

School Board Evaluation Programs: Case Studies of Nine Boards in Ontario*

Raymond Danley, Merl Wahlstrom, and Helen Hayes
ontario institute for studies in education

Dennis Raphael
university of western ontario

Le Ministère de l'Éducation de l'Ontario a demandé en novembre 1976 une étude des programmes d'évaluation administrés par divers conseils régionaux de la province. On a donc choisi neuf programmes pour en faire l'analyse détaillée. Par le moyen d'interviews, on a recueilli des données qui aident à décrire le développement et l'organisation des programmes, quels tests font partie de ces programmes, comment on en utilise les résultats, et ce qu'on attend du ministère sous forme de direction ou de facilités. Les neuf études de cas font voir le fonctionnement des programmes de plusieurs points de vue, ce qui est important si l'on veut en estimer non seulement la valeur pédagogique mais aussi les bienfaits politiques.

In recent years, Ontario's approach to educational evaluation has tended to move away from uniform standardized testing practices towards a greater reliance on the judgment of individual teachers. This approach may be seen as a corollary to the policy of local curriculum development which has characterized education in Ontario since the late 1960s. While allowing for greater flexibility, it has also resulted in greater variability among school systems in the province — a variability which has seemed to many to reflect uncertainty regarding standards of student achievement. The Ontario Teachers' Federation articulated this concern in its report to the cabinet in May 1976:

There appears to be widespread concern about levels of academic progress and behavioural standards being achieved in the schools. We may have reached the time when a comprehensive survey of school standards is needed so that remedial measures may be taken should these apprehensions prove to be well founded.

In response to these concerns, the Minister of Education announced in November 1976 that a "work group" had been formed to devise a plan of action to improve testing and evaluation in Ontario classrooms. Three primary objectives were to direct the work of this group:

*The research on which this paper is based was funded under contract by the Ministry of Education, Ontario.

1. to help teachers to "better diagnose any specific deficiencies in the levels of pupil achievement, especially in the basic skills, to help them improve classroom instruction and to work with individual students as required"
2. to provide for all levels of the education system "practical information that will help determine whether curriculum objectives are being met, and to assist in bringing about improvement"
3. to give parents "clear and understandable information about their children's progress and achievement," so that they can assist and encourage them at home. (Note 1)

Not surprisingly, reaction to the minister's move was defensive, and it was realized that more detailed information was needed on the nature and functioning of evaluation programs within local boards. Accordingly, a study was outlined by the ministry to examine aspects of current evaluation programs in the province, in order to discover

- how programs are developed
- how they are organized
- what tests are currently being administered
- how results are used in the system, and
- how the role of the ministry in evaluation programs is perceived.

In this study of the evaluation programs of school systems, evaluation was conceptualized as an information-gathering process which serves educators in making decisions. This conceptualization is consistent with Stufflebeam's (1971) definition of the evaluation process as "data gathering for future decision making." Central administrators who have the responsibility of monitoring the quality of the total educational program develop system-wide evaluation programs to provide them with information about the system. They use the information as input for future program decisions and to be well-informed about the state of the system in discussions with trustees, ratepayers, and officials from the Ministry of Education.

This conceptualization makes three assumptions about evaluation: (a) that the primary purpose of a system-wide evaluation program is to serve the information needs of central administrators; (b) that evaluation programs vary according to the amount of pressure placed on administrators to provide data for accountability and standards; and (c) that the procedures for implementing the evaluation program vary according to the organizational structure of the system.

It was decided that a case-study approach would best accomplish the aims of the study. Interviewing personnel at all levels would enable the researchers to obtain detailed information and to document varied percep-

tions of a board's testing program. The boards selected for study would all have system-wide programs in operation, but they would represent a range of other characteristics. The study would thus provide insights into current evaluation practice and hence important background data for those who have been given the mandate to plan for improved testing and evaluation of pupil achievement in the classrooms of Ontario.

METHOD

The study involved a detailed examination of the testing practices of nine school boards in the province. An extensive interview procedure was used to obtain the in-depth information required by the Ministry of Education; it was felt that a survey type would not be suitable for this task. The nine school boards were not randomly selected, but were identified by ministry personnel from survey information available to them. These nine boards reflected various sizes, geographical locations, types of board (public/separate), and types of student (ethnic mix). All nine boards had system-wide testing programs, but appeared to vary in the types of measures used. Three of the boards were seen as relying primarily on norm-referenced tests, three as preferring criterion-referenced measures, and three were known to have developed some instruments of their own for local use.

Once the nine boards had been identified, the research team prepared to interview individuals representing the range of roles and responsibilities across each system, viz., directors, superintendents, measurement specialists, principals, teachers, and parents. The interview was designed to elicit information concerning current testing policy and practice, recent changes and trends in evaluation procedures, and methods of using and communicating test results. It was expected that interviewing a variety of individuals from each board would provide a multi-faceted view of testing policies and procedures.

There were three interview schedules: one for the administrative level (directors, superintendents in charge of testing, measurement experts or consultants), one for school personnel (principals and teachers), and one for parents. All interviews focused on the three areas of concern: the policy structure of the evaluation program and current practices, recent changes and trends, and the use and communication of evaluation results. For the administrative level, questions were designed to elicit a general description of the program and its rationale, program organization and responsibilities, and policy on uses of evaluation results, on program changes and on general adherence to the program. In the principal and teacher interviews, questions focused on perceptions of the nature and value of the current program, on the facets of achievement being measured, and on the ways in which test results are being used and communicated. The parent interview examined parents' perceptions of the formal

testing program, of program changes, and of the availability of evaluation results.

In co-operation with the ministry, the project team contacted board directors through letters and telephone calls, and interviews were scheduled over a two-day period in each system. Most interviews were tape recorded after securing permission from the person being interviewed. This allowed for a relaxed session and made possible a subsequent analysis of interview content. The length of the interview combined with the short duration of the study made complete transcription infeasible, but the relevant information was noted. Findings from the nine case studies were then synthesized to give an overview of evaluation practices in Ontario.

RESULTS

The findings of the case studies will be presented here in summary form, in answer to the questions which formed the basis of the study:

1. How are the testing programs developed?
2. How are the programs organized?
3. What tests are currently being administered?
4. How are results used in the system?
5. What is the perceived role of the ministry in evaluation programs?

In the discussion of each aspect, reference will be made to the prior categorization of boards as oriented towards norm-referenced, criterion-referenced or locally prepared tests. In this manner, examination of possible differences among the selected boards can be made.

1. How are the testing programs developed?

Frequently, the development of an evaluation program has been a natural concomitant to the development of a local curriculum. The case studies of the nine boards also show the influence of the recent "back to basics" and accountability trends. The impetus for development of an evaluation program, however, comes from a variety of sources. At one end of the continuum, a decision can come from someone in authority desiring a particular type of information or acting in response to a variety of perceived pressures. At the other end of the continuum, a program can have risen from a "grass roots" movement among school staffs and/or parents. Most of the boards reportedly had the main impetus for development from a single source, usually the director. The director may be seeking evidence that the county is measuring up to the expectations of the ministry, or is achieving at levels comparable to other boards. He may perceive a need to obtain information upon which administrative decisions can be based.

Boards whose perceived orientation is towards norm-referenced testing tend to have programs which are initiated at the administrative level. Boards selected as being oriented towards criterion-referenced testing tend to report more "grass roots" pressure and involvement than other boards. Cases in which parents and/or trustees have applied pressure occur only in two of these boards. One is a separate school board, and the other serves a very affluent upper middle-class community. Boards selected for their locally prepared programs report a variety of influences: an administrator's desire, a long-functioning special services division, or a research and development branch responsive to the needs of principals and teachers.

2. How are the testing programs organized?

Organization structures varied across the boards. In some cases, committees responsible for the evaluation program consisted solely of administrative personnel, and the duties of these committees tended to be limited to the selection of standardized tests in response to directives from the administration. Boards oriented towards norm-referenced testing tended to have this type of structure. The testing program is also relatively stable with little immediate change anticipated.

The boards oriented towards criterion-referenced measures appeared to be less centralized and had many committees comprising personnel from all levels. These committees were not only responsible for selection of readily available tests, but were also involved in developing locally prepared measures. Further development along these lines is envisaged if the programs are well received.

Quite striking in their characteristic structure were the boards which emphasized locally prepared tests. All three of these boards had highly developed and relatively autonomous evaluation sections involved not only in local test construction but also in a number of research projects. The two boards most actively involved in local test development had enthusiastic testing directors who were very competent and appeared to enjoy good relationships with the administration. These boards plan to continue the development of local testing and evaluation programs, especially to measure attitudinal and affective spheres of behavior and to evaluate new locally prepared curricula.

Concern was expressed in some boards about resources available for developing local evaluation programs. Developing and implementing evaluation programs stretch the resources of a board, and the future of local evaluation depends on the availability of finance.

3. What tests are currently being administered?

Almost all the boards studied had mixed use of norm-referenced and locally prepared tests, though the sophistication level of these latter types

of tests differed widely across boards. Boards which used norm-referenced achievement tests were most likely to be using the Canadian Test of Basic Skills, and tended to be highly dependent on these measures. Yet in all three of these boards, a locally prepared mathematics test had been developed, apparently because of dissatisfaction with the currently available achievement tests and a need to evaluate the local mathematics curriculum.

Boards oriented towards criterion-referenced testing tended to be in a period of development and piloting of measures. One of these boards had contracted Science Research Associates to construct such tests. Another was in the process of constructing and modifying a variety of criterion-referenced measures. The third board had developed a number of measures, but continues to use many commercial tests as well.

The boards primarily involved in the local preparation of tests used a mixture of norm-referenced and criterion-referenced measures. One board described its tests as criterion-referenced; another had tests which combined the characteristics of both. The third board was unique in having developed a self-concept scale which, along with a variety of attitude scales, was in widespread use across the system. All of these three boards were heavily oriented towards further local test development.

4. How are results used in the system?

Results of evaluation programs serve a variety of purposes. They are used on a system level to evaluate the functioning of an entire board by comparing achievement to available norms. They can be used to identify problem schools or areas, or to indicate difficulties in program or instruction at a school or individual teacher level. Results are also used to diagnose students' strengths or weaknesses and to prescribe remedial or enrichment programs. In the boards examined, these various uses were frequently mentioned, though the degree or relative extent of these uses appeared to differ widely.

At the administrative level, results were perceived as providing a check on system-wide functioning, and on student achievement in general. The board's program is evaluated in this manner, and schools in need of special assistance identified. The measurement specialists used results as a basis for refining test instruments, for designing workshops, and for providing consultative services for principals and teachers.

It appears from the interview data that most boards use results in a variety of ways at the school level. Tests were seen as a useful adjunct to teachers' classroom evaluations. Teachers and principals seemed to appreciate having objective measures against which to compare their impressions. Teachers and principals tended to use results for grouping and placement and, to a lesser extent, for program changes. Results of most tests cannot be used for specific diagnostic purposes, but they do serve as

a signal at times to help identify children in need of either enrichment or remediation.

Some teachers felt that, because of their own lack of training, the results were not as well used as they might be. There was also a concern expressed on some boards that results might be used for judging and evaluating teachers and principals without taking into account specific characteristics of students and localities.

There was a tendency among boards oriented towards norm-referenced testing to use results for system and school evaluation. The boards oriented towards criterion-referenced testing tended to perceive the main usefulness of their programs as providing information for individual diagnostic purposes. One of the boards stressing locally prepared tests saw the main use of its tests to be for diagnostic purposes, while another used its locally prepared tests to keep track of local program and curriculum. The third used results for a wide range of purposes, including system-wide checks, assessment of individual school achievement, and evaluation of a number of ongoing curriculum projects.

Results are not communicated directly to parents unless requested, and then in a general manner in the parent-teacher interview situation. Two main reasons were given for this. The first reason concerned parents' perceived inability to understand and appreciate the meaning of raw results. The second reason given was that parents would possibly read too much meaning into the scores. Personnel at all levels stressed the fact that test results are but one part of the process of evaluating education; it was thought that parents might not appreciate this fact. The resistance to communicating results of standardized tests to parents was found among all levels of personnel in all types of boards. However, not surprisingly, there appeared to be a general desire among the parents interviewed that results be communicated to them.

5. *What is the perceived role of the ministry in evaluation programs?*

In most interviews, reservations were expressed regarding the possible implementation of a ministry-sponsored province-wide testing program. It was felt that such a program would lead to a standardization and mechanization of education. It was also felt that results from a province-wide program might be interpreted at face value, without taking into account diversities of student populations.

There were some cogent arguments expressed in support of a ministry-sponsored program, however. One primary teacher contended that a single carefully designed program would avoid wasteful duplication of efforts and resources. A number of teachers and parents stated that the highly transient nature of much of Ontario's population would make such province-wide evaluations valuable, since standards would be available.

There was a general request for financial assistance from the ministry to support development of local evaluation programs. This request came particularly from those boards that have already expended effort in the development of local programs. These innovative boards also request technical assistance. A recurring plea is for the creation of an item bank which would allow for the selection and norming of items. Some boards would also like broad information concerning available tests, their uses and purposes, and other related information.

SUMMARY

It was seen that a move towards board-wide evaluation through use of tests occurred in response to three influences: a need to provide accountability data to external clients and critics, a desire to measure local curriculum objectives, and a wish for diagnostic-prescriptive data at the individual level. Norm-referenced measures appear to be the typical starting point and foundation of system-wide programs. Results from these measures provide understandable comparative data that can be used for accountability purposes. Since they tend to test the "basics" and the "basics" are part of all local curricula, they do test, at least in part, the local program. Finally, results serve to designate in a rather global manner the child's strengths and weaknesses.

Some boards, however, were not satisfied with the degree to which these measures were related to curriculum objectives, and this dissatisfaction can be seen as responsible for the development, in all boards interviewed, of locally prepared measures. In some cases, a board may let out a contract for assistance in the development of new test instruments.

No board indicated that there would be a move towards more norm-referenced standardized testing. Most boards felt that there would be an increased emphasis on local tests that measure local curriculum objectives. One factor in the degree to which boards will be able to achieve these goals will be the resources and technical assistance provided by the ministry. Across all interviews there were requests for both technical assistance and funds for evaluation purposes.

One final but pertinent comment: development of evaluation programs is a costly procedure in terms of both time and money. Boards which can afford to maintain well-staffed evaluation sections appear to have developed and to be using more sophisticated measures. Extra funds might well be required if a provincial mandate were to require that boards conduct student evaluation programs beyond those currently being administered. One of the benefits weighed against cost in any evaluation program appears to be what might be termed a "reassurance" factor. One may recall, for example, the concern expressed when teachers were perceived as no longer teaching the basic skills because time was being expended in a variety of educational activities. A core curriculum and

more structured guidelines are now perceived to be emphasized. It is against such factors as student and teacher morale, parent satisfaction, public reaction, and calls for accountability that the cost of evaluation programs must be weighed.

REFERENCE NOTE

- ¹ Wells, Hon. Thomas L. Speech to the Ontario Legislature, November 26, 1976.

REFERENCES

- Stufflebeam, D. (Chair), *Educational evaluation and decision making*. Phi Delta Kappa National Study Committee on Evaluation. Itasca, Illinois: F. E. Peacock Publishing, 1971.
- Wahlstrom, M. W., Danley, R. R., & Raphael, D. *Measuring achievement at the primary and junior levels: An analysis of tests and practices in Ontario*. Toronto: Ontario Ministry of Education, 1977.

Raymond Danley is a Research Officer in the Department of Curriculum at the Ontario Institute for Studies in Education, Toronto, Ontario M5S 1V6.

Merl Wahlstrom is an Associate Professor in the Department of Curriculum and Measurement, Evaluation and Computer Applications at the Ontario Institute for Studies in Education.

Helen Hayes is a doctoral student and Research Officer in the Department of Curriculum at the Ontario Institute for Studies in Education.

Dennis Raphael is an Associate Professor in the Faculty of Education at the University of Western Ontario, London, Ontario N6G 1G7.