

Ross, J. A. (2004). Effects of running records assessment on early literacy achievement: Results of a controlled experiment. *Journal of Educational Research*, 97(4), 186-194.

Effects of Running Records Assessment on Early Literacy Achievement:
Results of a Controlled Experiment

John A. Ross

Ontario Institute for Studies in Education

University of Toronto

Corresponding author:
Dr. John A. Ross,
Professor and Field Centre Head,
OISE/UT Trent Valley Centre,
Box 719, Peterborough, ON K9J 7A1
jross@oise.utoronto.ca

Effects on Early Literacy Achievement of Running Records Assessment:

Results of a Controlled Experiment

Abstract: Recent research on effective schools (e.g., Pressley et al., 2001) identified consistent associations between students' literacy achievement and teacher practice. This study extended these correlational findings by conducting a controlled experiment to test the claims about one practice recommended by recent effective schools research, systematic classroom assessment, represented here as the use of running records to plan instruction. Schools assigned to the Running Records treatment outperformed schools assigned to a near-treatment condition (Action Research). After controlling for prior school achievement and collective teacher efficacy, the Running Records intervention accounted for 12% of the between-school variance in Reading and 7% in Writing, confirming the correlational finding from effective schools research.

Effects of Running Records Assessment on Early Literacy Achievement:
Results of a Controlled Experiment¹

Research on effective schools/teachers, in its heyday in the 1970s and 1980s, is resurging, particularly in the field of early literacy. The rich generalizations emerging from these correlational studies suggest a variety of instructional improvements that might plausibly contribute to higher student achievement. In this study we examined the student achievement effects of one of them, i.e., the finding that teachers in effective schools are more likely than teachers in ordinary schools to use classroom assessments like running records to diagnose student needs and monitor progress. The intervention was assessed in a controlled experiment involving all the schools with grade 3 students in one school district.

Motivation for the Study

Schools and Teachers with Effective Literacy Programs

In the 1970s and 1980s, researchers identified a cluster of practices (e.g., frequent monitoring of student progress) that consistently predicted higher levels of student literacy. This research described correlations between student achievement and school practices without providing any clues as to how effective schools got that way. It was not clear what consumers of the research should do with the results. Agents of school improvement overgeneralized the findings, transforming moderately positive correlations into one to one correspondences of teaching strategies with student outcomes. Training programs that used transmission strategies to implement the findings (e.g., Hunter, 1982) produced mixed results at best (e.g., Azumi & Mitchel, 1989; Corbitt, 1989). Interest in the correlates of achievement tended to fade, although Doyle (1987) argued that lists of effective teaching techniques have value in that they act as stimulants for teacher reflection and as topics in teachers' professional discourse.

In the 1990s researchers returned to the process-product paradigm with important changes in their approach: qualitative procedures involving relatively small samples of high achieving and ordinary schools replaced large scale quantitative studies; data were interpreted within constructivist perspectives; researchers described instructional practices in much greater detail and clustered these practices within broad models of teaching, drawing especially on balanced literacy models that combine skills development (e.g., decoding strategies) with authentic, integrated Reading and Writing activities. For example, Pressley et al. (2001) identified 103 behaviors that distinguished exemplary from average Reading teachers in 15 paired comparisons of grade 1 teachers. Pressley et al. further distinguished 7 categories of action that distinguished the five best from the five poorest teachers.

Both the earlier era of effective schools research (reviewed in Hoffman, 1991) and the more recent instances highlight assessment practices as contributors to student achievement in Reading and Writing. In these studies the key element is classroom assessment as opposed to external assessments. Pressley et al. (2001) found that the most effective teachers constructed running records (as described in Reading Recovery) during one-to-one Reading. Taylor, Pearson, Clark, and Walpole (2000) found that of 14 schools studied, the moderately effective schools operated classroom-based assessment systems to monitor individual student progress and shape instruction. Classroom-level performance results were shared with the principal and other teachers. The most effective schools in Taylor et al. shared assessment system data 3-5 times per year. Frequent classroom assessment was also a key feature in the effective literacy schools studied by Maine (2000), Matsumura, Patthey-Chavez, Valdes, and Garnier (2002), Wharton-McDonald et al. (1997), and Wray, Medwell, Fox, and Poulson (2000).

Running Records

Among the many approaches to classroom assessment, running records are of particular interest. A running record is a literacy assessment technique that is formative, i.e., it provides information that can be used to improve students' Reading. When the teacher administers a running record the student sits or stands beside the teacher so that both can see the text. As the child reads, the teacher codes each word, reporting the percentage of words correctly read, the self-correction ratio (the ratio of errors + self-corrections divided by the number of self-corrections), and the categories of errors made (meaning, visual, or structure). After the reading the student retells the story and answers questions about the story's meaning. A running record is successfully completed when the student has correctly responded to the questions about meaning and has read at 90-94% accuracy. The level of the passage read and the types of errors made by the student guide subsequent instruction (Clay, 1993).

Although running records are frequently used in research, there is little psychometric data available about the procedure. Evidence about reliability is mixed. For example, substantial discrepancies have been reported between the book levels measured by Reading Recovery teachers and the book levels subsequently calculated for the same students by their regular classroom teachers using the same procedures (Chapman, Tunmer, & Prochnow, 2001; Glynn, Crooks, Bethune, Ballard, & Smith, 1989). Other researchers report adequate test-retest reliability (Clay, 1993) or acceptable internal consistency (e.g., Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994). Validity evidence is abundant: running records correlate with other early literacy measures and discriminate among treatment conditions (e.g., Pinnell et al., 1994). The consequential validity of running records (e.g., the extent to which the assessment procedure contributes to higher achievement) has not been addressed because the effects of the assessment

have not been disentangled from the instructional treatments (such as Reading Recovery) in which the assessment is embedded.

Running records assessment is recommended by national curriculum authorities, (e.g., in New Zealand: Limbrick, 1999) and is the method of choice of many specialist teachers (Bean, Cassidy, Grumet, Shelton, & Wallis, 2002 found that 62% of the members of the International Reading Association who identified themselves as Reading teachers use it). Given the consistent evidence that the use of formative classroom assessment like running records distinguishes exemplary from ordinary teachers can we assume that promotion of the technique will improve student achievement? Taylor et al. (2000) suggest caution:

When all is said and done, we are examining natural correlations between program and teaching factors on the one hand and student performance on the other. These correlations...cannot be used to identify causes for improvements (or decrements) in student achievement. For that, more systematic experimentation is needed, including control groups, randomization, and careful analyses of growth over time. (p. 160)

The purpose of our study was to conduct a controlled experiment in which a sample of schools implemented running records as a strategy for aligning literacy instruction with students' need. We compared the provincial test scores of the running records schools to schools that implemented an alternative school improvement strategy (action research).

Nature of the Running Records Intervention

The running records treatment had two key components: teacher in-service and principal support. Teacher in-service consisted of six 60-minute after-school workshops on how to administer running records and use the information to focus instruction on individual student needs. Sessions were held at each school by literacy teachers (i.e., regular classroom teachers

seconded for the in-service role). Procedures were demonstrated using videotapes of typical children and teachers practiced the running records procedure, comparing their assessments to those of the in-service leaders and teachers.

Previous research on attempts to change assessment practices (e.g., Briscoe, 1993) indicates that even highly motivated teachers provided with ample information on new assessment methods have difficulty implementing the change and may revert to their previous procedures (Briscoe, 1993). Black and Williams (1998), in a review of 250 studies of teachers' assessment practices published since 1988, found that teacher use of assessment for formative purposes (i.e., to obtain information to modify instruction) was much less frequent than assessment for other purposes, most notably accountability. Changing assessment practices is difficult because assessment is tightly bound to deep-seated teacher beliefs about evidence, student motivation, and instructional effectiveness. To support teachers in the use of running records we worked with principals to create a collaborative culture of literacy assessment.

The in-service sessions for principals (3 sessions X 0.5 days) provided principals with specific knowledge on literacy assessment (a demonstration of how to conduct a running record with opportunities to practice the technique) and information on how to support teacher change based on a conception of transformational leadership (Bass, 1985; Bass & Avolio, 1994; Podsakoff, MacKenzie, Moorman, & Fetter, 1990). We defined transformational leadership as dedication to fostering the growth of organizational members and enhancing their commitment by elevating their goals (Burns, 1978). We followed Leithwood, Jantzi, and Steinbach (1999) in recommending six dimensions of transformational leadership relevant to schools: symbolizing good professional practice, providing individualized support, providing intellectual stimulation, holding high performance expectations, fostering a vision and collaborative decision making.

The in-service identified specific ways in which principals could apply each of these dimensions to the running records intervention. We adopted a transformational approach to leadership because the district was committed to the approach and because of evidence of the effects of transformational leadership on student achievement (Verona & Young, 2001) and on organizational learning, effectiveness, and culture (Authors, 2003; Leithwood et al., 1999). Organizational theorists attribute these effects to social identification, which enables followers to transcend their self-interests for the good of the group (Bass, 1985; Bass & Avolio, 1994; Leithwood, 1993; Walumbwa, Wang, & Lawler, 2003).

The specific actions taken by principals in the project were to identify at least two teachers who would contribute classroom data for the project, secure resources (e.g., purchase leveled Reading materials), facilitate teacher access to the in-service, coordinate the data collection in the school (two rounds of running records), and reflect on the results. Principals were encouraged to devise their own strategies for sampling their teachers, e.g., selecting teachers who provided leadership in literacy teaching in their school or teachers who might particularly benefit from close interaction with the principal on literacy teaching. Each teacher selected a sample of students (usually 5) for whom running records data would be submitted. Teachers used two broad selection strategies: a representative sample of students in the class or the five neediest children who did not qualify for the district's early literacy intervention. Principals were encouraged to involve all their teachers in the project and virtually all teachers in all schools attended the in-service activities. Teachers administered running records to the student sample on two occasions, before and after the focused Reading intervention. The assessments were 6-8 weeks apart. Principals collected summaries of the running records administered by teachers in the sample and submitted them to university faculty analyzing the

data on behalf of the district. Each school received from us a report, confidential to the school, that summarized each school's results and outlined the pre-post differences across the district. There was no ranking of schools and the school reports urged schools not to compare their results to the district as a whole because the sampling procedure was not standardized. Principals discussed the results with their teachers and with other principals in their family of school groups

Research Hypothesis

We predicted that schools assigned to the running records treatment would have higher student achievement in Reading and Writing than schools assigned to an alternate treatment.

We designed a controlled experiment in which schools were randomly assigned to two school improvement strategies. To ensure a fair comparison we used scores from the mandated provincial assessment as the dependent variables and we included in the comparison prior school achievement and a school variable, collective teacher efficacy, which is a powerful predictor of achievement. We also tested the credibility of the claim that the running records intervention had an achievement effect by examining an alternate explanation for any differences that might appear.

Method

Sample

Supervisory officers in a single Ontario, Canada school district randomly assigned schools in their areas to two treatments: Running Records and Action Research. Violations of randomness occurred when a few principals persuaded their supervisor to place them in a different treatment condition than the one to which they had been assigned. Schools were included in this study if: 1) they submitted a summary of their project to the collection of Action Research reports published in June, 2002 (Hannay, Telford, & Bray, 2002) OR submitted

running records data to the Running Records treatment database compiled at the same time (Ross & Hogaboam, 2002); 2) they had students who wrote the grade 3 Reading and Writing assessments in May, 2002; and 3) their teachers completed a survey measuring collective teacher efficacy in May, 2002. Seventy-three of the 75 K-6/8 schools in the district met the criteria, producing a sample consisting of 39 Running Records schools and 34 Action Research schools. Of the 2800 students in the study, 3% were classified as English as a Second Language, 20% were receiving special education support, and 6% were exempted from the provincial assessment. The mean family income of the district was US\$35,040. No data were available on racial distribution.

Sources of Data

Student achievement consisted of school scores for grade 3 Reading, Writing, and Mathematics collected in May of 2001 (*prior achievement*, a covariate) and 2002 (*achievement*, the dependent variable). The assessment was administered by teachers following the directions of an independent testing organization for two hours per day for five days. Reported for each school were the percentages of students who met the provincial standard of level 3 or 4, based on the global score for method 1 (all students included). The global score was calculated by the testing agency by combining the raw scores for the performance assessment and the multiple-choice items, with a correction factor to equate scores from one year to the next (Education Quality and Accountability Office, n.d.).

Collective teacher efficacy (i.e., teachers' belief that the teachers in their school constitute an effective instructional team) was included as an additional covariate because it is a strong predictor of school achievement (Bandura, 1993; Goddard, 2001; 2002b; Goddard & Goddard, 2001; Goddard, Hoy, & Hoy, 2000; Goddard, Hoy, & LoGerfo, 2003). It consisted of 14 items

reflecting two dimensions of collective teacher efficacy: the 7 items with the highest loading on the perceptions of the task factor.(e.g., “Drug and alcohol abuse in the community make learning difficult for students here”) and the 7 with the highest loading on the perceptions of teaching competence factor (e.g., “Teachers in this school really believe every child can learn”), reported by Goddard et al. (2000). Although the two-factor structure of the variable was maintained for face validity reasons, the two factors are highly correlated and, as in previous research, we combined the items into single scale. The items were in Likert format with a 6-point response scale anchored by strongly disagree and strongly agree.

Experimental Conditions

The Running Records treatment was as described above. In the Action Research treatment principals, in collaboration with a team of teachers, identified a professional issue related to Early Literacy to work on, devised improvement activities, implemented the action plans, reflected on and reported the results. There were four in-service sessions for principals (one full day and three half-day) in the Action Research treatment. In each session principals met with peers from the same family of schools, except that in the third session they attended with 3-4 teachers who were on their school improvement team. The in-service worked through a multi-phase process consisting of identifying a measurable goal, designing action plans, collecting and analyzing data on the effects of the intervention, and reflecting on the results. Teachers in the Action Research treatment had access to the same set of teacher in-service sessions as the teachers in the Running Records schools and had the same budget for purchasing early literacy resources.

The specific projects undertaken in the Action Research schools varied across sites. For example, one school developed the action research question, “how can we use response journals

to improve the communication strand of literacy?” Teachers implemented response journals along with the 3 R’s strategy (retell, relate, reflect). A systematic sample (every fifth student) was drawn from each K-3 class for the school’s data collection. Teachers used the rubric of the provincial testing organization and provincial exemplars to assess one piece of student writing at the beginning of the project and a second piece produced six weeks later. Other schools focused on questions less directly focused on Early Literacy, for example, “are students transferring their knowledge about writing to mathematics?” The school improvement strategy consisted of holding Mathematics workshops for all teachers and providing a collection of readings. Teachers were asked to select and try out a strategy and report the results to a staff meeting. Teachers also identified one student in each class at each of four levels of performance and had those students complete a mathematics performance assessment that was marked by the principal and vice-principal using the provincial Mathematics rubric (which contained a dimension referring to communication of mathematical ideas).

Analysis Procedures

We used the General Linear Model (GLM) program of SPSS to conduct a multivariate analysis of covariance. The dependent variables were the percentages of students in each school meeting the provincial standard (level 3 or better) in Reading and Writing in 2002. Included as covariates were two predictors of achievement, prior achievement and collective teacher efficacy. The independent variable was assignment to the Action Research or Running Records treatment. The unit of analysis was the school. As a further test of our hypothesis, we also examined whether there were any treatment differences in 2002 Mathematics achievement, controlling for prior mathematics achievement and collective teacher efficacy. If we were to find achievement differences in Mathematics as well as in literacy it might suggest that some factor

other than the Running Records treatment was contributing to the differences between the two groups of schools.

Results

Table 1 displays the means and standard deviations for both groups of schools for each of the two years for all three subjects. Despite the violations to randomization during the sample assignment procedure, there were no significant differences between the two groups of schools in prior achievement in Reading [$t(71)=.934$, $p=.354$], Writing [$t(71)=-.027$, $p=.978$] or Mathematics [$t(71)=1.796$, $p=.077$]. There were also no differences in collective teacher efficacy [$t(71)=-.821$, $p=.415$].

Table 1 About Here

Reading and Writing scores were highly correlated within years ($r=.79$, $N=73$, $p<.01$ in 2001 and $r=.74$ in 2002), suggesting that a multivariate analysis was more appropriate than analyzing the effects of treatment assignment separately for each subject. The correlations were lower within subjects between years: $r=.48$ for Reading and $r=.56$ for Writing but were statistically significant, suggesting that 2001 scores should be used as covariates in the analysis. (The correlation matrix for all the variables in the study is shown in the Appendix, Table 4.)

Since 2001 Reading and 2001 Writing were highly correlated ($r=.79$) we anticipated that only one of these variables would be a significant predictor of 2002 scores. Since the correlations of 2001 Writing with 2002 achievement were slightly stronger than the correlations of 2001 Reading with 2002 achievement, we anticipated that 2001 Writing would enter the equation first and take all the variance shared by the two covariates. Preliminary analysis (not reported) indicated that this was case: 2001 Writing but not 2001 Reading was a significant predictor of

2002 Reading and Writing achievement. We decided to use only 2001 Writing as a covariate in the analysis to avoid wasting a degree of freedom.

The multivariate test with Reading and Writing combined explained 56.5% of the variance in achievement. Student achievement in 2002 was significantly influenced by prior achievement in Writing [$F(2,68)=15.05$, $p<.001$, $\eta^2=.307$], by collective teacher efficacy [$F(2,68)=3.75$, $p=.028$, $\eta^2=.099$] and by treatment [$F(2,68)=4.827$, $p=.011$, $\eta^2=.124$].

Table 2 displays the univariate effects. The results were virtually the same for both Reading and Writing. Collective teacher efficacy and prior achievement were each significant predictors of 2002 achievement. Table 2 shows that after the effects of these covariates had been removed, treatment had a significant effect on 2002 student achievement. Schools in the Running Records treatment outperformed schools in Action Research. There was a large effect for Reading (12% of the variance) and a medium effect for Writing (7% of the variance), using Cohen's (1988) rules for interpreting effect sizes.

Table 2 About Here

We examined the effects of treatment on Mathematics achievement because if we found treatment differences in achievement for another subject it would suggest that there was something other than or in addition to the literacy activities contributing to the changes in Reading and Writing scores. We found that 2002 Mathematics achievement was significantly influenced by 2001 Mathematics achievement [$F(1,69)=19.540$, $p<.001$, $\eta^2=.221$] and by collective teacher efficacy [$F(1,69)=5.563$, $p=.021$, $\eta^2=.075$] but there were no treatment effects [$F(1,69)=2.864$, $p=.095$, $\eta^2=.040$]. A further test of alternative explanations would be to examine changes in school processes but we had only one year of data. We found there were no significant differences between the groups of schools in collective teacher efficacy in 2002

[$t(71)=-.821$, $p=.415$] but we have no way of knowing whether the two groups were equivalent prior to the assignment.

These results indicate that participation in the Running Records treatment had a greater positive effect on achievement in Reading and Writing than participation in the Action Research condition. Table 3 shows the district-province achievement gap. The achievement gap was obtained by subtracting the percentage of district students at the provincial standard from the percentage reaching that standard in the province as a whole. The provincial average for 2002 (50% in Reading and 56% in Writing) was used for both years in order to facilitate comparisons. The table shows that the 39 schools that were in the Running Records treatment were below the provincial average in 2001 in both Reading and Writing. In 2002 the same schools exceeded the provincial average in Reading and had reduced the gap in Writing by two-thirds. The 34 Action Research schools were also below the provincial average in both subjects in 2001. However, these schools fell further behind the province in both Reading and Writing in 2002.

Table 3 About Here

Discussion

The main contribution of this study is that it confirmed in a controlled experiment a key correlational finding from recent qualitative research comparing high achieving literacy sites to ordinary schools. The effective schools studies consistently report positive associations between high student literacy and engagement in systematic classroom assessment procedures, particularly running records (Maine, 2000; Matsumura et al., 2002; Pressley et al., 2001; Taylor et al, 2000; Wharton-McDonald et al., 1997; Wray et al. 2000). In a correlational study, causality cannot be demonstrated. Researchers have no way of knowing whether particular assessment practices lead to achievement, whether achievement encourages certain kinds of assessment,

whether the two are the product of some other factor, or whether assessment and achievement are substantively related at all. In this study schools that were assigned to the Running Records treatment improved their Reading and Writing achievement and outperformed schools from the same district assigned to a near-treatment condition. By manipulating the conditions to which schools were assigned, controlled experimentation greatly strengthens the credibility of the claim that schools that implement systematic classroom assessment will have higher literacy achievement. The validity of the causal argument is enhanced by methodological features of the study: the demonstration that schools were equivalent on relevant variables prior to the intervention; the selection of the school as the unit of analysis; the use of appropriate covariates identified in previous research as predictors of achievement; the use of multivariate procedures to avoid inflating Type I error; the search for disconfirming evidence by testing for effects on Mathematics; the use of a near-treatment alternative rather than a no-treatment control as the comparison group; the six week duration of the study; and the use of an external assessment for the criterion variables. The study met all of Slavin's (1987) requirements for inclusion of a study in a best-evidence synthesis.

All of the Action Research projects gave explicit attention to the measurement of the effects of project activities and several contained substantial student assessment elements that appeared to be similar to measurement in the Running Records treatment. For example, many of the Action Research projects had teachers applying provincial assessment rubrics to student writing. But these rubrics simply classify student writing—they do not provide specific strategies for moving students from one level to another. In contrast the Running Records treatment met the four criteria identified by Black and William (1998) for an effective feedback system: Running Records provide (a) data on the actual level of a measurable attribute (book level

accurately read), (b) the reference level of the attribute (there are 30 book levels), (c) mechanisms for comparing the actual performance to a meaningful scale and generating information about the nature of the gap (the First Steps continuum (Commonwealth of Australia, 2003) relates running records data to specific Reading competencies), and (d) a mechanism by which the information can be used to alter the gap (identification of appropriate books for the child to read and specific reading skills to be learned).

The plans of the Running Records schools were focused on a relatively narrow set of assessment and instructional activities. The narrow focus contrasts with the content of Ontario school improvement plans reviewed by Nagy (2000). Plans for improving scores on provincial assessments are required of all schools. Nagy found that large scale assessments provide general indicators of success that result in generalized action plans, plans that could easily be developed without reference to assessment data. At the other end of the scale, Mintrop and MacLellan (2002), in a review of Maryland schools placed on probation, found school improvement plans to be overwhelmed by minutiae—the average school listed 50 activities for the year. Most of the planned activities were incremental adjustments to existing school practices that skirted substantive change. The Running Records treatment avoided both the overly general and excessively particular. It focused teacher attention on a specific assessment procedure that accessed an elaborate but feasible approach to literacy instruction.

Limitations of the Study

In this study, the Running Records intervention was well-aligned with the actions of other agents. The goal of improving literacy matched the priorities of the province, the provincial assessment agency, the district, and the teachers themselves. Of particular importance was the leadership role of the principal in coordinating assessment activities in the school, including the

data that the school submitted to the district. The effective schools literature of the 1970s and 1980s highlighted the role of the principal in setting high expectations and holding staff accountable for meeting them; i.e., it promoted transactional leadership practices. In this study, principals were encouraged to adopt a transformational style, i.e., they attempted to develop the capacity of their staff, transforming the culture of the school. In this study we did not measure variations in principal actions, so we are unable to determine how much of the impact of the Running Records treatment was dependent upon principal actions. It could have been considerable because in previous research we found that transformational leadership contributed to collective teacher efficacy and to teacher commitment to professional norms (Ross & Gray, 2003), factors that are likely to have increased student achievement.

A second limitation of this study is that randomization procedures were violated in some cases by principals who petitioned their superintendent to place them in a different condition than the one to which they were assigned. We believe that these violations were few in number; we demonstrated that the groups were equivalent on two measures (prior achievement and collective teacher efficacy) that predict achievement; and we included both measures as covariates in the analysis.

Conclusions

The school effectiveness literature of the 1970s and '80s identified correlates of achievement that were robust across subjects, neighborhoods, and levels of schooling (see e.g., Reynolds & Teddlie, 2000). The production of this knowledge did not lead to improved schools because the research failed to translate the correlates into feasible interventions and did not address the conditions that needed to be in place for instructional improvement to occur. The study reported here suggests that the new wave of school effectiveness research is likely to be

more useful because it identifies through qualitative investigations specific behaviors occurring in naturalistic settings. This specificity enables school improvement researchers to construct ecologically valid interventions to test the external validity of particular findings. In this study implementation of running records, a classroom assessment system, contributed to higher achievement in Reading and Writing, as predicted by effective schools research. An essential element of the intervention was a support system for teachers developed in collaboration with principals working from a transformational leadership stance and coordinating their actions with other change agents. The results of this study leave us optimistic that weaving the findings of school effectiveness research into viable change processes will lead to the improvements in achievement of the kind that motivated effective schools researchers, funding agencies, and practitioners.

References

- Azumi, J., & Mitchel, L. (1989). *The impact of a district-wide staff development project on the attitudes and behaviours of teachers*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- Bass, B. (1985). *Leadership and performance beyond expectations*. New York, NY: The Free Press.
- Bass, B. M., & Avolio, B. J. (1994). *Improving organizational effectiveness through transformational leadership*. Thousand Oaks, CA: Sage.
- Bean, R. M., Cassidy, J., Grumet, J. E., Shelton, D. S., & Wallis, S. R. (2002). What do reading specialists do? Results from a national survey. *The Reading Teacher*, 55(8), 736-744.
- Briscoe, C. (1993). Using cognitive referents in making sense of teaching: A chemistry teacher's struggle to change assessment practices. *Journal of Research in Science Teaching*, 30(8), 971-987.
- Burns, J. M. (1978). *Leadership*. New York: Harper & Row.
- Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2001). Does success in the Reading Recovery program depend on developing proficiency in phonological-processing skills? A longitudinal study in a whole language instructional context. *Scientific Studies of Reading*, 5(2), 141-176.
- Clay, M. M. (1993). *An observation survey of early literacy achievement*. Auckland: Heinemann.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd edition. Hillsdale, NJ: Erlbaum.
- Corbitt, E. (1989, April). *The three R's of staff development: Reality, relevance, and relationships*. Paper presented at the annual convention of the Council for Exceptional

Children, San Francisco, CA.

Department of Education, Commonwealth of Australia (2003) *First Steps Reading*

Developmental Continuum [Web Page]. URL

http://www.myread.org/monitoring_first.htm [2003, June 26].

Doyle, W. (1987). Research on teaching effects as a resource for improving instruction. In M.

Wideen & I. Andrews (Eds.), *Staff development for school improvement: A focus on the teacher*. New York, NY: Falmer Press.

Education Quality and Accountability Office. (n.d.) *Explanatory paper: Grades 3 and 6*

assessment of reading, writing, and mathematics, 1999-2000 [Web Page]. URL

http://www.eqao.com/eqao/home_page/pdf_e/00/00P063e.pdf [2003, June 18].

Glynn, T., Crooks, T., Bethune, N., Ballard, K., & Smith, J. (1989). *Reading Recovery in*

context. Wellington, NZ: Department of Education.

Goddard, R. D., Hoy, W. K., & LoGerfo, L. (2003, April). *Collective teacher efficacy and*

student achievement in public high schools: A path analysis. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Goddard, R. D. (2001). Collective efficacy: A neglected construct in the study of schools and

student achievement. *Journal of Educational Psychology*, 93(3), 467-476.

Goddard, R. D. (2002b). A theoretical and empirical analysis of the measurement of collective

efficacy: The development of a short form. *Educational and Psychological Measurement*, 62(1), 97-110.

Goddard, R. D., & Goddard, Y. L. (2001). A multilevel analysis of the relationship between

teacher and collective efficacy in urban schools. *Teaching and Teacher Education*, 17, 807-818.

- Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Education Research Journal*, 37(2), 479-507.
- Hannay, L., Telford, C., & Bray, C. (2002). *Aligning school-district actions to promote school improvement and accountability*. Kitchener, ON: OISE/UT Trent Valley Centre.
- Hoffman, J. V. (1991). Teacher and school effects in learning to read. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research* (Vol. 2pp. 911-950). New York: Longman.
- Hunter, M. (1982). *Mastery teaching*. El Segundo, CA: TIP Publications.
- Leithwood, K. A. (1993, October). *Contributions of transformational leadership to school restructuring*. Paper presented at the Convention of the University Council for Educational Administration, Houston, TX.
- Leithwood, K. A., Jantzi, D., & Steinbach, R. (1999). *Changing leadership for changing times*. Buckingham: Open University Press.
- Limbrick, L. (1999, September). *The literacy debates: What are the issues in New Zealand?* Paper presented at the British Educational Research Association, Brighton, UK.
- Maine Department of Education Center for Inquiry on Literacy (2000). *A solid foundation: Supportive contexts for early literacy programs in Maine schools*. Augusta, Maine: author. (ERIC Document Reproduction Service No. ED 458 537)
- Matsumura, L. C., Patthey-Chavez, G. G., Valdes, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *Elementary School Journal*, 103(1), 3-26.
- Mintrop, H., & MacLellan, A. M. (2002). School improvement plans in elementary and middle

- schools n probation. *Elementary School Journal*, 102(4), 275-300.
- Nagy, P. (2000). The three roles of assessment: Gate keeping, accountability, and instructional diagnosis. *Canadian Journal of Education*, 25(4), 262-279.
- Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A. S., & Seltzer, M. (1994). Comparing instructional models for the literacy education of high-risk first graders. *Reading Research Quarterly*, 29(1), 9-39.
- Podsakoff, P., MacKenzie, S., Moorman, R., & Fetter, R. (1990). Transformational leaders' behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *Leadership Quarterly*, 1(2), 107-142.
- Pressley, M., & et al. (2001). A study of effective first-grade literacy instruction. *Scientific Studies of Reading*, 5(1), 35-58.
- Reynolds, D., & Teddlie, C. (2000). The future agenda for school effectiveness research. In In C. Teddlie, & D. Reynolds (Eds.), *The international handbook of school effectiveness research* (pp. 322-343). London: Falmer.
- Ross, J.A. & Gray, P. (2003) *Transformational leadership and teacher commitment to organizational values: The mediating effects of collective teacher efficacy*. Peterborough, ON: OISE/UT Trent Valley Centre.
- Ross, J. A., & Hogaboam-Gray, A. (2002). *School improvement capacity building in Kawartha Pine Ridge DSB: 2002 quantitative report*. Peterborough, ON: OISE/UT Trent Valley Centre.
- Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal*, 101(2), 121-166.

- Verona, G. S., & Young, J. W. (2001, April). *The Influence of principal transformational leadership style on High School Proficiency Test results in New Jersey comprehensive and vocational-technical high schools*. Paper presented at the annual meeting of the American Educational Research Association, Seattle.
- Walumbwa, F. O., Wang, P., & Lawler, J. J. (2003) *Exploring new frontiers: The role of collective efficacy in the relations between transformational leadership and work-related attitudes* [Web Page]. URL
<http://cobacourses.creighton.edu//MAM/2003/papers/walumba.doc> [2003, May 6].
- Wharton-McDonald, R., Pressley, M., Rankin, J., Mistretta, J., Yokoi, L., & Ettenberger, S. (1997). Effective primary-grades literacy instruction=Balanced literacy instruction. *The Reading Teacher*, 50(6), 518-521.
- Wray, D., Medwell, J., Fox, R., & Poulson, L. (2000). The teaching practices of effective teachers of literacy. *Educational Review*, 52(1), 75-84.

Table 1 Percentage of Students Meeting the Provincial Early Literacy Standard by Subject, Strand, and Year

	2001		2002	
	Mean	SD	Mean	SD
Reading				
Assessment Strand (n=39)	48.44	16.37	52.70	17.73
Action Research Strand (n=34)	45.03	14.56	42.60	18.55
Writing				
Assessment Strand (n=39)	48.67	16.56	53.48	15.83
Action Research Strand (n=34)	48.76	14.00	47.66	12.98
Mathematics				
Assessment Strand (n=39)	61.03	16.50	48.85	20.50
Action Research Strand (n=34)	53.38	19.85	58.95	20.64
Collective Teacher Efficacy				
Assessment Strand (n=39)	NA	NA	4.63	.34
Action Research Strand (n=34)	NA	NA	4.70	.38

Table 2 Effects of Strand Assignment on Grade 3 Reading and Writing, Controlling for Prior Achievement and Collective Teacher Efficacy: Univariate Effects

Source	Subjects	Significance Tests	Partial Eta Squared
Corrected Model	Reading	$F(3,69)=14.222, p<.001$.382
	Writing	$F(3,69)=16.009, p<.001$.410
Intercept	Reading	$F(1,72)=2.490, p=.119$.035
	Writing	$F(1,72)=1.168, p=.284$.017
Collective teacher efficacy	Reading	$F(1,72)=5.598, p=.021$.075
	Writing	$F(1,72)=6.398, p=.014$.085
Writing 2001	Reading	$F(1,72)=20.786, p<.001$.232
	Writing	$F(1,72)=26.915, p<.001$.281
Experimental Condition	Reading	$F(1,72)=7.163, p=.009$.092
	Writing	$F(1,72)=4.314, p=.041$.057

Table 3 District-Province Achievement Gap by Treatment Condition and by Year

	Achievement Gap	
	2001	2002
Reading		
Assessment Strand (n=39)	-2.17%	+2.7%
Action Research Strand (n=34)	-4.94%	-7.40%
Writing		
Assessment Strand (n=39)	-8.10%	-2.52%
Action Research Strand (n=34)	-6.97%	-8.34%

Appendix Table 4 Correlation Matrix of Variables in the Study (N=73 Schools)

	Reading 2001	Reading 2002	Writing 2001	Writing 2002	Math 2001	Math 2002
Reading 2001	--					
Reading 2002	.48***	--				
Writing 2001	.79***	.51***	--			
Writing 2002	.47**	.74***	.56***	--		
Math 2001	.89***	.52***	.75***	.45***	--	
Math 2002	.47***	.87***	.51***	.67***	.52***	--
CTE	.08	.33*	.27*	.36**	.17	.30*

* $p < .05$ ** $p < .01$ *** $p < .001$

CTE=Collective teacher efficacy

Endnotes

¹ The research was funded by the Social Sciences and Humanities Research Council of Canada and the Ontario Ministry of Education and Training. The views expressed in the report do not necessarily agree with those of the Council or Ministry. Peter Gray contributed to the data analysis. An earlier version of this article was presented at the annual meeting of the American Educational Research Association, April, 2004 in San Diego.