

Measuring the Effectiveness of Professional Development

Executive Summary of a Two-Year Pilot Study Conducted by the Schultz Center for Teaching and Leadership in conjunction with the Duval County Public Schools and the College of Education at the University of North Florida

January 2007

Introduction

This is a report of work at the Schultz Center for Teaching and Leadership to demonstrate the relationship between student achievement gains on the FCAT and the level of in-service training of their teachers. The focus of this work was to develop a model for evaluation that would stimulate improvements in professional development, enhance teacher performance, and increase student achievement. A second aspect of the study was to develop a design template which could be used to determine the return on financial investment made in delivering professional development. This summary reprints and borrows heavily from the December 2005 CEPRI Report: *In-Service Education: The Challenge of Determining Cost and Effectiveness*.

Background

Over the past four years, there has been a concentrated effort in Florida to provide more structure for in-service education. The Florida School Community Professional Development Act (*Florida Statute 1012.98*) required the Department of Education to design methods by which school districts may evaluate and improve their professional development systems. The evaluation process required by the new legislation is entitled the Professional Development System Evaluation Protocol. The system is based on standards established by the National Staff Development Council (NSDC) and is designed for use by the DOE during on-site visits in school districts, by district staff in preparation for site visits, and by school staff and school advisory councils for self-assessment. All school districts participate in reviews of their professional development systems by the Department of Education through a multi-year site visit schedule. Reviews began in March 2003. Once a school district has gone through a review, a number of resources are available to aid efforts to enhance professional development programs.

These recent initiatives provide a mechanism for assessing professional development programs. However a methodology for measuring the effectiveness of specific professional development programs in improving student learning continues to be elusive.

Development of an In-Service Education Evaluation System

Numerous factors limit the ability to assess the return on the state's investment in professional development of teachers and the coherence and resulting effectiveness of the programs. These factors include

- lack of records of the exact training received by each teacher;
- fragmentation of programs which may result in teachers receiving inconsistent or conflicting training;
- lack of coherent training programs which are internally consistent and which include all of the important elements identified in the literature and the U.S. DOE principles for effective professional development.; and
- lack of attempts to relate teacher training to student performance.
-

Essential Components of an In-Service Accountability System

The Schultz Center for Teaching and Leadership was established to provide a regional center for providing best practices training for teachers and principals in a state-of-the-art, high technology environment. The Center established a research-based training system that includes a systematic method for monitoring the type and amount of teacher in-service activities and the degree to which training is actually implemented in the classroom. With this unique framework, it was possible to start development of a system to quantifiably measure effectiveness of teacher training on student outcome.

The Schultz model has several key components.

1. A centralized electronic registration system that all teachers must use in order to receive continuing education credits toward recertification. The centralized system allows for the collection and analysis of in-service activity data by school and by individual teacher.
2. A formal training process at the district level, composed of a series of workshops on a research-based curriculum conducted by district standards coaches who are veteran teachers trained by nationally certified trainers in effective literacy instruction.

3. Follow-up support for teachers by school-based standards coaches who have the primary responsibility of facilitating teachers' application of literacy training to their classroom practices.
4. Monthly teacher training sessions by district standards coaches for school-based standards coaches, as well as follow-up site visits, classroom observations, and modeling in actual school settings. Specific job descriptions for the district and school standards coaches can be found in Appendix E.
5. Training for administrators on both the standards of classroom literacy and program implementation. The training features assessing observable teacher and student behaviors and contributes to the effectiveness of performance monitoring.
6. A comprehensive assessment system in which teachers complete a self-assessment, school standards coaches complete an assessment of teachers, and the district standards coaches assess the teachers and the school standards coaches. Observable teacher behaviors are rated on a scale from 1 (no evidence) to 4 (fully evidenced). When all three assessments are complete, differences are reconciled.
7. Additional, follow-up training based on the results of the assessments.

The first step of the study was to extend this model by adding the evaluation of student performance as a method for identifying training programs in need of improvement and of additional training (Figure 3).

The Schultz Study: Can Student Achievement Be Correlated to Teacher In-Service Education?

The “gold standard” for evaluation research is the use of rigorous designs, preferably involving random assignment of subjects to groups and the assignment of subjects (or groups or schools) to conditions. If a prospective study were feasible, then experimental conditions could be more closely controlled and statements about whether certain “programs” caused results could be made with a substantial degree of confidence. However, such a study is very difficult to manage for school districts under normal operating circumstances. The Schultz Center also had limited resources for conducting the study. Nevertheless, because the Schultz Center provides standards based–best practice training in a technology intensive environment and is a major provider of in-service education in Duval County, this system provided the components and data from which an assessment could be made of the potential impact of professional development on teachers, and in turn, the potential impact on the students being served by those teachers. The study project contemplated a simple theory of action described in Figure 4.

Figure 3: An In-Service Model for Continuous Improvement

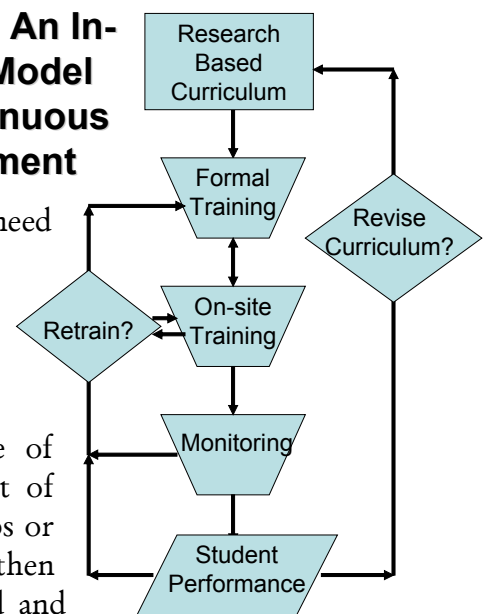
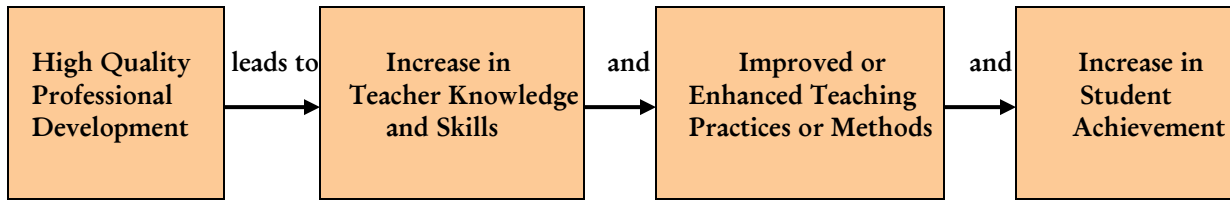


Figure 4: Theory of Action of In-Service Education



Design of the Study: Accounting for Potential Variables

Since the project necessarily involved a post hoc analysis and could not plan to systematically include specified teachers or schools, it was necessary to assure that the study population involved a representative selection of teachers and students. Analysis of the data showed a fairly even distribution of elementary teachers from across the district who had participated in a series of literacy courses. Considerable effort was then made to assure that the schools and teachers involved were representative of the district as a whole, and that the students whose scores were eventually used were representative of the racial and socioeconomic makeup of the district.

Due to long-standing desegregation plans and cross-district busing and magnet program strategies, students from all racial groups and socioeconomic levels attend schools throughout the district. All sections of the district were represented in the schools and teachers included in the study. In order to try to eliminate statistical bias, the study attempted to ensure that student data was reflective of the range of student achievement as a whole, and that teachers participating in standards-based literacy training came from schools from across the district and that students came from representative racial and socioeconomic subgroups of students. Although the project was necessarily limited in time and scope, the study populations (students and teachers) were stratified to provide confidence in the results.

Defining Categories

Teachers were sorted into three categories, based on their degree of completion of Schultz Center Standards-Based Literacy Training (SBLT).

- **Non-Professional Development:** (NPD) Teachers with students having FCAT scores but **who had not yet completed SBLT professional development courses at the Schultz Center.** Literacy training was a major district priority during this period. **It cannot be overemphasized that the NPD group included a large number of teachers receiving the gamut from no professional development, to those participating in a variety of traditional in-service offerings.**
- **Some-Professional Development:** (SPD) Teachers with students having FCAT scores and **who had completed at least one SBLT professional development course** at the Schultz Center. See the above comment related to other training.
- **Critical-Professional Development:** (CPD) Teachers with students who had FCAT scores **who had completed the ‘critical mass’ of SBLT professional development hours at the Schultz Center (at least 80% of the training).**

Figure 5: Number and Percentage of Students by Teacher Category

Classification of Training	Frequency*	Percent
Non-PD	117	14.8
Some-PD	242	30.6
Critical-PD	433	54.7
Total	792	100.0

*These figures represent the number of students.

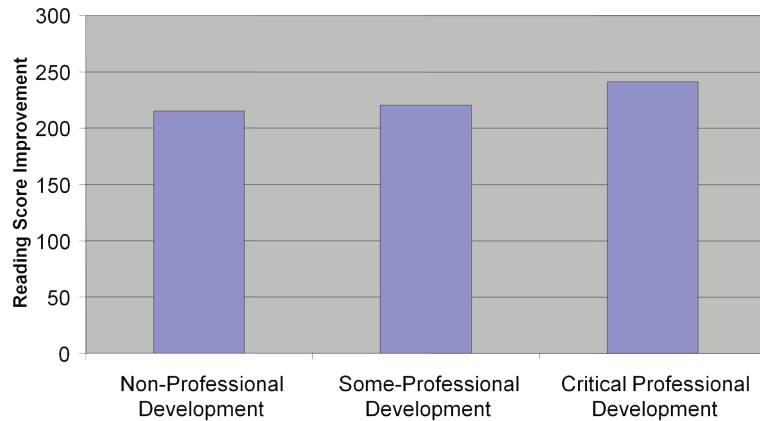
Teachers participating in Year 1 represented achievement results for 792 students from all regions of the Duval County school district. This population was limited to fourth and fifth grade teachers whose students had at least two years of FCAT Reading-DSS scores. Teachers were fairly evenly split between fourth and fifth grades, and only schools labeled as ‘elementary schools’ were included. The impact on students for each group of teachers was then evaluated based on student learning.

Year 1 Results

District wide, all fourth and fifth grade students had strong learning gains over their FCAT results from their previous year. Figures 3 and 4 illustrate the improvements for students in grades 4 and 5 based on the amount of SBLT training received by their teachers. While all students showed learning gain averages, there was more improvement among students of teachers with the maximum levels of participation in SBLT. Fourth graders showed a range of average learning gains in reading from 215 points for teachers with no training to 241 points for teachers with full training (Figure 3). The improvement of the students in this latter category exceeded the Duval County district average of 223 points by approximately 8%.

Sources: Duval County School District and the Schultz Center.

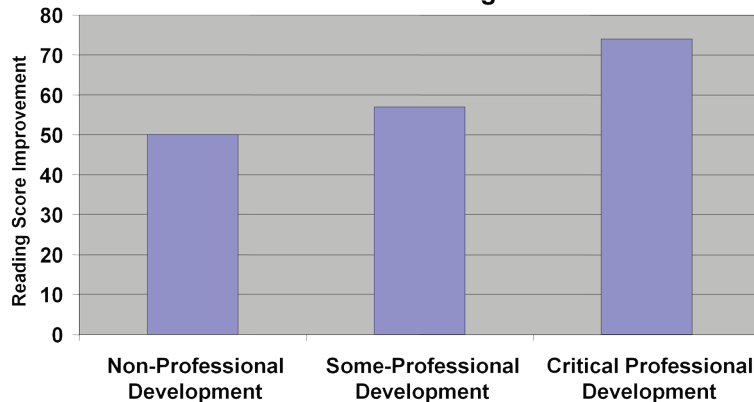
Figure 3 : 4th Grade Reading Score Improvement Based on Level of Teacher In-Service Training



Fifth grade students showed even stronger results when taught by a teacher who had received full SBLT training (Figure 4). The 74-point gain by students from teachers with the maximum levels of training exceeded the district average of 60 by over 20%.

Sources: Duval County School District and the Schultz Center.

Figure 4 : 5th Grade Reading Score Improvement Based on Level of Teacher In-Service Training



The second year of the pilot continued with more work to document the impact of professional development on elementary grade results. Florida educators and decision-makers alike understand the critical need for children to become able readers, and the long term impact on academic achievement this ability will have on students. In 2005-2006, two changes were made to the focus of attention in the impact study: first, we sought ways to the impact of teacher training on third grade students – an important group representing the first point at which students are comprehensively assessed on the FCAT, and second, we considered other bases for comparison of student results to the district as a whole because ending the third year of district wide emphasis on SBLT is reaching larger numbers of teachers to the point where focus now seems appropriate to shift to the district as a whole rather than evaluating the impact strictly on those participating. The investment and logic behind carrying out

systematic and standards-based training for teachers is based on a belief that it will enhance the abilities of all teachers to improve the academic achievement of students; therefore, the include analysis on comparison group in the long run must expand to the observed performance of students district-wide. While resources for this research have not allowed full implementation beyond working with pilot data, we are hopeful that legislative funding will permit building capacity and the system mechanisms necessary to build impact evaluation into the expectations for all literacy training, and eventually all forms of professional development delivered in Duval County.

Year 2 Results

In Year 1 (2004-2005) participation data concentrated mostly on teachers at grades 4 and 5 who had completed two years of training, resulting in data for 55 teachers and 792 students. In Year 2, the pilot focused on grades 3, 4, and 5 with 123 teachers and 2,409 students involved. For third grade, the challenge was in finding an adequate means of evaluating impact since spring of grade 3 is the first time at which the FCAT Reading test is administered state-wide. However, in Duval County (and many other districts in the state), a series of progress monitoring (benchmark) assessments are given which are modeled after the test blueprints for the Florida Department of Education’s FCAT test. Using the results of such assessments give districts predictive information about how children are currently doing toward mastering the skills that will later be assessed on the FCAT, and are used to provide diagnostic information to teachers for planning instruction. These tests are given twice a year (fall and mid-year) and in our case, we can use the results of these assessments, given in the fall to understand the functioning level of third grade students who had not yet taken an FCA T. Because this study did not allow for randomly selecting teachers to include in training, and children are not assigned to teachers in pre-determined ways, this fall benchmark test can give us a way to control for the pre-existing differences that children will have in reading ability as they begin third grade instruction. To do this, it is necessary to know that the benchmark test being used is highly correlated to the later outcome being measured (FCAT Reading).

An analysis of the Duval County Benchmark Reading Tests was conducted to determine the degree to which we could use this measure with confidence as an early indicator of later FCAT performance. Figure 8 below reports the correlation coefficients for both fall and mid-year benchmark reading tests given to Duval County students. This reflects the degree to which the benchmark tests can be thought to measure the same general content as is measured by the spring FCAT reading test.

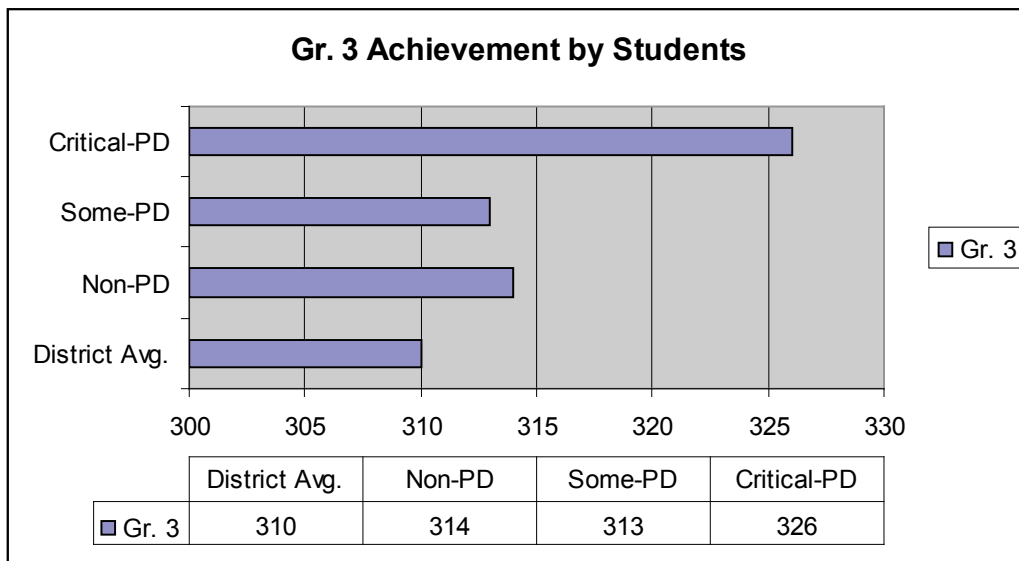
Figure 8: Relationship of Duval County Benchmark Reading to FCAT Reading

Duval Benchmark Reading Tests	FCAT Reading (Scale Score)		
	Grade 3 FCAT	Grade 4 FCAT	Grade 5 FCAT

Fall Benchmark 1	+ .72	+ .72	+ .74
Mid-Year Benchmark 2	+ .73	+ .75	+ .74

By using this information for third grade students (benchmark fall results), we were able to account for pre-existing differences among students to a large extent, and determine the impact on achievement levels of third grade students taught by teachers who had completed the desired levels of professional development. This statistical procedure is called analysis of covariance, and results in achievement level scores for the FCAT that have been adjusted for differences in students' beginning levels of achievement. Figure 9 below is a graph illustrating the results for third grade students in Year 2 (2005-2006):

Figure 9: Achievement of Gr. 3 Students by Category of Teacher Professional Development

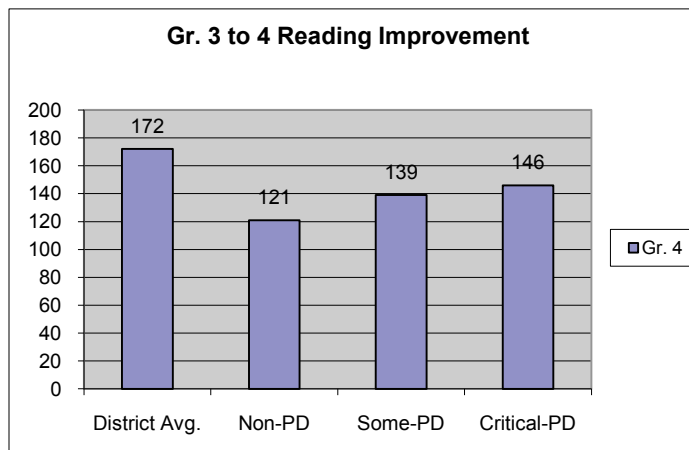


There are two items to note about the graph above showing third grade adjusted achievement. First, we are pleased to see the same trend noted for third grade students of teachers receiving “critical” doses of professional development. The FCAT scale score for students of teachers who completed both courses of SBLT was 326, compared to 313 and 314 for students of teachers completing still in process, or not having completed a single course. The second note is a reminder that the “Non-PD” group classification does not imply no training. This classification label was used to indicate the student achievement for: 1) teachers who are in the process of receiving training, and 2) a random sample of students from all other teachers of

similar classes in the district. This was based on our decision for future analyses to raise the level of comparison to include performance of students from across the district, not limiting the comparison to the students in the pool of teachers in training. If we are interested in systemic, district-wide reform, using this wider frame lens to evaluate our progress is necessary. A complicating factor to this approach will be that using the entire teaching corps as our frame of reference will incorporate a group of teachers who individually, and in selected schools, had been trained intensively as part of the *NCEE America's Choice* school design initiative. The District Average shown (310), on the other hand, is the average score received by all third grade students in the spring of 2006. There is no statistically significant difference between the score of the first two categories (Non-PD or Some-PD), and the point is that over time, we expect the performance of students whose teachers complete district-endorsed professional development will match and exceed that of previously trained standards-based or typically-trained teachers.

Finally, to continue the examination of learning gains in grade levels which do allow the possibility of comparing progress year to year (third to fourth, and fourth to fifth), the Year 2 pilot included that analysis as well. There were insufficient numbers of fifth grade teachers completing training this year to produce stable and reliable estimates of gain scores, but fourth grade data was available. Figure 10 below illustrates the gains made by fourth grade students compared with their own third grade achievement, classified by the same labels used before:

Figure 10: Year 2 Learning Gains of Grade 4 Students



Fourth grade students showed a range of average learning gains in reading from 121 points for teachers with no training or not completing a single course, to 146 points for teachers with full training. The improvement in learning gains for students achieved by teachers completing the full sequence of SBLT reflects a 21% improvement over the gains achieved by teachers not

yet trained. The Year 1 pattern of improved performance by groups receiving increasing levels of training is repeated in Year 2, although reading performance for the fourth grade overall was lower this year than last, both in Duval County and in the state as a whole. The District Average in this graph reflects all students in Duval County with matched third and fourth grade test scores – once again including all teachers receiving earlier training by NCEE and the wide-screen view of district-wide performance that will now be used to monitor and track systemic standards-based improvement in Duval County over time.

Conclusion

The Schultz model has shown promising results in Years 1 and 2. **Although the limited scope and scale of this study does not permit definitive conclusions, it would clearly appear to validate the principle that a high quality, systematic, standards-based training program enhances teacher education and performance and, in turn, student achievement.** A framework and methodology has been established that can be built upon to establish a more refined and statistically comprehensive analysis. We are hopeful that legislative support for this initiative, and the granting of special funding to support this work, will allow us to more fully implement impact evaluations for all student groups being targeted (elementary and secondary), and make the protocol for outcomes research being developed by the Schultz Center a model that can be shared and disseminated statewide. As the number of teachers reach “critical mass”, and training continues to grow and saturate to all schools and grade levels, statistically valid and replicable evaluation techniques will allow a definitive assessment of results. Of even more importance is that as students are taught by successive teachers using consistent, research-based techniques, improvement in student performance should be accelerated. As a result of improved teacher performance, exposure of students to successive teachers trained in the same techniques, and taught with a consistent and aligned curriculum, it can be expected that a definitive answer to the question of the effect of training on student performance will be forthcoming.