators.

The Impetus for Monitoring Systems

School boards, administrators, and teachers rely heavily on their working knowledge to make decisions (Sproull and Zubrow, 1981; Williams and Bank, 1984). The 'working knowledge' of decision-makers includes the large body of facts, principles, and perceptions that determine their particular view of the world. It is influenced by their values, attitudes, and beliefs, and by the steady stream of information gathered from friends, colleagues, clients, and the public. But decision-makers often view their working knowledge as inadequate for confronting certain issues: it is too subjective or shallow, or biased because it does not represent the opinions of a wide constituency. Thus they regularly seek 'specific knowledge'. They appoint committees, hire consultants, conduct evaluations, commission research, and attend courses to obtain information focused on specific problems or activities. Although specific knowledge is often more objective and relevant, obtaining it requires time and resources. Also, specific knowledge is sometimes at odds with working knowledge, and does not always fulfil decision-makers' needs (Husen and Kogen, 1984).

Monitoring information can fill some of the gaps between working and specific knowledge. If data are collected routinely from a variety of sources, then many regular information needs can be met more quickly. Monitoring data on some topics can be more objective than those obtained through in-

¹ The administrative units for governing schools at the regional level are generally called local education authorities (LEAs) in England and Wales, education authorities (EAs) in Scotland, and school districts in North America. For sake of brevity I use the term 'school districts' to cover all of these, unless I am referring specifically to administrative units in the UK.

formal observations and conversations. Monitoring data tend to cover a wider variety of topics than those obtained through evaluations, consultancies, or commissioned research. In some cases monitoring information is insufficient for addressing particular problems, but it can provide a basis for the kind of specific knowledge required.

The collection of monitoring data also can serve several functions directly pertinent to improving schooling and reducing inequities. The information can be used to identify problem areas in the schooling system, so that corrective action can be taken. It can assist administrators in determining the best allocation of resources. It can be used to diagnose strengths and weaknesses in pupils' mastery of curricular objectives, and therefore guide curriculum and instruction. It can be used to assess the effects of interventions implemented at the state, district, or school level. It can stimulate discussion about the goals of schooling, and give rise to new ideas that affect policy and practice.

Monitoring data can also motivate administrators and teachers to improve performance and reduce inequities. This function is not necessarily achieved through top-down accountability tactics. In many cases the provision of information itself stimulates the self-regulatory mechanisms already existing in most schools. Research on school and teacher effectiveness suggests that effective schools generally have some system for monitoring performance (e.g., Cohen, 1981; Odden, 1982; Purkey and Smith, 1983), and that effective teachers frequently test their pupils and conduct weekly and monthly reviews (Rosenshine and Stevens, 1986).

However, the movement to collect performance indicators is driven not just by a desire for better information. Many educators believe that the explicit intention of monitoring systems is to make schools accountable through market forces. Throughout Europe and North America a faith in market mechanisms pervades government offices. One of its doctrines is that publicly funded organizations should be held accountable by having to report regularly on their performance. In some public services, such as transportation, postal services, media services, and public utilities, performance is equated with profits or losses. Balance sheets of the public services are compared to those of private companies offering comparable services. Public services that are unprofitable become candidates for closure or privatization. But in areas such as education, health, social welfare, corrections, and the environment, the outcomes are less tangible. Value for money is more difficult, perhaps impossible, to ascertain. In these areas the push has been towards measuring performance on a number of criteria, and comparing results across organizations. The belief is that such comparisons will stimulate competition and motivate organizations to provide higher levels of service.

Some administrators and educators believe that the introduction of market mechanisms to education will significantly improve schooling. The view is that inter-school or inter-regional comparisons will bring pressures to bear on schools, particularly those performing below the average. These pressures will induce schools to perform better, and if not, the data will constitute objective grounds for closing schools or appointing new staff. Also, if indicators can be used to accurately and fairly assess the performance of individual teachers or principals, then they might be used as an objective basis on which to decide promotions, merit pay, and dismissals. Indicators of school

performance have been used in school award programs in California, Florida, and South Carolina (Mandeville and Anderson, 1987; Wynne, 1984). A few states have proposed that teachers be awarded cash bonuses for superior results based on indicator data.

Educators do not unanimously accept the view that market mechanisms will improve the education system. Landsberger, Carlson, and Campbell (1988) surveyed approximately 6000 administrators in England and Wales, West Germany, and the US to determine the most important policy issues facing educational administrators. The primary concern of these administrators was whether 'market mechanisms' should be built into the educational system. Opponents to monitoring argue that there is not consensus about the goals of education, the characteristics of an effective school or teacher, or the nature of educational achievement. They believe that monitoring restricts what is taught in schools, displaces valuable teaching time, and reduces the autonomy of teachers.

Even if administrators do not intend to use monitoring data explicitly for purposes of accountability, the collection of data by itself unleashes subtle and indirect market forces. For example, the results of inter-school comparisons might be used only to supplement other evaluative data, and to support schools in their self-evaluations. But schools directly or indirectly compete for the best teachers and pupils, and monitoring results affect schools' reputations, which eventually influence teachers' decisions about where to teach, and parents' decisions about the best area in which to live. These market forces are supported by policies that promote greater choice and diversity in schooling, such as open enrolment plans that allow parents to choose schools outside their designated catchment areas. Parents sometimes use monitoring results to exert pressures through their locally elected school boards. The pressure can be considerable in areas with declining enrolments, where some schools are threatened with closure.

Some of the impetus for monitoring has come from the fear that monitoring data collected at national or state levels are inadequate or will be used inappropriately. Smith (1988) suggested that the widespread interest in performance monitoring in the US stemmed from the discontent of analysts and policy-makers with existing national data, and a determination on the part of the federal government to use performance indicators for purposes of accountability. The government's intention to continue publishing its 'wall chart' of state-by-state comparisons, and the criticism the document received from some of the states, induced the Council of Chief State School Officers (CCSSO) to develop a more comprehensive and valid system for monitoring performance at the state level. In turn, some school districts have created monitoring systems as a means of protection against criticisms that might stem from state assessments on a narrow set of schooling outcomes.

Finally, the collection of indicator data is consistent with a more general trend amongst governments and other administrative bodies to amass extensive amounts of data and to compile descriptive statistics. This trend has been supported by rapid advances in the technology for collecting, storing, and analyzing data. These activities serve not only an administrative function, but also a political one. Statistics describing the health of education systems can be used to demonstrate the need for reform arising from the poor management

of a previous administration, or to demonstrate improvements stemming from reforms of the administration in power. Some critics contend that analysts choose to report, depending on their political purposes, statistics describing absolute levels of performance, changes in levels of performance, levels or changes for a particular subsample of the population, or comparisons with other districts, states, or countries. Porter (1988) argues that performance indicators are merely a political tool designed to strengthen the hand of those favouring centralized control of the process and products of teaching.

Purpose of the Book

This book is intended to guide the many decisions entailed in developing a monitoring system. Its purpose is to specify the kind of data that might be routinely collected by a school district or by individual schools, and to describe appropriate methods for analyzing and reporting data. No single design for a monitoring system could be appropriate across all districts or schools, and a district-level design would not necessarily serve the requirements for individual schools. The guide begins therefore with a more general discussion of the main issues pertaining to performance monitoring, and sets forth some general principles concerning the design of a monitoring system.

The guide does not describe qualitative approaches to educational evaluation, such as those proposed by Eisner (1985), Fetterman (1988), Hammerley and Atkinson (1983), Lincoln and Guba (1985), and Patton (1980). This decision was not intended to disparage these methods; it was made simply to limit the scope of the book. The multilevel models and methods proposed in this book provide a framework for describing the variability in schooling outcomes between and within schools. This framework is potentially useful for guiding qualitative study in that it invites one to think about how policies and practices at different levels of the system affect schooling outcomes, and whether their effect varies for pupils with differing backgrounds. The framework also serves to contextualize the findings of qualitative studies (e.g., see Raffe and Willms, 1989).

The development of a system for monitoring school or district performance is not an easy task. If a monitoring system is to be useful for planning and decision-making, it must be based on a sound theory about how schools achieve their effects, and it must have a clearly defined purpose which is understood by educators and policy-makers at all levels of the schooling system. It must cover a wide range of educational goals, and be responsive to changes in the educational system. Yet it cannot be too costly in terms of pupil, teacher, or administrative resources.

Several technical issues concerning the measurement of schooling outcomes and the assessment of school effects also must be addressed. Perhaps the most difficult issue concerns the identification of goals that are common across schools in a system. This is complicated by the fact that schools often have different goals for pupils with differing interests and levels of ability. In addition, test developers find it difficult to construct measures that span the entire range of achievement for pupils at a particular grade level, particularly in the later years of schooling, and yet cover the curriculum in enough detail to be useful for diagnostic purposes.

Even with adequate solutions to the measurement problems, the task of separating the effects of school practices and policies from factors that lie outside the school is complicated and requires extensive data. Research on school effectiveness at the Centre for Educational Sociology (University of Edinburgh) showed that pupil attainment at the secondary level is related to pupils' socioeconomic status (SES), their prior level of ability, the composition of their family, the type of neighbourhood they live in, the overall ability and socioeconomic composition of their school, and the level and structure of local employment (Garner and Raudenbush, 1991, McPherson and Willms, 1986; Raffe and Willms, 1989; Willms, 1986). The multilevel modeling techniques discussed later in this book allow one to make statistical adjustments to the school means on an outcome measure to take account of factors that lie outside the school. Estimates of the adjusted means provide a better basis for making comparisons between schools; however, the accuracy of the estimates depends on the amount and type of data available. The accuracy depends also on the assumptions made about the relationships between outcome measures, policy and practice variables, and measures describing outside influences.

All of these theoretical, administrative, and technical issues are inextricably tied to political issues concerning the professional autonomy of teachers, the nature of the curriculum, and the control of resources. The development of a monitoring system requires hundreds of little decisions about what kind of data should be collected, how data will be analyzed, who will have access to data, and what actions will be taken on the basis of findings. These decisions are affected by how those developing the system view the purposes of monitoring, and the amount of resources that can be devoted to the enterprise. If a guide is to establish a standard for performance monitoring, it must either attempt to take the competing interests of several groups into account, or set forth its own biases.

Initial Premises

One could feasibly write an entire volume discussing the political issues concerning performance monitoring. This is not my purpose. However, I will not skirt the central issue of whether such a book should be written at all. Those who decry monitoring may view the book as an attempt to accelerate it. They oppose its acceleration because they believe monitoring may curb educators from questioning the purposes of schooling, and from critically examining what they teach and how they teach it. They would argue also that monitoring systems help institutionalize organizational structures and practices aimed at goals that have not been justified or accepted by the educational community. The opponents of monitoring include hundreds of teachers and administrators who feel that monitoring places unrealistic demands on their time and resources, and that ultimately it will reduce their authority and be used against them.

Advocates of monitoring would counter that both administrators and teachers need objective information to make sound educational decisions. They would argue that market mechanisms have a positive effect on schooling, or at least that monitoring motivates teachers and administrators. They

would point to Gallup Poll results for the US that suggest over 80 per cent of parents with children in school want standardized national testing programs to measure the academic achievement of pupils attending public schools in their community (Elam and Gallup, 1989). The advocates of monitoring might concede that data derived from standardized tests and questionnaires have limitations, but would maintain that such data are better than no data at all.

The debate in the UK has taken a different turn. The Government strived to involve teachers in setting standards and constructing tests. The standardization of the tests has been based on teachers' judgements of what a child should be able to accomplish at various stages, rather than on statistical criteria pertaining to how well test items discriminate amongst pupils. The tests also incorporate several different types of assessment tasks, including performance-based tasks that require pupils to use higher-order thinking skills. Thus, in many respects, the national testing program increases the professional autonomy of teachers. Despite these efforts, however, the pilot testing of the national tests has met with widespread resistance. Many teachers are unhappy because the Government has not made it clear how the results will be used. In particular, it has not specified whether school comparisons will be made, and if so, on what basis. A number of parents are sceptical too. They are unsure whether they want their children tested because they fear the results may be used to make decisions about the type of school program suitable for their children. More generally, there is mistrust of the Government's political agenda connected with national tests and monitoring.

My own position on monitoring is that the benefits *can* outweigh any negative consequences. I believe also that some of the dangers of monitoring can be avoided, or at least minimized. This position has several antecedents. First, it is derived from my interest in studying the causes and consequences of inequities in schooling. Through my research on public and private schools in the US (Willms, 1985a), and my research with McPherson on the effects of comprehensive schooling in Scotland (McPherson and Willms, 1987), I saw the potential for using systematically collected data for examining questions about equity between racial, ethnic, and social-class groups, and between the sexes. I also learned about some of the limitations of monitoring data. If educators are to argue for equality of outcomes, they must be prepared to specify and defend some set of outcomes. My position on monitoring stems as well from having witnessed inappropriate uses of performance indicators. In many cases comparisons between schools are made without making statistical adjustments for the types of pupils entering schools. Thus, the findings frequently suggest that the best-performing schools are those with the most favourable pupil intakes. These conclusions are often unwarranted. The third antecedent to my position on monitoring is a genuine curiosity about how schools work. As an educational researcher I see great potential in using monitoring data to further our understanding about the relationships between schooling processes and pupil outcomes.

Monitoring systems have acquired some momentum; to a large extent I take them as a *fait accompli*. I hope that this guide will help administrators and teachers use indicator data more fairly, and remind them of the limitations of performance indicators. I believe that monitoring can further our

understanding of the effects of educational policies and practices on schooling outcomes, and can help determine whether particular educational interventions have a desirable effect.

However, my endorsement of monitoring is not without qualifications. I also make some assumptions about what could be, even though its realization in practice is difficult. To begin, therefore, I am setting out a list of premises as a means to clarify my position. They are presented in the spirit of Cronbach *et al.*'s ninety-five theses of evaluation (Cronbach *et al.*, 1980); that is, I hope they will provoke further discussion of the issues:

- Monitoring systems can contribute to the working knowledge of both teachers and administrators. They can serve a number of functions relevant to improving schooling and reducing inequities.
- Monitoring data are not a substitute for other kinds of data. Monitoring data should be used in conjunction with data collected both from discussions with staff and pupils, and through detailed observations of school and teacher practice.
- Monitoring systems can induce debate about school policies and practices. Their usefulness in raising questions can be as great as it is in answering them.
- Monitoring systems will not turn the social control and management of schools into a technology. The fear that this will occur presumes that administrators are ignorant of the complexities of schooling and the limitations of monitoring data, and that teachers and pupils are fully submissive to administrative authority.
- One of the dangers of monitoring is that it can restrict the goals of education to a set of objectives defined centrally rather than locally. This can be alleviated by devolving the design and control of most aspects of monitoring to the level at which decisions about policy and instruction are made.
- Administrators must make decisions concerning accountability, such as decisions about school closures, the dismissal of teachers, or staff promotion. If monitoring data are available, they will inevitably be used, either formally or informally, in these decisions. Some of the anxiety concerning accountability can be lessened if administrators use monitoring data to provide constructive feedback about policy and practice. Those monitoring school performance need to specify clearly how monitoring data will be used in summative decisions — what analyses will be undertaken, how findings will be reported, and who will have access to the findings.

In writing this book I continually encountered two difficulties. I wanted to describe what I consider to be *best practice* of monitoring school performance, rather than describe the benefits of monitoring or belabour its pitfalls as it is currently practiced. The problem is that the degree to which best practice can be accomplished is always circumscribed by political considerations, and these tended to temper my view of the ideal. Thus, I was continually caught in having to decide between what constitutes ideal practice and what may be practical. Generally, I attempted to prescribe standards for the ideal and

discuss attendant practical considerations — rather than attempting to assess what may be practical in most situations, and discussing limitations because the practical was less than ideal. The second difficulty, related to the first, is that even the best practice of monitoring has limitations. In attempting to set a standard for good practice I wanted also to delineate the shortcomings of monitoring. However, the requirements for best practice may be overwhelming to administrators wanting to get started, such that they decide monitoring is too costly given their resources. There is the danger too that by providing detailed descriptions of the shortcomings of monitoring, I would give the impression that there are so many problems with it that it may be best not to do it at all. I chose to present an optimistic picture of what could be accomplished through monitoring, but have not ignored the many limitations. I hope that the reader will view the work as simply a guide for better assessment, and not prematurely judge the benefits or limitations of monitoring.

Overview of the Book

The next chapter examines the above premises in the light of current reforms and policy initiatives in the UK and the US. The two systems are very different, especially at the secondary level, in part because of the long history of national examinations in the UK. These examinations to a large extent drive the curriculum, and ensure a degree of uniformity. The US curriculum is characterized by diversity more than uniformity, which poses special problems for the development of monitoring programs. In both schooling systems there is an explicit agenda to develop indicators for purposes of accountability.

Chapter 3 describes the input-process-output model, the theoretical model on which systems of monitoring performance are based. It also discusses how three different types of monitoring systems, defined by their purposes, are related to this model. I suggest three ways in which the model can be strengthened. These concepts underlie many of the arguments in the chapters that follow regarding the kind of data to be collected and the approach to analysis.

Chapter 4 describes four ways that researchers and educators use the term 'school effects'. Using the definition relevant to the comparison of schools, I suggest there are two types of school effects that should be considered. These are defined and a model for their estimation is presented. I also discuss some of the technical issues raised by the estimation of school effects and the comparison of schools.

Chapters 5, 6 and 7 outline the substantive and technical issues concerning the measurement of schooling inputs, processes, and outcomes, respectively. In the first two of these chapters I distinguish between schooling inputs and processes. 'Schooling inputs' is used to refer to factors exogenous to the schooling system; that is, factors associated with pupils' family backgrounds, and the social, economic, and political factors that affect schooling outcomes but lie outside the control of teachers and educational administrators. 'Schooling processes' is used to refer to factors directly related to school policies and practices. This distinction is not always a comfortable one because many factors are related to school policy and practice and are influenced also by

forces outside of schooling. The distinction is important mainly for the statistical modelling of school effects; Chapters 5 and 6 discuss the reasons for this distinction in detail.

The purpose of Chapter 5 is to discuss the role of input measures in analysis and to make recommendations for their measurement. I accomplish this to some degree by employing data describing pupils' family backgrounds, cognitive abilities, and schooling outcomes for a large sample of pupils that attended primary and secondary schools in Fife. The chapter also discusses strategies for handling missing data.

Chapter 6 attempts to specify a 'best set' of process indicators. But here I attempt to accomplish this by reviewing the literature and proposing some criteria associated with their coverage of the domain, their usefulness, and their measurability. The measurement of schooling processes is in many ways more difficult than the measurement of inputs or outcomes, and therefore the chapter includes considerable discussion about the problems entailed in measuring and interpreting school process data.

Chapter 7 describes some of the outcome measures that can be included in a monitoring system. I argue that monitoring systems should be based on a wide range of outcome measures. The chapter discusses issues concerning the identification of the goals of schooling, and outlines the major considerations in selecting appropriate tests and constructing indicators. One important issue concerns whether indicators emphasize *equity* or *excellence* in schooling. For example, an indicator of the percentage of pupils that achieved some minimum level of competency underscores the importance of equity more than an indicator of the percentage of pupils that achieved outstanding results on a statewide achievement test. A discussion of validity and reliability of outcome measures is included here. Readers unfamiliar with these terms may wish to read Chapter 7 before Chapters 5 and 6.

The discussion in Chapters 4 to 7 suggests that it is impossible to specify a set of definitive principles on how to develop a system for monitoring school or district performance. The development of a system requires many inter-related decisions, most of which have political ramifications. At the end of each of these chapters I offer a set of guidelines for the development of a monitoring system. These should be considered guidelines, not definitive principles.

Chapter 8 presents a design of a system for monitoring schools at the district or EA level. This design is based on earlier designs set out for a school district in Canada and an educational authority in Scotland. Variants of this design are now being implemented in these settings. The purpose of the chapter is not to present a fully comprehensive design, but rather to provide some starting points for a school district or education authority beginning the process. The chapter proposes the kind of data to be collected at various levels, and discusses the problem of confidentiality. It also specifies the stages for developing a system and a time line.

Chapter 9 does two things. It delineates the types of analyses that could be included in an annual report, and describes the statistical and graphical techniques connected with each type of analysis. These techniques could be used for describing the performance of individual schools or entire school districts. Most of the analyses can be done with commercially available

statistical software packages such as SPSS/PC or SYSTAT, and graphical packages such as Harvard Graphics.

Chapter 10 discusses how information from a monitoring program can be used to develop a district-level research program aimed at answering research questions relevant to a district's needs. I begin by specifying four basic questions that pertain to nearly all research on school effectiveness. These questions provide a framework for discussing the strengths and limitations of various designs. The first type of design discussed requires only cross-sectional data; it could be employed after the first year of operation of a program. After two or three years of operating a monitoring system, longitudinal data could be used to assess the effects of particular interventions, or to examine whether certain policies or practices improve outcomes or reduce inequalities between high- and low-status groups. The chapter includes examples from the research programs in Scotland and Canada. This is the most technically demanding chapter. I have strived to make it easier for the reader who prefers words over equations by moving the technical description of multilevel modelling to an appendix.

The final chapter provides an executive summary of the material covered in the first ten chapters. It also suggests how administrators and policy-makers might tackle some of the political issues concerning accountability, reduction of the curriculum, and teachers' professional autonomy.

