

PRIORITIES AND VALUES IN ACCOUNTABILITY PROGRAMS

Philip Nagy

Henk Demeris

Christina van Barneveld

Ontario Institute for Studies in Education, The University of Toronto
Toronto, Ontario

Abstract: The authors examine relationships among three principal elements of educational accountability: assessment of achievement, collection of educational indicators, and setting of educational policy, using a Venn diagram to portray interactions among the three. Their main argument rests on the importance of understanding value perspectives in accountability. Because different value positions are largely unexamined, important data remain uncollected, and much that is collected remains under-analyzed. As a consequence, important issues remain un-addressed, and accountability fails to serve the needs of education.

Résumé: Les auteurs examinent les relations entre trois éléments principaux de la responsabilité d'enseigner: l'évaluation des accomplissements, le recueil des indices révélateurs éducatifs, et l'identification des politiques éducatives, en se servant d'un diagramme de Venn pour représenter les interactions parmi les trois éléments. Leur raisonnement repose sur le fait primordial de la compréhension des perspectives mises en valeur par la responsabilité. Parce que les positions en valeur diffèrent et sont inexaminées en grande partie, certaines données ne sont pas recueillies, et beaucoup des données ramassées ne sont pas analysées à leur juste valeur. Conséquemment, certaines questions ne sont pas examinées, et la responsabilité ne dessert pas les besoins éducatifs.

■ The purpose of this article is to examine relationships among three principal elements of educational accountability: assessment of achievement, collection of educational indicators, and setting of educational policy. In much of the current discussion of accountability, important data remain uncollected, data that is collected remains under-analyzed, and relationships among achievement, indicators, and policy are not clearly examined. As a

consequence, important issues remain unaddressed and accountability fails to serve the needs of education.

Increasing public concern over the quality of education and demands on limited resources have given rise to the educational accountability movement. Relevant data can and should contribute important information for decision making and policy setting. However, the design of assessments, selection of indicators, and analysis and interpretation of data are value-laden, that is, decision makers impose their own values on the accountability process and only a narrow range of value positions is served. In this article, we highlight the role values play in relationships among achievement, indicators, and policy. A clearer focus on values can broaden and enhance the quality of educational discussion and the utility of accountability data for school improvement.

Two examples from outside the educational system may clarify our concern. When unemployment drops and workers are in shorter supply, the price of labour rises, profits suffer, and stock prices drop. Those doing the work and those making the investment, who may indeed be some of the same people, have competing interests. If the health of the economy is viewed from the perspective of the labour force, the same data lead to a different conclusion than if it is viewed from the perspective of the investor. Similarly, the World Bank demands to set the economic rules within a developing country in exchange for financial assistance. If the criterion is an improved international balance of payments, then these rules seem to improve things. If the criterion is the quality of daily life, then these rules seem to be a detriment. Thus, different value perspectives suggest different data and yield different conclusions, indeed different policies.

The relevance of accountability systems to the entire educational community depends on the extent to which all stakeholders' interests and values are reflected in them. Different stakeholders hold multiple and often competing values, which lead to multiple and competing expectations for the use and application of data. A re-examination of current practices and the values they reflect is required. In the first section of the article, we define key terms. In the second section, we describe the relationships among them using an Assessment, Indicators and Policy (AIP) model. In the third section, we illustrate the importance of value conflicts in data collection, indicator selection, analysis, and interpretation. Finally, we review

the consequences of the limited value positions reflected in current educational assessment programs and discuss alternatives that address this issue and may be implemented in future educational assessment programs with a minimal allocation of additional resources.

DEFINING TERMS

Accountability

Concerns over the quality and cost of education over the last two decades have led to calls for greater accountability and improvement in the education system and a corresponding increase in efforts to monitor and evaluate the education system (McEwen, 1996). The function of an accountability system “is to oversee (monitor and evaluate) the performance of the education system and propose needed changes to policymakers” (Wohlstetter, 1991, p. 31). Effective accountability mechanisms use performance indicators to assess the efficacy of the education system. Darling-Hammond and Ascher (1991) noted that accountability occurs only when processes exist for interpreting and acting on the information provided by the performance indicators.

Earl (1998) concurs with these same two purposes of reporting and improvement, but also makes the distinction between *accounting*, the mere collection of information, and *accountability*, which she describes as forward looking, and “taking the information to make judgments — about quality, about how good is good enough, and, most importantly, about how to make changes that will enhance and extend student learning” (p. 21). It is this view that supports the thrust of our argument, in effect that what was intended as accountability has become merely accounting, and narrowly value-laden accounting at that.

Achievement

We define achievement as student accomplishment in cognitive, psychomotor, or affective areas, as measured by test or other assessment instrument (e.g., a portfolio). Excluded from our definition of achievement are many components of a typical assessment program (e.g., student attitudes, demographic information), ancillary data that are collected simply for convenience at the same time as testing. These we put under the umbrella of *Indicators*, defined next.

Indicators

Although there is no single definition for indicators, there is some agreement on their purpose. Indicators are designed to provide information about the state of the educational system and to assist in policy analysis, evaluation, and formation. Issues of value arise because indicator systems can be structured in one of two ways. As Earl (1993) points out, indicators can be used simply to collect selective pieces of information as requested by policy makers or organized as a set of interrelated indicators based on a conceptual model. Indicator systems involve multiple sources of information about the complexity of student achievement and resources, demographics, policies, structures, and processes that characterize schools and school systems. To demonstrate the importance of values to what we consider an appropriate conception of indicators, we turn to history.

Indicator systems arose from the realization that achievement results alone are inadequate as a measure of school quality and as a guide to corrective action. The birth of the indicator movement can be seen in the *Coleman Report* (Coleman et al., 1966) and in the commentary and criticism that it generated (e.g., Berliner & Biddle, 1995; Freudenthal, 1975).

Consider *Education Counts* (National Center for Education Statistics, 1991) as an example of an early indicator framework. This document laid out an "overall conception of how an indicator system should be developed" (p. 3). It called for data collection in six major areas: learner outcomes (defined above as *achievement*), quality of educational institutions, readiness for school, societal support for learning, education and economic productivity, and equity. These categories capture, at least in principle, virtually all that could be asked of an indicator system. However, its very comprehensiveness and complexity highlight the difficulty of putting it into practice. Value-laden compromises are required in selecting what and how much data will be collected and examined to achieve the goals of such comprehensive systems within a reasonable budget. As Kaagan (1990) has pointed out, such a system is a framework "into which an array of value-laden statistics or indicators are placed for review and analysis, leading to necessary modifications of policy and practices" (p. 55).

Policy

In this article, we view policy simply as decisions made by educational authorities, either elected or appointed, that impact on the operation of schools.

Values

Our definition of a value is a perceived *good*, balanced where appropriate by aspects of cost, as in a cost-benefit ratio. The range of possible responses to Herbert Spencer's classic question "What knowledge is of most worth?" provides a case in point. There are those who espouse education that is largely focused on job training (Coalition for Education Reform, 1994; Economic Council of Canada, 1992) and those who take a much broader view of the purpose of schools, considering education to be a way of serving the whole person and passing on our cultural heritage (e.g., Greene, 1995; Miller, 1999). The differing implications of these positions with respect to both funding and assessing education are obvious.

Our own values are transparent in what follows, and we have chosen to strive for balance rather than neutrality. In fact, this article is motivated in part by our objections to the value position that focuses entirely on costs, in effect dismissing benefits as irrelevant.

Values are reflected in the priorities that are set by educational assessment programs. Based on our observations of educational assessment systems, we present a four-stage continuum of assessment priorities:

- **High Priority:** Some data are collected and analyzed in the initial report of an assessment. These are clearly of primary importance and highest value to those responsible for the program. For example, the Council of Ministers of Education (1996) report on the School Achievement Indicators Project (SAIP) includes a breakdown of results by province and results by gender. Some ancillary data are given, but are not related to achievement in their report;
- **Lower priority:** Some analyses are not done in the initial report, but shortly after, either by or under contract to the assessment agency. For example, Lawson, Penfield, and Nagy (1999) examined gender-related issues after the release of the first Ontario Grade three assessment report (EQAO,

1997). While this work was done at the behest of EQAO, these results are judged of slightly lower priority and value for the agency than those included in their initial report;

- Permitted: Some analyses are not done by or paid for by the assessment agency, but by others with agency permission. For example, Nagy (1997) examined patterns of achievement of those who speak different languages at home and school using the Council of Ministers assessment data (CMEC, 1994). These results were of considerably less importance, and less value, to those who commissioned the original data collection; and
- Impossible: Some analyses cannot be done at all because the data were not collected. For example, data in the CMEC (1993, 1994) projects that would allow a strong analysis of the relationship between socio-economic conditions and achievement were not collected for political reasons. Such analyses are of even lower value to the sponsoring organization.

These four levels of priority serve as a backdrop for our discussion of relationships among achievement data, indicators, and policy as viewed from a values perspective. To describe these potential relationships, we developed the AIP model described below (see Figure 1 for a graphical representation of this model).

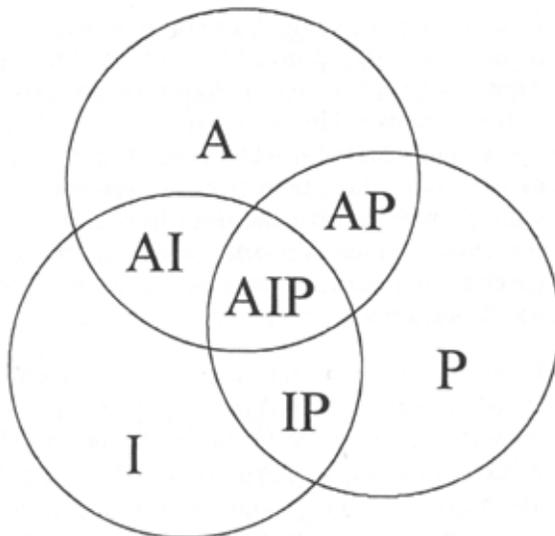
THE AIP MODEL

Achievement (A), indicators (I), and policy (P) can operate together or separately, as pictured in the seven zones of Figure 1. Examples for each zone are provided below:

- A only — Achievement data are sometimes collected and left unused, playing no role in either an indicator system or a policy context. Assessment data collected in advance of the establishment of an indicators program, such as the EQAO (1997) program in Ontario, could be considered in this category, even though schools were asked to develop action plans on the basis of the data (see PA below). This same program can be faulted for spending too many resources on collecting data (i.e., testing every student in a grade every year), so that little time or money is left for effective assimilation and action.

- I only — There are two different kinds of examples to be noted here. The first is that considerable indicator data is unuseable. For example, the 1996 report of the Pan-Canadian Indicators project (Council of Ministers, 1996) contains achievement data and considerable contextual data, but they remain unrelated. There are, for example, detailed graphs showing the increasing rates of immigration over a number of years, and presumably the reader is to draw some inferences about education. There are no data on the relationship between ESL and achievement; the report amounts to accounting, not accountability. Second, indicator data can be ignored. For example, the Ontario government's educational funding formula (*Education Quality Improvement Act*, 1997) takes no account of the fact that poor children have special needs simply because of their poverty, despite ample evidence of the relationship between socio-economic conditions and achievement (summarized, for example, in Nagy, Demeris, Lawson, Penfield, and van Barneveld, 1999). We are not questioning, of course, that many poor children do succeed in school.

Figure 1
The AIP (Achievement, Indicator, Policy) Model



- P only — Policy is often made without use or consideration of either achievement or indicators. For example, recent reforms to Ontario education have been made on the basis of a presumed *drop* in both public confidence and achievement levels. However, there is no evidence to support either contention. Evidence that achievement is not as high as many would like is not evidence that it is declining. As Maehr and Maehr (1996) put it, “Schools aren’t as good as they used to be; they never were” (p. 21). Considering public perception, as Livingstone (1998) puts it, “In spite of extensive rhetoric about the decline of education, public opinion about the quality of the schools has remained quite stable over the past decade” (p. 19).
- PA only — The use of achievement data for policy making in the absence of contextual information is a serious concern, but a rarity. “Horse-race” reporting of large-scale assessment, in which the media publish ranked lists of the best and worst schools or districts, based entirely on achievement, come to mind, but this hardly qualifies as policy. The request by many Ontario jurisdictions for schools to formulate action plans based on the results of the 1997 grade three assessment (EQAO, 1997) is closer to policy, but thanks to the unspecific nature of the data and the good sense of school staffs, these plans are uniformly generic and bear little relationship to high or low achievement.
- IA only — Some indicator programs, even with appropriate achievement data, have little potential for impact on policy. This is highlighted by the four levels of assessment priority outlined above. The first two levels, *high* and *lower priority*, reflect the interests of policy makers, but even here achievement data tend to be under-analyzed. That is, not enough is done with the data in the initial report and immediately ensuing reports. So much time and effort go into getting the data that few resources are left for understanding it and acting on it.
- IP only — One of the most ironic and unfortunate situations concerns the making of policy based on indicator information, specifically perceptions, in the absence of achievement data. Berliner and Biddle (1995) document attempts by some groups in the US to paint the state of American public education much worse than it is for their

own tax-saving purposes. In Canada, in an effort to sway public opinion toward a “back to the basics” curriculum, the Coalition for Education Reform (1994) has set up caricatures of various positions they disapprove of — whole language, social promotion, self-esteem — and exaggerated the beliefs of educators with respect to these issues. This is not a pure IP example, however; this group have also attempted to seriously distort achievement data (Nagy, 1994).

- IPA — Policy decisions made in an achievement *and* indicator context are rare. One place to look is in the field of value-added analysis (Meyer, 1997; Thomas and Mortimore, 1996), where attempts are being made to address the question of how well schools are teaching in a broader and more appropriate context. In value-added analysis, multiple regression and hierarchical linear modeling procedures are used to identify which schools are “adding value” to their students. Previous achievement and school and socio-economic conditions are included in the models, so that the effects of the school that year can be isolated. The procedures are far from well-developed, but their intent is to identify schools that are doing a good (or poor) job given their circumstances, to find out why, and then to try to generalize to other contexts. Achievement and indicators are used together to directly inform policy. Our own work examining achievement of “schools like me” focuses on this problem using Ontario data (Nagy et al., 1999). For example, one part of our analysis shows the distribution of schools in eight groups of several hundred schools each, based on average household income as reported in the 1996 Canada census.

Thus far in the argument, we have laid out conceptions of assessment priorities and issues arising in the relationships among assessment, indicators, and policy decisions. We have pointed out examples of value conflicts, such as attempting to convince the public that public education is not functioning and does not deserve tax support. We turn now to some more systematic examination of value conflicts.

VALUE CONFLICTS IN EDUCATION

Different values point to the possibility of examining the success of the education system in different ways. These can involve both the process and the outcomes of education.

The Second International Mathematics Study (Robitaille and Garden, 1989) serves to exemplify the tension between enrolment and retention and average attainment. Nagy (1996) discusses a systematic pattern across jurisdictions; those who offer senior high school mathematics programs to elite students only (England, Japan, Hong Kong) have much higher average achievement than those who encourage higher participation in such programs (British Columbia, Hungary). Generally, if expectations are raised, fewer students will attain them. If we look at enrolment, or if we look at achievement, different countries win the "horse-race." There are alternative ways to measure the success of schools, with different policy implications.

For a second example, consider the grade three assessment programs of Ontario's Education Quality and Accountability Office (e.g., EQAO, 1997). Students were given scores on several four-point scales, but a few students did not produce enough material for scores to be produced, and in effect, were given "0" on the scale(s). A heated discussion ensued over whether these students should be included in the denominator when calculating such data as the percent of students in a school (or district) at or over a given level. To exclude the student makes the school look better, but it also hides a problem of numbers of quite needy students who must be served by the school. Different value positions are reflected in how the data are portrayed. EQAO decided to produce the calculations both ways.

Finally, consider different views of the curriculum. Since assessment of achievement tends to focus on basic skills, there is a tendency to relegate other aspects of a full education (e.g., art, music, second language) to second place. Apart from that, even among those who feel that the only, or at least primary, purpose of education is job preparation, there is debate. Some (e.g., Coalition for Education Reform, 1994) take a very narrow basic skills approach, while others (Conference Board of Canada, 1992) take a broader view. The Coalition calls for an entirely academic curriculum, with extensive practice, focused on first language (only), mathematics, history and geography, and sciences (Coalition for Education Reform, 1994, p. 48). In contrast, the Conference Board, funded entirely by Canadian business, includes motivation for life-long learning, positive attitudes and behaviours, responsibility, adaptability, and the skills required to work with others. The debate occurs in other jurisdictions. Even countries who have traditionally done very well on international tests, such as Singapore, are rethinking the narrowness of their curricula (Broader Criteria for Schools, 1999).

When we turn to the process of education, again different value positions arise. The clearest examples come from work on different views of the educational reform process. The top-down view is that people can be led to change by command or coercion. Examples abound, from attempts to introduce accountability programs in the face of teacher opposition (Black, 1994) to the current notion that the mere existence of a testing program will cause improvement in test scores (Hansen, 1993). The opposite views are that improvement of any sort requires the active co-operation and support of those directly involved (Fullan, 1991) and that actions based on test scores, not the scores themselves, can lead to improvement (Hansen, 1993). Differences in approach are highlighted by several writers in a special issue of *Orbit*, a publication of the Ontario Institute for the Study of Education of the University of Toronto, on the top-down approach of the Ontario government's "Bill 160" (Hargreaves, 1998).

Consider the clear dichotomy between the current, simplistic views on school finance prevalent in some government circles and the real world of school finance. For example, the Ontario government has recently imposed a funding formula which, within rather narrow limits, allows the same amount for the education of every child in the province. To be fair, the funding formula does make adjustments for elementary versus secondary students and for special education. Despite the huge literature showing the strong relationship between poverty and educational outcomes (e.g., Thomas and Mortimore, 1996), the formula does not take this into account. The value position taken is akin to objections that progressive tax systems are unfair to the rich. Contrast this with the view of Hanushek (1997), "Simple resource policies hold little hope for improving school outcomes" (p. 141). Different values leads to different resource allocations.

Value questions become prominent when we consider Berliner and Biddle's (1995) argument that at least some of the criticism of public education by the business community is an attempt, implicitly dishonest in their view, to lower input costs. These attempts focus both on reducing public support and thus taxation levels and on shading curriculum emphases in the schools in order to reduce workforce training costs. In the Canadian context, the Economic Council of Canada (1992) documents both the dissatisfaction of business with the results of schooling and the high percentage of firms who do not engage in any formal staff training programs (p. 23).

Any discussion of values cannot be complete without a brief comment on aspects of our culture that relate to our education system. Nagy (1994, 1996) has argued that North Americans exhibit contradictory attitudes about childhood, mathematics and science, and work. He presents three examples:

- While Korean teenagers are expected to be “responsible for their own serious behavior” (Lapointe, Askew, & Mead, 1992, p. 24), we have much more relaxed expectations for young teens.
- While we call for more scientifically and technically trained people, we brand mathematicians and scientists as egg-heads; is it any wonder that we have difficulty interesting more of our youth in mathematics and science?
- While some members of the public call for more homework and a longer school year for children and teachers, as a society we agitate for shorter work weeks and longer vacations for ourselves.

Our point is not to argue that we should all become lazy or that children do not need to learn mathematics and science. It is simply to point out that if we are asking the schools to produce harder-working children who love mathematics and science, our culture makes it an uphill battle.

CONCLUSIONS

Our argument, in summary, is that some important indicator data are not collected and that they are under-used when interpreting student achievement results. Current data collection, indicator selection, analysis, and interpretation reflect a narrow value position, while other valid value positions are ignored or minimized. Policies are being developed that serve only a fraction of actual stakeholders.

On the one hand, we recognize that collecting data is expensive. We should use the data we collect. On the other hand, we can't analyze data we don't have. We need achievement data, indicator data, and policy to be interfaced more effectively as indicated in our AIP model. Multiple-value positions must be considered and reported. Examining only one value position, for example, focusing almost exclusively on either only costs or only benefits, is inappropriate. We need an interpretive conceptual framework that goes beyond simply collecting data.

Achievement scores alone cannot capture all of the important features of education systems. We proposed an Assessment, Indicators, and Policy (AIP) model to identify assessment priorities and issues arising in the relationships among these three. This model will assist in selecting indicators that will be most useful in making important judgments. In contrast, much current indicator data are selected because the information is already available or in response to the current political agenda.

Since indicator systems send messages or signals, the choice of signals should be carefully made. Many promising indicator systems fail because they are too expensive or too politically costly to implement. Therefore, further research is recommended to examine the values and priorities in assessment beyond the cost-benefit ratio.

REFERENCES

- Berliner, D.C., & Biddle, B.J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Don Mills, ON: Addison-Wesley.
- Black, P.J. (1994). Performance assessment and accountability: The experience in England and Wales. *Educational Evaluation and Policy Analysis, 16*, 191-203.
- Broader criteria for schools. (1999, March 18). The (Singapore) Straits Times.
- Coalition for Education Reform (1994). *Could do better: What's wrong with public education in Ontario and how to fix it*. Toronto: Author.
- Coleman, J.S., Campbell, E.Q., Hobson, C.F., McPartland, A.M., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). *Equality of educational opportunity*. Washington, DC: Department of Health, Education, and Welfare.
- Conference Board of Canada. (1992) *Employability skills profile*. Ottawa: Author.
- Council of Ministers of Education, Canada. (1993). *School Achievement Indicators Project: Report on mathematics assessment, 1993*. Toronto: Author.

- Council of Ministers of Education, Canada. (1994). *School Achievement Indicators Project: Reading and writing*. Toronto: Author.
- Council of Ministers of Education, Canada. (1996). *Education indicators in Canada*. Toronto: Author.
- Darling-Hammond, L., & Ascher, C. (1991). *Creating accountability in big city schools*. Urban Diversity Series No. 102. New York: National Center for Restructuring Education, Schools, and Teaching. ERIC No. ED 334 339.
- Earl, L. (1993). Developing indicators: Answering the quality question. Research report 93-94/04. Scarborough, ON: Board of Education.
- Earl, L. (1998). Developing indicators: The call for accountability. *Policy Options*, 19(6), 20-25.
- Economic Council of Canada. (1992). *A lot to learn: Education and training in Canada*. Ottawa: Author.
- Education Quality and Accountability Office. (1997). *Provincial report on achievement: English-language schools*. Toronto: Queen's Printer.
- Education Quality Improvement Act*. (1997). www.edu.gov.on.ca/document/nr/97.09/septbcgr.html
- Fruedenthal, H. (1975). Pupils' achievements internationally compared: The IEA. *Educational Studies in Mathematics*, 6, 127-186.
- Fullan, M.G. (1991). *The new meaning of educational change*. New York: Teachers College Press.
- Greene, M. (1995). *Releasing the imagination: Essays on education, the arts, and social change*. San Francisco: Jossey-Bass.
- Hansen, J.B. (1993). Is educational reform through mandated accountability an oxymoron? *Measurement and Evaluation in Counseling and Development*, 26, 11-21.
- Hanushek, E.A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, 19, 141-164.

- Hargreaves, A. (Ed.). (1998, Special Issue). *Orbit*, 29(1).
- Kaagan, S.S. (1990). State education indicators: A mandate for action. *NASSP Bulletin*, 74(527), 54-65.
- Lapointe, A.E., Askew, J.M., & Mead, N.A. (1992). *Learning science*. Princeton, NJ: Educational Testing Service.
- Lawson, A., Penfield, R., & Nagy, P. (1999). Relating attitudes, gender and student achievement in grades 3 and 6. Technical Report #1 for the Education Quality and Accountability Office. Toronto: EQAO.
- Livingstone, D.W. (1998). Public opinion. *Orbit*, 29(1), 17-22.
- Maehr, M.L., & Maehr, J.M. (1996). "Schools aren't as good as they used to be; They never were." *Educational Researcher*, 25(8), 21-24.
- McEwen, N. (1996). Student outcome indicators in Canada: Uses, issues and solutions. *International Journal of Educational Research*, 25, 219-230.
- Meyer, R.H. (1997) Value-added indicators of school performance: A primer. *Economics of Education Review*, 16(3), 283-301.
- Miller, J. (1999). *Education and the soul: Toward a spiritual curriculum*. Buffalo: SUNY Press.
- Nagy, P. (1994). *National and international comparisons of student achievement: Implications for Ontario*. A Report to the Ontario Royal Commission on Education.
- Nagy, P. (1996). International comparisons of student achievement in mathematics and science: A Canadian perspective. *Canadian Journal of Education*, 21(4), 396-413.
- Nagy, P. (1997). Secondary analysis of data from Canada's indicators program. Presented at the annual meeting of the American Educational Research Association, Chicago.
- Nagy, P., Demeris, H., Lawson, A., Penfield, R., & van Barneveld, C. (1999). Improving school comparisons using "Schools Like Me." Final report to the Education Quality and Accountability Office.

- National Center for Educational Statistics. (1991). *Education counts: An indicator system to monitor the nation's educational health*. Washington, DC: US Government Printing Office.
- Robitaille, D.F., & Garden, R.A. (Eds.). (1989). *The IEA study of mathematics II: Contexts and outcomes of school mathematics*. New York: Published for IEA by Pergamon Press.
- Thomas, S., & Mortimore, P. (1996). Comparison of value-added models for secondary-school effectiveness. *Research Papers in Education*, 11(1), 5-33.
- Wohlstetter, P. (1991). Accountability mechanisms for state education reform: Some organization alternatives. *Educational Evaluation and Policy Analysis*, 13(1), 31-48.