

Research Note

A. Sidiq Ali

Peel District School Board, Mississauga

A Special Education Program Evaluation for Accountability Purposes: An In-Depth Case Study

Introduction

The Peel District School Board (PDSB), a large, ethnically and geographically diverse school district west of Toronto, has a specific special education program for students with learning disabilities. The Communication Program provides specialized instruction to students who are able to learn how to read and write to a level where integration back into regular classes (with some support) is feasible. After five years of program operation, a formative evaluation was started.

Context

The evaluation of the Communication Program serves as the first in-depth evaluation of a special education program in the PDSB's new Special Education Accountability Framework (Ali & Favaro, 2004). The accountability framework is built on the premise that successful program development cannot occur without program evaluation (Sanders, 2000). A typical program evaluation involves the systematic collection and analysis of information about program activities, characteristics, and outcomes, to improve program effectiveness and/or inform decisions about future programming (Patton, 1997).

In the accountability framework, however, this evaluation also serves to introduce to the PDSB a performance measurement system focused on student achievement. The performance measurement system was to be established during the evaluation and to continue after its completion. Use of program evaluation for such purposes is not uncommon in business (McDavid & Hawthorn, 2006), but here new ground was being broken in our field and organization.

Following principles of successful accountability as put forth in O'Day (2002) and Baker and Linn (2002), accountability data, performance-related or not, must be the basis for program improvement. To facilitate this type of data use, another key to the success of the performance measurement system for the purposes of the evaluation and beyond was the buy-in from program managers—in this case, school administrators and teachers—as they must view the performance measurement as nonthreatening, in their control, and for their use in program improvement (McDavid & Hawthorn, 2006). Hence a standardized test of student reading achievement was used for evaluation and

A. Sidiq Ali is the Assessment and Accountability Officer for Special Education Support Services. He can be reached at sid.ali@peelsb.com.

accountability. Data from this performance measure, however, comprised only a portion of the evaluation data.

A mixed method (Teddlie & Tashakkori, 2002) research design is employed for the evaluation. Building from over half a decade of archival data; two years of student-level quantitative achievement data; affective data from administrators, teachers, students, and parents; semistructured focus group interviews; document analyses and observations of key program personnel, this evaluation examines the congruency of policies, procedures, and programs, while providing academic profile status of learners in the program over a three-year period. Analyses proposed for this evaluation are presented in the context of accountability: information produced by this evaluation will be useful in describing, modifying, and refining the program. For the purpose of these research notes, however, particular focus is on the student-level quantitative achievement data.

Method

Because basic reading skill development is the main focus of the Communication Program, a standardized test, the Kaufman Test of Educational Achievement (2nd ed., Kaufman & Kaufman, 2004, KTEA-II), which contains measures of reading achievement, was the instrument chosen to help gauge students' performance in reading and reading-related skills. Information about students' reading and related skills may help inform policy decisions regarding the delivery of the Communication Program, especially in the area of instructional design.

For the purposes of the evaluation, the first administration of the KTEA-II occurred in late fall 2004. A second administration of the KTEA-II took place in the late fall of 2005. Beyond the evaluation, the performance testing will occur every fall.

Participants

Students in the Communication Program in 2004 ($N=518$) and students in the Communication Program in 2005 ($N=522$) are the participants. Many of these students from the first year to the next are the same, yet some students have left the program and some have joined it over the two years.

Analysis

Because the performance measurement data are to be used for accountability purposes, a quasi-longitudinal (Linn, 2004) approach will be taken when analyzing and presenting scores from the KTEA-II. In this accountability model, gains in the mean performance of all tested students from one year to the next, for example, the mean test score for students in 2005 minus the mean test score for students in 2004, are the focus. Including multiple years in the analysis of performance data has the advantage of reducing variability due to measurement and sampling error by increasing the number of students in the analyses. Further, this model does not depend on the comparability of successive groups of students, which has been shown to be problematic (Linn). In basic research design terms, analysis of the KTEA-II data for the evaluation and beyond will follow a trend analysis, where samples from a population whose members may change are tested at various times (Fraenkel & Wallen, 2006).

We must, however, be careful not to attribute gains or the lack thereof solely to the program, as a number of factors such as socioeconomic status, which are beyond the influence of the teachers, are at play when examining students' performance. So a multilevel regression model (Bock, 1989) will be tested with hierarchical linear modeling (Raudenbush & Bryk, 2002) using the KTEA-II gain scores as the outcome measure and predictors will include measures of school-level socioeconomic status (Peel District School Board, 2004).

In addition, relationships between KTEA-II subtests will provide a profile of the types of learners in the program. Correlations between subtests will be evaluated using the effect size guidelines in Cohen and Cohen (1983). Further, relationships between reading skill achievement and affective characteristics of students such as their reading attitudes and self-efficacy may also inform program development.

Discussion

The key to making the evaluation a success in terms of introducing a performance measurement system for accountability purposes is obtaining teacher and administrator buy-in of the system, but also ensuring that data are reliable and validly interpreted so that policymakers and practitioners may act accordingly. The outlined measurement and analysis endeavor to provide reliable data for valid interpretation.

References

- Ali, A.S., & Favaro, P. (2004). *A special education accountability framework*. Paper presented at the annual conference of the Canadian Society for Studies in Education, Winnipeg.
- Baker, E.L., & Linn, R.L. (2002). *Validity issues for accountability systems*. CSE Technical Report 585. Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing.
- Bock, D.E. (Ed.). (1989). *Multilevel analysis of educational data*. Toronto, ON: Academic Press.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analyses for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Fraenkel, J.R., & Wallen, N.E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw-Hill.
- Kaufman, A.S., & Kaufman, N.L. (2004). *Kaufman Test of Educational Achievement* (2nd ed.). Circle Pines, MN: AGS Publishing.
- Linn, R.L. (2004). Accountability models. In S.H. Fuhman & R.F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 73-95). New York: Teachers College Press.
- McDavid, J.C., & Hawthorn, L.R. (2006). *Program evaluation and performance measurement: An introduction to practice*. Thousand Oaks, CA: Sage.
- O'Day, J. (2002). Complexity, accountability, and school improvement. *Harvard Educational Review*, 72, 293-329.
- Patton, M.Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: Sage.
- Peel District School Board. (2004). *Pathway schools: Elementary and secondary social risk index*. Mississauga, ON: Author.
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models for social and behavioral research: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Sanders, J.R. (2000). *Evaluating school programs* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Teddlie, C., & Tashakkori, A. (Eds.). (2002). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.