Baseline Study of Selected Professional Development Schools in West Virginia

Christina L. Gilchrist, Ph.D.
Dawn M. Salgado, M.A., Research Consultant
Joseph L. Holloway, B.S.

December 2005

Appalachia Educational Laboratory (AEL)
at

EDVANTIA™
Partners in education. Focused on results.

Baseline Study of Selected Professional Development Schools in West Virginia

By

Christina L. Gilchrist, Ph.D.
Dawn M. Salgado, M.A., Research Consultant
Joseph L. Holloway, B.S.

December 2005

Appalachia Educational Laboratory at Edvantia, Inc.
Charleston, West Virginia

Edvantia was founded in 1966 as the Appalachia Educational Laboratory, Inc. (AEL); on September 1, 2005, AEL became Edvantia, Inc. The Regional Educational Laboratory for the Appalachian region is known as the Appalachia Educational Laboratory at Edvantia.

Edvantia is a nonprofit education research and development corporation, founded in 1966, that partners with practitioners, education agencies, publishers, and service providers to improve learning and advance student success. Edvantia provides clients with a range of services, including research, evaluation, professional development, and consulting.

For information about Edvantia research, products, or services, contact

EDVANTIA™

P.O. Box 1348, Charleston, WV 25325 • 304.347.0400 • 800.624.9120 • fax 304.347.0487
One Vantage Way, Suite D-210, Nashville, TN 37228 • 615.565.0101 • fax 615.565.0112
info@edvantia.org • www.edvantia.org

Edvantia maintains policies and procedures that govern the Institutional Review Board in ensuring the welfare and protecting the rights of human subjects who choose to participate in research and evaluation activities.

This publication is based on work sponsored wholly or in part by the Institute of Education Sciences (IES), U.S. Department of Education, under contract number ED-01-CO-0016. Its contents do not necessarily reflect the positions or policies of IES, the Department, or any other agency of the U.S. government.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Background. In West Virginia, African American and lower-income White children begin school less prepared to read and learn than their classmates—an indicator of an achievement gap. House Bill 4669 was signed into law in March 2004. It authorized the creation of a Professional Development Schools (PD Schools) demonstration program in counties with enrollments of 5% or greater African American or low-socioeconomic students. The program was designed to address achievement gap concerns.

The West Virginia Department of Education (WVDE) was, therefore, required to both design and implement the Professional Development Schools (PD schools) demonstration program and report its progress annually to the state board of education. In August 2004, the WVDE requested that the Appalachia Educational Laboratory at Edvantia, Inc. assist with the evaluation component of the demonstration project.

The WVDE proposed a framework with six primary initiatives that encompass many comprehensive strategies for closing the achievement gap: (1) implement an awareness communication plan, (2) provide administrative support for counties (known as Closing the Achievement Gap Specialists or CAGS), (3) conduct needs analysis, (4) provide technical assistance that enables counties to focus on improving social and economic policies that enable children to come to school ready to learn, (5) involve parents and families, and (6) evaluate project effectiveness.

Purpose. In August 2004, the WVDE asked the Appalachia Educational Laboratory at Edvantia, Inc. to evaluate the demonstration project, in order to meet the evaluation and reporting requirements set forth in the law. According to Fink (1993), baseline evaluation data allow for subsequent analyses of the effect or impact of a program. The purpose of this study is to set a baseline prior to full implementation of the PD schools framework. Two evaluation-oriented goals were set: (1) to compare the perception of professional staffs at PD and matching schools across several school quality dimensions and (2) to see if there are any statistically significant differences between PD and matching schools in student achievement. If differences between PD and matching schools exist, then the evaluators can statistically adjust for such differences in subsequent analyses. Furthermore, statistical significance testing during the baseline year will also allow the evaluators to examine whether PD and matching schools were significantly different prior to full implementation of the PD schools framework.

Sample. Thirty PD schools were selected by the county superintendents and were confirmed by the state superintendent. The PD schools that were selected met the specified minimum enrollment of 5% African American or low-socioeconomic students. Those PD schools are intended to become demonstration schools for the proposed WVDE PD school framework. Lab staff selected a matching school for each PD school based on three criteria—(1) having a similar school level (e.g., elementary, middle, or junior high school); (2) having a similar percentage of African American students and being within the same (or from a nearby)region; and (3) having a similar school size. Because of the unique nature of the selected PD schools, only approximate matches could be found in some cases.

Two main sources of data were tapped during the 2004-2005 school year—professional staff surveys and student achievement data. Three surveys were administered to professional staffs of both PD and matching schools during the 2004–2005 school year. Results of the Measure of School Capacity for Improvement (MSCI), Perceptions Of School Culture (POSC), and Continuous School Improvement Questionnaire (CSIQ) provide a context to student achievement data. Student achievement data consisted of WESTEST Mathematics (Math) and Reading/Language Arts total subscale scores for the 2004-2005 school year from all students of participating PD and matching schools.

Surveys. The data were collected in cooperation with the WVDE and the Closing the Achievement Gap Specialists (CAGS) identified and hired by the WVDE. The CAGS administered the surveys in the PD schools; other WVDE personnel administered the surveys in the matching schools. Those surveys were returned to the laboratory for processing and analysis. The WVDE provided lab evaluators with a database of the test scores; as required by law, all identifiable student information was removed to protect the privacy of the students.

Response rate score reliability estimates were calculated for this set of survey responses to assess the fitness for further statistical analyses. Almost twice as many PD as matching school staff members responded (PD = 65.3%; matching = 34.7%). Cronbach alphas were computed and ranged from .78 to .96. This suggests that the data were reliable and further analysis was appropriate.

Individual school subscale means were compared with the subscale means of all other participating schools for each survey. School-level profiles for each of the survey sets were developed and provided directly to each school's leadership for use as they saw fit.

Separate Multivariate Analyses of Variance (MANOVA) test statistics were computed for each instrument using its subscale scores as the dependent variables; school type, school level, and county were used as the independent variables.

The MANOVAs of survey data indicated significant differences between school types (PD v. matching schools) on the majority of the instruments' subscales, with small to medium strengths of association, ranging from .001 to .200, respectively. Furthermore, follow-up ANOVAs indicated that PD school professional staffs assigned lower ratings to items than did their matching counterparts. This can be interpreted as professional staffs from PD schools having lower perceptions regarding their schools' capacity for improvement, school culture, and continuous school improvement dimensions than the associated matching schools.

The PD schools, which are starting to implement the professional development framework, were rated by their professional staffs significantly lower than matching schools on nearly all subscales related to potential to become high-performing learning communities (MSCI), overall school culture (POSC), and staff perception of the school's performance and improvement (CSIQ). Although the statistical analyses adjust for differences within and between groups, an ideal study would have similar response rates from both groups. Because matched

schools had no incentive to participate (other than receiving school-level profiles), the basis for comparison was not ideal. Thus, interpretations must be made with caution.

In the first year of the professional development schools demonstration project, these results confirm superintendents' selections of the designated professional development schools. PD schools had higher percentages of African American and low-socioeconomic students. Also, PD schools scored significantly lower than matching schools across all of the surveys and on many of the subscales within those surveys. This can be interpreted as each PD school's staff having a significantly lower opinion of their school, its culture, and its capacity to improve than the matching schools. This should not be a surprise, as the superintendents intended to select schools with higher percentages of at-risk students to become model schools.

Furthermore, the lower ratings on both the capacity for improvement and perceptions of school culture would naturally be associated with lower ratings on continuous school improvement. That is, professional staff members who perceive that their school has a low capacity for improvement and does not have a positive culture are not likely to rate their schools as meeting the high standards set for continuous school improvement. This is also expected in baseline studies.

Student Achievement. The MANOVAs of student achievement data indicated that there were no statistically significant multivariate effects of school type (PD vs. matching schools) on the combined set of 2004-2005 WESTEST Math and Reading/Language Arts subscales scores, except for the third and fourth grades. However, there were significant multivariate effects of comparison county for almost all grades (specific county differences are summarized by grade).

Within the sample of 60 schools, there are no mean differences between PD and matching schools. Typically, this would be expected when matching schools were purposefully selected. However, based on this subset of schools, this finding might not support the program theory underlying the House Bill—there might not be any statistically significant differences between the selected PD and matching schools. There are other competing explanations for this phenomenon. It may also be that the relatively small percentages of African American and low-socioeconomic students are being masked by the percentages of low-socioeconomic students from the corresponding matching school. Or, perhaps, if there are theoretically significant differences between the selected PD and matching schools, the differences do not show up within the sample of 60 schools, but might show up on a state-level analysis.

Students' performance on WESTEST Math and Reading/Language Arts subscales was related more strongly to the county in which they live, rather than attendance at a PD or a matching school. This phenomenon might occur because circumstances are fairly uniform across the county. So perhaps the academic difficulties experienced by African American students from PD schools, however real, may not significantly differ from the academic difficulties experienced by low-socioeconomic students in matching schools. Therefore, the significant differences that surfaced among students were associated with the county means and might relate more to the disadvantages shared by all students of a county than to ethnic or socioeconomic factors.

Recommendations. The first and primary recommendation, especially if the WVDE intends to continue an evaluation throughout the life of the demonstration program, is to find a way to provide incentives for matching schools to participate in the surveys. The closer the samples are in size and composition, the more accurate interpretations of any data analyses will be.

Survey results reported here indicate a need for policies and initiatives aimed at improving the school programs and related perceptions. Specifically, interventions should emphasize (1) increasing staff communication and collaboration, and students' levels of motivation and capacity to be successful; (2) increasing student, parent, and community involvement; (3) creating an environment that enhances students' ability to learn and teachers' ability to teach effectively; and finally, (4) increasing teachers' abilities to impact student learning through needed resources, increased training, and/or developing a more supportive and collaborative teaching community among staff.

The stated purpose of this baseline study was to investigate whether there are any significant differences between PD and matching schools' student achievement, so evaluators could account for such differences in future evaluations. The evaluators suggest that other perspectives could be used to examine the data. If possible, analyses could investigate differences among both student- and school-level variables. It would be desirable to identify the characteristics within the counties that highly correlate with achievement.

Finally, this evaluation assesses only a subset of the variables within the PD school framework. Additional constructs are outside the purview of this baseline year. As program components develop, data could be gathered about their implementation and effectiveness.

Now that a baseline has been established, it is important to continue data collection so trends can be identified. Annual snapshots of staff perceptions and student achievement can yield data for identifying areas of strength and weakness, as well as areas of decline or improvement. Such information could be invaluable in informing changes to the program that would enhance the end product. This is consistent with the diagnostic intent of formative evaluation approaches.

INTRODUCTION

This report presents a baseline study of selected Professional Development and matching schools in West Virginia. This section consists of the background, purpose, audience, and overview of the report.

Background

In West Virginia, African American students have been overrepresented in special education categories and underrepresented in Advanced Placement courses (Education Trust, n.d.; Ginsberg, 2004; Kusimo, Petty-Wilson, & Body, 2004). According to Ginsberg (2004), African American and lower-income White children begin school less prepared to read and learn than their classmates, which indicates an achievement gap for these children.

To ameliorate the situation in 2003, the governor of West Virginia created the Governor's Minority Students Strategy Council to investigate achievement gap issues in the state and make recommendations (State of West Virginia Governor's Minority Students Strategy Council, n.d.). The council identified gaps and proposed legislation to implement professional development schools (Kusimo et al., 2004). House Bill 4669, which was signed into law in March 2004, authorized the creation and designation of professional development schools in counties with 5% or greater enrollments of African American or low-socioeconomic students (Southern Regional Education Board, 2004; West Virginia Education Association, 2004). The West Virginia Department of Education (WVDE) was, therefore, required to both design and implement the Professional Development Schools (PD schools) demonstration program and report its progress annually to the state board of education.

In August 2004, the WVDE requested that the Appalachia Educational Laboratory at Edvantia, Inc. assist with the evaluation component of the demonstration project.

Purpose

According to Fink (1993), baseline evaluation data allow for subsequent analyses of the effect or impact of a program. The purpose of this study is to set a baseline prior to full implementation of the PD schools framework. Two evaluation-oriented goals were set: (1) to compare the perceptions of professional staffs at PD and matching schools across several school quality dimensions and (2) to see if there are any statistically significant differences between PD and matching schools in student achievement. If differences between PD and matching schools exist, then the evaluators can statistically adjust for such differences in subsequent analyses. Furthermore, statistical significance testing during the baseline year will also allow the evaluators to examine whether PD and matching schools were significantly different prior to full implementation of the PD schools framework.

Audiences

The primary audiences of this report are the WVDE staff, who must report annually to the state department of education regarding the PD schools demonstration, and participating schools' professional staffs. A secondary audience consists of the professional staffs from West Virginia schools that have enrollments of 5% African American or low-socioeconomic students but were not designated as PD schools. Finally, the Institute of Education Sciences at the U.S. Department of Education receives a copy of this report, pursuant to fulfilling the requirements of Edvantia's Regional Educational Laboratory (REL) contract.

Overview

Four main sections compose this evaluation report: (1) focus of the evaluation, (2) description of the methods used to conduct this evaluation, (3) summary of the findings, and (4) conclusions and recommendations.

FOCUS

This section describes the WVDE Professional Development Schools Demonstration and presents the evaluation approach.

WVDE Professional Development Schools Demonstration

West Virginia House Bill 4669 was designed to address the achievement gaps experienced by both minority and low-socioeconomic students in the state (West Virginia Education Association, 2004). It proposed a 5-year pilot project called the Professional Development Schools demonstration program (PD schools) in which demonstration schools implement various strategies for closing the achievement gap. The bill proposed having schools work with education and community organizations to address achievement gaps and find solutions that could then be used at other schools with achievement gaps (Southern Regional Education Board, 2004).

When House Bill 4669 was signed into law in March 2004 (W.V.C. §18-2E-3g), a new professional position was created within the West Virginia Department of Education (WVDE) to oversee the implementation of the PD schools project. In response, the WVDE staff presented various strategies for closing the achievement gap: adding preschools to schools; lowering student-teacher ratios; assigning reading/literacy teachers to each school; requiring one-on-one reading instruction for 30 minutes a day for all first- and second-grade students; focusing on reading, writing, and math; establishing 90-minute English classes in middle schools; extending the kindergarten school year to include summers; offering tutoring on Saturdays; and fostering health care.

These proposed strategies were further distilled into six primary initiatives: (1) implement an awareness communication plan, (2) provide school-level support for counties (known as Closing the Achievement Gap Specialists or CAGS), (3) conduct a needs analysis, (4) provide technical assistance that enables counties to focus on improving social and economic policies that enable children to come to school ready to learn, (5) involve parents and families, and (6) evaluate program effectiveness.

While details of the framework were being refined, schools were chosen to implement the framework. The law required that three public elementary, middle, or junior high schools (from each county with enrollments of at least 5% African American or low-socioeconomic students or more) become demonstration schools. Ten counties met this criterion and a total of 30 demonstration schools were nominated by county superintendents and were accepted by the state superintendent of schools. By the end of the demonstration project, these schools will have piloted curriculum methods and instructional techniques aimed to improve academic achievement among minority and low-socioeconomic students (Boyles, 2004).

Evaluation Approach

In August 2004, lab staff were asked to assist with the mandated evaluation component of the PD schools demonstration program. This request was received after the preliminary PD schools framework was developed and after the PD schools were selected. However, this request was received prior to full implementation of the PD schools framework. Because PD schools were to begin implementing the PD schools framework during the 2004-2005 school year, a formative evaluation seemed appropriate.

Another issue that warrants attention is the law's requirement for a *summative* evaluation of the PD schools framework at the end of the five-year demonstration period. So, it seemed logical that the first year's evaluation should serve as a baseline. According to Fink (1993), baseline evaluation data allow for subsequent analyses of the effect or impact of a program. Given this content, this study did not focus on program implementation or development; rather, it focused on collecting student achievement and school quality indicators by which to judge the effectiveness of the PD schools framework in the future. As a result, this study was designed in a manner that would enable at least a quasi-experimental design in subsequent years.

The resulting evaluation questions align with the instruments selected to gather school quality and student achievement data:

1. **Measure of School Capacity for Improvement.** Did the perceptions and attitudes toward equity in practice, expectations for student performance, differentiated instruction, improving program coherence, peer-reviewed practices, coordinated curriculum, and technical resources differ between professional staffs from schools that will be implementing the professional development schools framework and professional staffs from schools that will not be implementing the professional development schools framework?

2. **Perceptions Of School Culture.** Did the perceptions of the school culture on six dimensions differ between professional staffs from schools that will be implementing the professional development schools framework and professional staffs from schools that will not be implementing the professional development schools framework?

3. **Continuous School Improvement Questionnaire.** Did the ratings of school performance on the seven dimensions of continuous school improvement differ between professional staffs from schools that will be implementing the professional development schools framework and professional staffs from schools that will not be implementing the professional development schools framework?

4. **Student Achievement Data.** Did the Math and Reading/Language Arts mean subscale scores on the WESTEST differ between students from schools that will be implementing the professional development schools framework and students from schools that will not be implementing the professional development schools framework?

METHODS

This section describes the perspective, design, sampling, data collection, and analysis procedures used in this evaluation.

Perspective

This study was intended to describe professional staffs' perceptions of school quality and student achievement using quantitative data that can provide a baseline for future evaluations.

Design

The evaluation design that was used to compare PD and matching schools' student achievement and school quality indicators was pre-experimental in nature. According to Shavelson (1996), an intact-group comparison design is appropriate for collecting baseline data because it may "provide useful insights that can be incorporated into other research designs" (p. 25); Gay (1996) also concurs. The following diagram illustrates this design:

NR	X	O_1
NR		O_1

Note: NR = Non-random assignment; X = Designated professional development schools; O_1 = MSCI, CSIQ, and POSC (see Appendix A).

The use of comparison groups in this baseline year will enable quasi-experimental designs in future evaluations.

Time Frame

This study was conducted during the 2004–2005 school year. Planning for this data collection began in fall 2004. Data were collected in spring 2005.

Sample

Different sampling procedures were used for PD and matching schools. The WVDE selected 30 PD schools and lab staff suggested 30 matching schools (totaling 60 schools statewide) to participate in this study. The sampling procedures follow.

Selecting PD Schools

According to the criteria established in the legislation, 10 counties qualified to participate in the demonstration project because they had enrollments of at least 5% African American or low-socioeconomic students. The law required that three schools from each county (elementary, middle, or junior high schools) become demonstration schools (n = 30). Although random selection and assignment of a purposefully selected sample would have been ideal, HB 4669 vested authority to select and assign schools in the WVDE or county superintendents.

Selecting Matching Schools

Matching schools were selected for each PD school (n=30). The statistics used to compare the PD schools to other schools in West Virginia were downloaded from the National Center for Education Statistics (NCES) Web site (n.d.). The statistics available for comparison from this source were grade level, locale, type, number of students, number of teachers, number of students by gender, number of students per racial group, number of students qualifying for free or reduced-price lunch, and number of migrant students. Other statistics included whether the school was a charter, magnet, Title I school, and/or Title I schoolwide. The percentage of African American students in a school was calculated by dividing the number of African American students by the total number of students. The following primary criteria were then applied in selecting the matching schools:

First criterion. PD schools were matched to other schools of similar grade levels (e.g., elementary, middle school, or junior high schools).

Second criterion. The percentage of African American students in each PD school was compared to other schools in the district. (Three different groups for comparison were made: group one, 1-33%; group two, 34-66%; and group three, 67-100%). If there were no matches within the district, then the PD schools were compared to other schools outside the district but within the region.

Third criterion. The number of students enrolled in the PD school was compared to other schools in the district. (Three different groups for comparison were made: group one, 1-300; group two, 301-600; and group three, 601-900.) Again, if there were no matches within the district, then the PD schools were compared to other schools outside the district but within the region.

When a school matched all three criteria above, then the other statistics were compared. If the schools also had reasonably similar characteristics, they were noted as a match. If no school within the region fit the above criteria, then the closest match within a nearby region was chosen. The resulting list of recommended matching schools was approved by the WVDE.

Descriptive Statistics

Table 1 presents data used to match schools with the PD schools. The state department of education chose the 30 PD schools from 7 of the 8 West Virginia Regional Education Service Agency (RESA) districts. The table also shows the variables used to select the matches. Because of the unique nature of the selected PD schools, only approximate matches could be found in some cases.

Table 1: Numbers or Percentages of African American, White, Free or Reduced-Price Lunch, and All Students by RESA, School Type, School Name, School Level, and Locale

RESA	School Type	School Name	% A.A. Stds.	Level	# All Stds.	Locale	Std./ Teach. Ratio	A.A. stds.	White stds.	Free or Reduced-Price Lunch
I	PD	Beckley/Stratton Middle	22%	Mid	760	Rural, outside CBSA*	14.9	167	570	404
	Match	Park Middle	22%	Mid	383	Small Town	12.6	86	289	181
	PD	Stratton Elementary	66%	Elem	242	Small Town	9.3	160	81	163
	Match	Crescent Elementary	19%	Elem	310	Small Town	15.5	59	244	136
	PD	Beckley Elementary	63%	Elem	362	Small Town	9.4	229	124	251
	Match	Kimball Elementary	36%	Elem	219	Rural, outside CBSA*	11.5	78	141	188
	PD	Fall River Elementary	1%	Elem	228	Rural, outside CBSA*	12	2	226	197
	Match	Montcalm Elementary	1%	Elem	406	Rural, outside CBSA*	15.6	3	401	319
	PD	Mount View Middle	31%	Mid	282	Small Town	14.1	88	193	228
	Match	Summers Middle	6%	Mid	405	Small Town	16.5	24	377	299
	PD	Sandy River Middle	0%	Mid	383	Rural, outside CBSA*	18.2	0	382	306
	Match	Pineville Middle	1%	Mid	296	Rural, outside CBSA*	13.5	3	293	155
	PD	Bluefield Intermediate	34%	Elem	399	Small Town	13.3	136	261	282
	Match	Straley Elementary	12%	Elem	181	Small Town	12.1	21	158	119
	PD	Bluefield Middle	22%	Mid	574	Small Town	12	124	447	339
	Match	Princeton Middle	6%	Mid	624	Small Town	14	40	571	329
	PD	Lashmeet/Matoaka Elementary	1%	Elem	439	Rural, outside CBSA*	14.4	3	434	329
	Match	Athens Elementary	1%	Elem	491	Rural, outside CBSA*	15.6	7	481	227

(Table 1 continues)

Table 1 (continued)

RESA	School Type	School Name	% A.A. Stds.	Level	# All Stds.	Locale	Std./ Teach. Ratio	A.A. stds.	White stds.	Free or Reduced-Price Lunch
II	PD	Spring Hill Elementary	40%	Elem	480	Mid-Size Central City	13.5	191	276	207
	Match	Meadows Elementary	23%	Elem	221	Mid-Size Central City	13.4	50	167	88
	PD	Cammack Middle	27%	Mid	358	Mid-Size Central City	13	98	251	171
	Match	Enslow Middle	10%	Mid	307	Mid-Size Central City	13.1	30	274	249
	PD	Peyton Elementary	25%	Elem	178	Mid-Size Central City	11.2	45	133	146
	Match	Miller Elementary	22%	Elem	139	Mid-Size Central City	12.4	31	107	63
III	PD	Chandler Elementary	59%	Elem	226	Mid-Size Central City	11.9	133	91	192
	Match	J E Robins Elementary	34%	Elem	257	Mid-Size Central City	14.7	88	167	174
	PD	Glenwood Elementary	77%	Elem	206	Mid-Size Central City	12.5	159	44	161
	Match	Piedmont Elementary	72%	Elem	288	Mid-Size Central City	12	206	76	233
	PD	Stonewall Jackson Middle	35%	Mid	668	Mid-Size Central City	13	237	423	398
	Match	Dunbar Middle	30%	Mid	464	Urban Fringe/Mid-Size City	15.5	140	318	253
IV	PD	Collins Middle	10%	Mid	742	Small Town	16.1	76	656	419
	Match	Mullens Middle	5%	Mid	211	Rural, outside CBSA*	14.1	11	200	151
	PD	Montgomery Middle	15%	Mid	244	Rural, outside CBSA*	13.9	37	207	118
	Match	Fayetteville Middle	4%	Mid	271	Small Town	14.6	10	261	153
	PD	Mount Hope High	15%	High	182	Rural, outside CBSA*	11.7	28	152	110
	Match	Fayetteville High	4%	High	298	Small Town	14.9	12	286	103

(Table 1 continues)

Table 1 (continued)

RESA	School Type	School Name	% A.A. Stds.	Level	# All Stds.	Locale	Std./ Teach. Ratio	A.A. stds.	White stds.	Free or Reduced-Price Lunch
VI	PD	Madison Elementary	28%	Elem	310	Mid-Size Central City	11.2	88	217	246
	Match	Elm Grove Elementary	9%	Elem	346	Mid-Size Central City	14.2	32	311	186
	PD	Ritchie Elementary	11%	Elem	262	Mid-Size Central City	14	30	231	214
	Match	Woodsdale Elementary	7%	Elem	334	Mid-Size Central City	13.5	24	305	144
	PD	Wheeling Middle	8%	Mid	227	Mid-Size Central City	12.8	18	209	105
	Match	Bridge Street Middle	7%	Mid	398	Mid-Size Central City	15.4	26	370	220
VII	PD	Dunbar Middle	15%	Mid	337	Small Town	16	51	283	140
	PD	Miller Junior High	14%	Mid	369	Small Town	13.7	52	313	129
	Match	East Fairmont Junior High	2%	Mid	398	Small Town	15.9	9	385	160
	PD	East Park Elementary	2%	Elem	445	Small Town	13.9	8	434	303
	Match	East Dale Elementary	1%	Elem	568	Small Town	17.8	6	557	250
	PD	Watson Elementary	28%	Elem	340	Small Town	13.1	96	243	201
	Match	Jayenne Elementary	11%	Elem	195	Rural, outside CBSA*	17.7	22	171	74
VIII	PD	Eagle Intermediate	24%	Elem	408	Mid-Size Central City	12.1	96	290	235
	Match	Orchard View Intermediate	13%	Elem	443	Mid-Size Central City	14.6	59	369	211
	PD	Martinsburg North Middle	19%	Mid	663	Mid-Size Central City	13	124	499	335
	Match	Moorefield Middle	3%	Mid	470	Rural, outside CBSA*	14.5	14	450	272
	PD	Martinsburg South Middle	12%	Mid	625	Mid-Size Central City	13.6	78	525	249
	Match	Hedgesville Middle	1%	Mid	991	Rural, inside CBSA*	15.5	13	963	324

(Table 1 continues)

Table 1 (continued)

RESA	School Type	School Name	% A.A. Stds.	Level	# All Stds.	Locale	Std./ Teach. Ratio	A.A. stds.	White stds.	Free or Reduced-Price Lunch
VIII	PD	Charles Town Middle	15%	Mid	847	Urban Fringe/Large City	15	130	683	280
	Match	Shepherdstown Middle	6%	Mid	451	Urban Fringe/Large City	14.5	28	411	139
	PD	North Jefferson Elementary	13%	Elem	301	Urban Fringe/Large City	10.9	38	247	179
	Match	Ranson Elementary	26%	Elem	383	Urban Fringe/Large City	13.1	99	263	234
	PD	Wright Denny Elementary	11%	Elem	418	Urban Fringe/Large City	15.9	48	353	127
	Match	Shepherdstown Elementary	11%	Elem	295	Urban Fringe/Large City	14.1	31	257	59

Note: A.A. = African American, *CBSA = Core Based Statistic Area

Data Collection

Two main sources of data were tapped—professional staff surveys and student achievement data. Data collection protocols, using these instruments, complied with corporate policies and procedures regarding projects involving human subjects. Therefore, Institutional Review Board approval was obtained prior to data collection.

Surveys

Three instruments—the Measure of School Capacity for Improvement (MSCI), Perceptions Of School Culture (POSC), and Continuous School Improvement Questionnaire (CSIQ)—were administered to participating schools' professional staffs by WVDE personnel; the instruments had already been developed, refined, and validated by lab staff. A brief description of each instrument is provided below; see Appendix A for further details.

Measure of School Capacity for Improvement. The Measure of School Capacity for Improvement (MSCI) is a 64-item self-report instrument that assesses the degree to which school staff believe their school possesses the potential to become a high-performing learning community.

Perceptions Of School Culture. The Perceptions Of School Culture (POSC) is a 62-item self-report instrument that measures how school staff perceive the culture of their school.

Continuous School Improvement Questionnaire. The Continuous School Improvement Questionnaire (CSIQ) is a 70-item self-report instrument that measures how the school staff rate their school's performance on seven dimensions related to continuous school improvement.

Student Achievement

Student achievement data consisted of WESTEST Mathematics (Math) and Reading/Language Arts (Reading/LA) total subscale scores for the 2004-2005 school year from all students of participating PD and matching schools. The data set included Math subscale scores, Reading/Language Arts subscale scores, county name, school name, various NCLB subgroups (such as gender, ethnicity, low socioeconomic status, limited English proficient, migrant, and special education), and a unique identifier assigned to students by the WVDE.

Appropriate privacy precautions were taken. The data were given to lab staff by WVDE staff without any identifiable information—identifiable student information was removed from the database, but the WVDE staff retained a key matching the confidential state student number that was associated the newly-assigned unique identifier specific to this project. Therefore, if the researchers should desire to match students subsequent test scores, then matches could be performed by the WVDE prior to forwarding the data to lab staff.

Analysis Procedures

Surveys

Lab staff received completed surveys, delivered either by mail or by the Closing the Achievement Gap Specialists (CAGS), and prepared them for processing. The data were scanned using the REMARK application and converted to a SPSS file for statistical analysis. The data were cleaned to detect and remedy any input or processing errors. Descriptive statistics, which were appropriate for the measurement scale of each variable, were calculated for all quantitative data. Response rates were calculated to provide a basis for comparison between PD and matching school respondents. Overall internal consistency reliability estimates were calculated for this administration of each instrument; only when the reliability of the data is established for survey responses are further analyses be warranted.

In order to see if there were statistically significant differences between PD and matching schools prior to full implementation of the PD schools framework, inferential statistics were calculated. Specifically, the researchers selected the Multivariate Analyses of Variance (MANOVA), primarily to protect against Type I errors when computing the various univariate dependent measures and to improve the power of the tests by considering correlations among the variables (Weinfurt, 1995).

Three separate MANOVAs—one for each instrument—using instrument subscale scores as dependent variables and using school type (PD v. matching schools) as the independent variable. The assumptions of MANOVA include multivariate normality, homogeneity of variance-covariance, linearity, and a lack of multicollinearity. These assumptions were tested in order to assure accuracy of the results.[1]

Student Achievement Data

Student achievement data were also analyzed using descriptive statistics and Multivariate Analysis of Variance (MANOVA) procedures. However, because the WESTEST items, content, and subscale ranges differ (at least theoretically) by grade, separate MANOVAS were computed for each grade. The dependent variables were Math and Reading/Language Arts total subscale scores. The independent variables included grouping (PD v. matching schools) and county. Because the evaluators had to select matching schools outside the actual county in five instances, the name of the comparison county (which is the county of the PD school that a matching school

[1] In the event that the assumption of homogeneity of variance-covariance was not met (indicated by a significant Box's *M* statistic), Pillai's Trace and its associated *F*-test were used rather than Wilks' Lambda and its associated *F*-test. Linearity and multicollinearity were assessed through the use of scatter plots and bivariate correlations of the dependent variables and were found to meet the assumptions of MANOVA. Bivariate correlations among instrument subscales ranged from .70 to .90 for the CSIQ, .30 to .80 for the MSCI, and .65 to .89 for the POSC, suggesting that the assumption of a lack of multicollinearity is met, using guidelines suggested by Tabachnick and Fidell (2001).

was paired with) was used. The same statistical diagnostics used to test the MANOVA assumptions for the survey analyses were also used in analyzing the student achievement data.[2]

The tables and figures provided in the Findings section of this report display the descriptive and statistical data; corresponding technical and statistical tables are found throughout the appendixes. The Discussion section (the final section of the body of this report) contains conclusions that were drawn, based on the findings, and offers recommendations.

[2] Pillai's Trace and its associated *F*-test were used, because the assumption of homogeneity of variances was not met for student achievement data. Linearity and lack of multicollinearity were met, based on a bivariate correlation of .894

FINDINGS

This section contains summaries of the findings of the surveys and student achievement data. The supporting technical information, which may not be of interest to every reader, is contained in the appendixes.

Surveys

Response Rates

The survey response rates for each school were calculated and displayed in Appendix B. Because the three surveys were all packaged in one envelope per teacher, there was a similar response rate across the surveys in most of the schools. There was, however, some variation. The response rates of the PD schools were compared to the response rates of the matching schools. The PD schools were found to have higher overall response rates. Overall 1,076 (65.3%) PD school and (34.7%) matching school staff members responded. Figure 1 below illustrates the differences in response rates between the PD and matching schools by instrument.

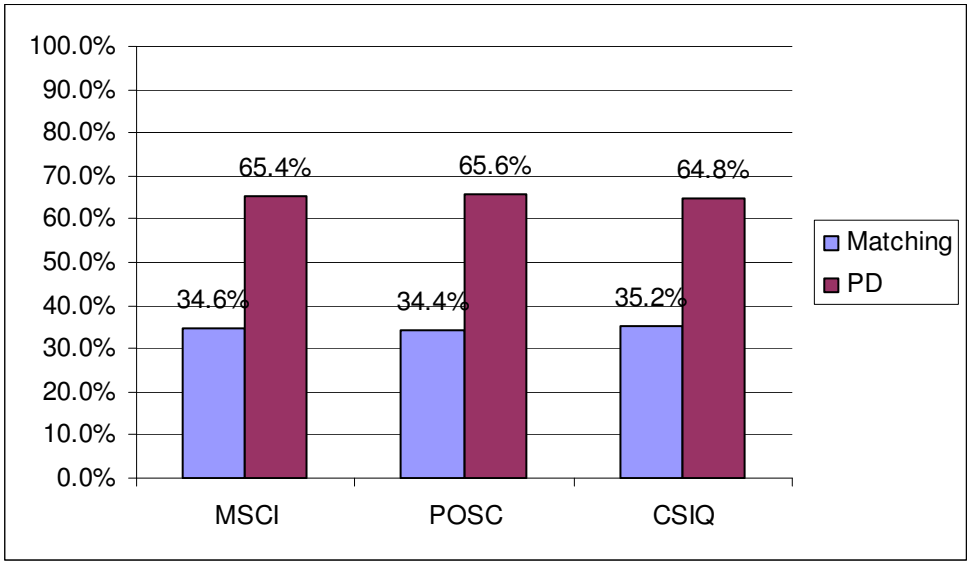

Figure 1: Statewide response rates by survey and school type

Score Reliability Estimates

Internal consistency reliability estimates (Cronbach alpha) were calculated for this administration of the three surveys. Cronbach alpha values ranged from .78 to .96. Fourteen of the alpha values were greater than .90; five were between .80 and .89; only one alpha value was .78. Thus, further analyses of these data are warranted. See Appendix C for further details.

School-Level Survey Profiles

Three separate school-level profiles were developed for each participating school—one for each survey instrument—and were reported by the surveys' corresponding subscales. The figures provided in each profile compared the mean subscale scores of that school with the overall subscale score mean of all other participating schools. These profiles, sent directly to the school leadership during the summer of 2005, were intended to highlight that school's likely strengths and weaknesses along the various dimensions measured. It was hoped that school leaders would share the profiles with the school staffs and use the results in school improvement or professional development planning, as they saw fit.

Statistical Analysis by Survey

Completed and returned surveys were scored and analyzed by school type (PD vs. matching). Descriptive statistics, correlations[3], and a Multiple Analysis of Variance (MANOVA) test statistic were computed using the responses from each instrument. Brief summaries are provided here; the associated technical information is provided in the appendixes.

Measure of School Capacity for Improvement

Descriptive Statistics. Means and standard deviations for each subscale of the Measure of School Capacity for Improvement (MSCI) are presented by school type (see Appendix D). PD school subscale means ranged from 3.42 to 4.98 (*SD*s from .10 to 1.38) and matching school subscale means ranged from 3.51 to 5.14 (*SD*s from .63 to 1.41).

Internal Consistency Reliabilities. Pearson Product Moment correlation coefficients were computed on the MSCI to examine the interrelationships among the MSCI subscales. Results indicate that all subscales are positively correlated with each other, with moderate to high correlations between the subscales. Therefore, it is appropriate to conduct further statistical analyses of the professional staffs' responses to the MSCI.

MANOVA. Results from the MANOVA indicate a statistically significant multivariate effect of school type (PD vs. matching schools) on the combined MSCI subscales (Pillai's

[3] If the dependent variables are correlated, then the results of the MANOVA will not have redundant information; this improves the power of the measures to detect any differences among groups (Weinfurt, 1995).

Trace= .051, F (7, 1049) = 8.13, p < .001), with a strength of association size of η_p^2=.051. Therefore, it was appropriate to calculate follow-up ANOVAs to see which school type had higher subscale means for each subscale.

Matching school professional staffs collectively had higher subscale scores than PD school professional staffs on Equity in Practice (F(1, 1055)=14.58, p= .000, η_p^2= .01), Expectations for Student Performance (F(1, 1055)=40.32, p= .000, η_p^2= .04), Differentiated Instruction (F(1, 1055)=15.57, p= .000, η_p^2= .02), Coordinated Curriculum (F(1, 1055)=5.764, p= .017, η_p^2= .005), and Technical Resources (F(1, 1055)=17.79, p= .000, η_p^2= .02). There is no significant difference between PD and matching schools for Improvement Program Coherence (F (1, 1055) =.3.32, p= .069, η_p^2= .003) or Peer-Reviewed Practice (F (1, 1055) =.0.89, p= .35, η_p^2= .001).

Perceptions Of School Culture

Descriptive Statistics. Subscale total means and standard deviations for each subscale of the Perceptions Of School Culture (POSC) are presented by school type in Appendix E. PD school subscale totals ranged from 15.67 to 50.91 (*SD*s from 3.86 to 10.39) and matching schools ranged from 17.09 to 55.24 (*SD*s from 3.67 to 10.29).

Internal Consistency Reliabilities. Pearson Product Moment correlation coefficients were computed on the POSC to determine the interrelationships among the POSC subscales. Results indicate that all subscales are positively correlated with each other, with moderate to high correlations between the subscales. Therefore, it is appropriate to conduct further statistical analyses of the professional staffs' responses to the POSC.

MANOVA. Results from the MANOVA indicate a statistically significant multivariate effect of school type (PD vs. matching schools) on the combined set of POSC subscales (Pillai's Trace= .058, F (6, 1013) = 10.31, p= .000), with a strength of association size of η_p^2=.058. Therefore, it was appropriate to calculate follow-up ANOVAs to see which school type had higher subscale means for each subscale.

Matching school professional staffs collectively had higher subscale scores than PD school professional staffs on Collaborative Working Relationships (F(1, 1018)=8.90, p= .003, η_p^2= .01); Student-Centered Vision, Mission, and Policies (F(1, 1018)=14.85, p= .000, η_p^2= .01); Student Responsibility for Learning (F(1, 1018)=52.63, p= .000, η_p^2= .05); Teacher Responsibility for Learning (F(1, 1018)=21.18, p= .000, η_p^2= .02); Inviting Physical Environment (F(1, 1018)=21.11, p= .000, η_p^2= .02); and Students and Parents as Decision Makers (F(1, 1018)=31.85, p= .000, η_p^2= .03).

Continuous School Improvement Questionnaire

Descriptive Statistics. Means and standard deviations for each subscale of the Continuous School Improvement Questionnaire (CSIQ) are presented by school type in Appendix F.

Internal Consistency Reliabilities. Pearson Product Moment correlation coefficients were computed on the CSIQ to determine the interrelationships among the CSIQ subscales. Results indicate that all subscales are positively correlated with each other, with moderate to high correlations between the subscales. Therefore, it is appropriate to conduct further statistical analyses of the professional staffs' responses to the CSIQ.

MANOVA. Results from the MANOVA indicate a statistically significant multivariate effect of school type (PD vs. matching schools) on the combined set of CSIQ subscales (Pillai's Trace= .075, F (7, 996) = 11.49, $p < .001$), with a strength of association size of η_p^2=.075. Therefore, it was appropriate to calculate follow-up ANOVAs to see which school type had higher subscale means for each subscale.

Matching school professional staffs collectively had higher subscale scores than PD school professional staffs on Learning Culture (F(1, 1002)=17.32, p= .000, η_p^2= .02), School-Family-Community Connections (F(1, 1002)=36.85, p= .000, η_p^2= .04), Shared Goals for Learning (F(1, 1002)=12.74, p= .000, η_p^2= .01), Purposeful Student Assessment (F(1, 1002)=21.94, p= .000, η_p^2= .02), Effective Teaching (F(1, 1002)=13.08, p= .000, η_p^2= .01), and Aligned and Balanced Curriculum (F(1, 1002)=16.67, p= .000, η_p^2= .02). There is no significant difference between PD and matching schools for Shared Leadership (F (1, 1002) =2.53, p= .11, η_p^2= .00).

Statistical Analysis of Student Achievement Data

Student achievement data are summarized at various levels—by comparison county, school type (PD vs. matching), and grade level. First, an overall sample snapshot provides a context for the results. Then, an brief overall grade-level analysis is presented. However, more thorough and technical presentations of grade-level results are presented in Appendix G.

The 2004-2005 WESTEST data that were provided by the WVDE are summarized by NCLB subgroup (see Table 2). Cell sizes from PD schools ranged from 6 to 6,235 students; matching school cell sizes ranged from 16 to 6,802 students.

Table 2: Total Numbers and Percentages of Students by NCLB Subgroup and School Type

NCLB Subgroup	School Type		Total Number	% of Total
	PD	Matching		
Gender				
Male	5376	4147	9523	18.9%
Female	4968	3724	8692	17.3%
Race				
Asian/Pacific Islander	61	47	108	0.2%
Black	2181	917	3098	6.2%
Hispanic	230	89	319	0.6%
Amercian Indian/Alaskan Native	18	16	34	0.1%
White	7854	6802	14656	29.1%
Other				
Low Socioeconomic Status	6235	4271	10506	20.9%
Migrant	*	*	40	0.1%
Limited English Proficiency	155	76	231	0.5%
Students With Disabilities	1831	1303	3134	6.2%

Note: * Cell size is smaller than $n = 15$; in order to protect the privacy of the students, small cell sizes are not reported in this table. However, the data are still used in the statistical analyses.

When analyzed by grade level, there was no significant multivariate effect of school type (PD vs. matching schools) on the combined Math and Reading/Language Arts subscales for most grades. Matching schools scored higher than PD schools in the third and fourth grades. However, there was a significant multivariate effect of comparison county on the combined Math and Reading/Language Arts subscales across most grades. Individual county comparisons were described within the previous grade-level results subsections. Finally, there were also significant multivariate effects of the interaction between school type (PD vs. matching schools) and county on the combined Math and Reading/Language Arts subscales. See Table 13 for further details.

Table 3: *F* Ratios, Probabilities, and Partial Eta-Squared Values on the Combined Set of Math and Reading/Language Arts Subscale Scores by Grade

Grade	School Type			County			School Type * County		
	F	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
3	5.067	.006	.006	3.145	.000	.019	*	*	*
4	4.683	.009	.005	5.429	.000	.044	3.814	.000	.031
5	*	*	*	3.645	.000	.013	1.633	.044	.006
6	*	*	*	3.785	.000	.009	3.483	.000	.008
7	*	*	*	3.518	.000	.008	3.169	.000	.007
8	*	*	*	4.001	.000	.009	2.129	.004	.005
10	*	*	*	*	*	*	*	*	*

Note: * = No significant multivariate effect was reported; *F* = F ratio statistic; *p* = exact statistical significance probability; η_p^2= partial eta-squared value.

CONCLUSIONS

This section contains a discussion of the statistical results that were presented in the Findings section and offers recommendations. Conclusions for both surveys and student achievement data are presented, followed by recommendations.

Surveys

Conclusions regarding professional staffs' perceptions of school quality are summarized by individual survey and by the overall survey set. However, the difference between the response rates of PD and matching schools makes it difficult to compare the survey results (PD schools = 1,076 or 65.3%; matching schools = 874 or 34.7%). It is likely that because matching schools did not have incentives to participate in completing the three surveys, some decided not to respond.

MSCI. Comparisons of ratings about school capacity for improvement resulted in lower mean ratings for schools implementing the professional development framework than for matching schools. Specifically, professional development schools scored lower on the extent to which faculty believe there are equitable practices in the school, the extent to which faculty believe all students can perform well academically, the extent to which faculty believe instructional techniques can be modified for students of all types, the extent to which staff believe the curriculum is coordinated within and across grade levels, and the extent to which staff believe they have access to technical resources. The PD schools were so designated because they have higher percentages of African American and low-socioeconomic students. Once designated, they were required to implement the professional development framework to reduce achievement gaps. The fact that PD school staffs perceived their schools' capacity to improve less positively than matching schools supports the theory underlying HB 4669—that schools with higher percentages of at-risk students may also merit additional support. An alternative explanation may be that nonrespondents from matching schools may have provided data that would be significantly different from those of respondents.

POSC. Comparisons of ratings about perceptions of school culture resulted in lower mean ratings for schools implementing the professional development framework than for matching schools. This was true of perceptions about the extent to which there were collaborative working relationships; a student-centered vision, mission, and policies; student and teacher responsibility for learning; and an inviting physical environment in which to learn and work when compared to matching schools. Specifically, professional development schools scored lower on staff perceptions of having a shared sense of responsibility and collaboration with other staff and the extent to which written and unwritten norms and standards are in alignment. Professional development school staffs also rated their schools lower than matching schools on faculty beliefs concerning the motivation levels of their students and their own ability to impact student learning. Finally, professional development school staffs rated their schools lower than matching schools on the overall appearance and feel of the schools in which they worked, and they also perceived parents and students at their schools to be less involved in

attempts to impact the school program overall. Thus, perceptions of school culture were significantly less positive at PD schools. The fact that PD school staffs perceived their schools' culture less positively than matching schools also supports the theory underlying HB 4669—that schools with higher percentages of at-risk students may merit additional support. An alternative explanation may be that nonrespondents from matching schools may have provided data that would be significantly different from those of respondents.

CSIQ. For subscales related to continuous school improvement, schools implementing the professional development school staffs rated their schools significantly lower than matching schools on staff dimensions related to the extent to which the school promotes learning by all, the extent to which parents and community members feel involved, the extent to which the school has clear and focused goals that are understood by all members of the community, the extent to which student assessment is meaningful, the extent to which teacher practice is aligned with research on effective teaching, and the extent to which staff perceive the schools' curriculum to be aligned and balanced. The fact that PD school staffs' perceptions regarding their school's continuous improvement are less positive than matching schools also supports the theory underlying HB 4669—that schools with higher percentages of at-risk students may also merit additional support. An alternative explanation may be that nonrespondents from matching schools may have provided data that would be significantly different from those of respondents.

Overall survey set. The PD schools, which are starting to implement the professional development framework, were rated by their professional staffs significantly lower than matching schools on nearly all subscales related to potential to become high-performing learning communities (MSCI), overall school culture (POSC), and staff perception of the school's performance and improvement (CSIQ). Although the statistical analyses adjust for differences within and between groups, an ideal study would have similar response rates from both groups. Because matched schools had no incentive to participate (other than receiving school-level profiles), the basis for comparison was not ideal. Thus, interpretations must be made with caution.

In the first year of the professional development schools demonstration project, these results confirm superintendents' selections of the designated professional development schools. PD schools had higher percentages of African American and low-socioeconomic students. Also, PD schools scored significantly lower than matching schools across all of the surveys and on many of the subscales within those surveys. This can be interpreted as each PD school's staff having a significantly lower opinion of their school, its culture, and its capacity to improve than the matching schools. This should not be a surprise, as the superintendents intended to select schools with higher percentages of at-risk students to become model schools.

Furthermore, the lower ratings on both the capacity for improvement and perceptions of school culture would naturally be associated with lower ratings on continuous school improvement. That is, professional staff members who perceive that their school has a low capacity for improvement and does not have a positive culture are not likely to rate their schools as meeting the high standards set for continuous school improvement. This is also expected in baseline studies.

Student Achievement Data

Within the sample of 60 schools, there are no mean differences between PD and matching schools. Typically, this would be expected when matching schools were purposefully selected. However, based on this subset of schools, this finding might not support the program theory underlying the House Bill—there might not be any statistically significant differences between the selected PD and matching schools. There are other competing explanations for this phenomenon. It may also be that the relatively small percentages of African American and low-socioeconomic students are being masked by the percentages of low-socioeconomic students from the corresponding matching schools. Or, perhaps, if there are theoretically significant differences between the selected PD and matching schools, the differences do not show up within the sample of 60 schools but might show up on a state-level analysis.

Students' performance on WESTEST Math and Reading/Language Arts subscales was related more strongly to the county in which they live, rather than attendance at a PD or a matching school. This phenomenon might occur because circumstances are fairly uniform across the county. So perhaps the academic difficulties experienced by African American students from PD schools, however real, may not significantly differ from the academic difficulties experienced by low-socioeconomic students in matching schools. Therefore, the significant differences that surfaced among students were associated with the county means and might relate more to the disadvantages shared by all students of a county than to ethnic or socioeconomic factors.

Recommendations

The first and primary recommendation, especially if the WVDE intends to continue an evaluation throughout the life of the demonstration program, is to find a way to provide incentives for matching schools to participate in the surveys. The closer the samples are in size and composition, the more accurate interpretations of any data analyses will be.

Survey results reported here indicate a need for policies and initiatives aimed at improving the school programs and related perceptions. Specifically, interventions should emphasize (1) increasing staff communication and collaboration, and students' levels of motivation and capacity to be successful; (2) increasing student, parent, and community involvement; (3) creating an environment that enhances students' ability to learn and teachers' ability to teach effectively; and finally, (4) increasing teachers' abilities to impact student learning through needed resources, increased training, and/or developing a more supportive and collaborative teaching community among staff.

The stated purpose of this baseline study was to investigate whether there are any significant differences between PD and matching schools' student achievement, so evaluators could account for such differences in future evaluations. The evaluators suggest that other perspectives could be used to examine the data. If possible, analyses could investigate

differences among both student- and school-level variables. It would be desirable to identify the characteristics within the counties that highly correlate with achievement.

Finally, this evaluation assesses only a subset of the variables within the PD school framework. Additional constructs are outside the purview of this baseline year. As program components develop, data could be gathered about their implementation and effectiveness.

Now that a baseline has been established, it is important to continue data collection so trends can be identified. Annual snapshots of staff perceptions and student achievement can yield data for identifying areas of strength and weakness, as well as areas of decline or improvement. Such information could be invaluable in informing changes to the program that would enhance the end product. This is consistent with the diagnostic intent of formative evaluation approaches.

REFERENCES

Boyles, J. (2004, June 12). Marple faces new challenges to bridge the achievement gap. *Charleston Daily Mail,* p. 2A.

Education Trust. (n.d.). *African American achievement in America.* Retrieved August 9, 2004, from http://www2.edtrust.org/NR/rdonlyres/9AB4AC88-7301-43FF-81A3-EB94807B917F/0/AfAmer_Achivement.pdf

Fink, A. (1993). *Evaluation fundamentals: Guiding health programs, research, and policy*. Newbury Park, CA: Sage.

Gay, L. R. (1996). *Educational research: Competencies for analysis and application* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Ginsberg, J. (2004, March 28). Working to narrow the gap. *The Charleston Gazette.* Retrieved August 25, 2004, from http://www.wvgazette.com/section/Series Brown+v.+Board of Education/200403278?pt=0

Kusimo, P. S., Petty-Wilson, P., & Body, T. O. (2004). *Governor's Minority Students Strategies Council: Minority students' achievement in West Virginia.* Charleston, WV: Office of the Secretary of Education and the Arts, State of West Virginia.

National Center for Education Statistics. (n.d.). *CCD public school data 2003-2004 school year.* Retrieved December 20, 2004, from http://nces.ed.gov

Shavelson, R. J. (1996). *Statistical reasoning for the behavioral sciences* (3rd ed.). Boston: Allyn & Bacon.

Southern Regional Education Board. (2004, May). *Legislative report, Report No. 3.* Retrieved August 25, 2004, from http://www.sreb.org/main/LegAction/legrept/2004Report/04S03-Leg._Report_3.pdf

State of West Virginia Governor's Minority Students Strategy Council. (n.d.). *Welcome to the Governor's Minority Students Strategies Council web page.* Retrieved August 26, 2004, from http://www.allwvstudents.org/

Tabachnick, B. G., & Fidell, L.S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn & Bacon.

Weinfurt, K. P. (1995). Multivariate analysis of variance. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 245 – 276). Washington, DC: American Psychological Association.

West Virginia Code §18-2E-3g (2004).

West Virginia Education Association. (2004). *WVEA legislative update.* Retrieved August 11, 2004, from http://www.wvea.org/government/legislativeupdate.htm

West Virginia Legislature. (March 13, 2004). *Enrolled H. B. 4669.* Retrieved August 23, 2004, from http://129.71.164.29/Bill_Text_HTML/2004_SESSIONS/RS/BILLS/hb4669%20enr.htm

APPENDIXES

APPENDIX A

Evaluation Instrument Descriptions

Measure of School Capacity for Improvement

The Measure of School Capacity for Improvement (MSCI) is a 58-item questionnaire that is administered to a school's professional staff—administrators, teachers, teachers' aides, librarians, counselors, and any other staff who have significant contact with students and parents. The instrument measures professional staff members' perceptions of how their school is faring in seven different areas related to capacity for improvement. Staff members use a 6-point Likert-type scale to state how true items are of their school or the frequency with which items are true for their school. Completing the MSCI requires approximately 20 minutes.

The 58 items of the MSCI compose seven subscales that encompass important aspects of capacity for improvement. Subscales, which have been validated in a large-scale, nationwide (U.S.A.) study, are composed of varying numbers of items. The MSCI and all its subscales possess strong reliability (Cronbach's alphas from .77 to .94 for the seven subscales, .97 for the overall instrument) and have demonstrated validity.

Equity in Practice. The 15 items that compose the *Equity in Practice* scale assess equitable practices in the school, specifically responsive pedagogy and antidiscriminatory practices. This scale examines the school's capacity to create an atmosphere of tolerance, cultural awareness, and equity for all learners.

Expectations for Student Performance. The *Expectations for Student Performance* scale contains 11 items. The items assess staff members' expectations of the students and their beliefs that all students can perform well academically.

Differentiated Instruction. *Differentiated Instruction*, formed by 11 items, addresses instructional practices and strategies for reaching students of diverse learning needs. The scale focuses on using or modifying instructional practices to be effective with students of all types.

Improvement Program Coherence. The *Improvement Program Coherence* scale is composed of nine items pertaining to the extent to which improvement initiatives and efforts at a school are coordinated. The items focus on the coordination of improvement programs or initiatives with existing initiatives and with school improvement goals. Items also focus on school-level support of and for improvement initiatives.

Peer-Reviewed Practice. The four items that make up the *Peer-Reviewed Practice* scale explore the observation and review by staff of their peers' work. All items assess the extent to which professional staffs in a school observe the work of their colleagues and give or receive relevant feedback about their performance.

Coordinated Curriculum. The *Coordinated Curriculum* scale, composed of four items, addresses the coordination of curriculum within and across grade levels at the school.

Technical Resources. The four items on the *Technical Resources* scale concern instructional resources and materials, including whether staff possess or have immediate access to adequate materials and resources to achieve instructional objectives.

Perceptions Of School Culture (POSC)

POSC includes 62 randomly ordered items that compose six subscales: four that contain 13 items each, and two that containing 5 items each. Each POSC item contributes to the score of only one subscale.

Collaborative Working Relationships (13 items). This subscale reflects the extent to which faculty work together, trust and respect each other, have open channels of communication, and share leadership and responsibility for problem solving and decision making.

Student-Centered Vision, Mission, and Policies (13 items). This subscale indicates the degree to which the school's vision, mission, goals, and policies are clear and consistent with each other; incorporate high expectations for all students; and are communicated to staff, students, and parents. It also indicates the extent to which the school uses measurable goals and data-based decision making.

Student Responsibility for Learning (13 items). This subscale measures faculty perceptions of their students' intrinsic motivation, persistence, awareness of their own learning strengths, and control over their own learning. It also indicates faculty perceptions of the strength of parents' belief in the importance of student effort and parent support.

Teacher Responsibility for Learning (13 items). This subscale reflects the degree to which faculty strive to improve teaching and learning, at both the individual and collective levels, and share responsibility for high levels of student learning. It also indicates the extent to which teachers accommodate students' different learning styles and encourage student collaboration and self-motivation.

Inviting Physical Environment (5 items). This subscale indicates the extent to which the school's physical environment is perceived as clean, safe, and attractive. It also reflects the degree to which the school makes visitors comfortable by having a welcoming entrance or helpful signs.

Students and Parents as Decision Makers (5 items). This subscale assesses the degree to which students and parents participate in planning and decision making that impact the school program. It also reflects the school's efforts to promote students' engagement with their own learning.

Continuous School Improvement Questionnaire (CSIQ)

This 70-item, machine scannable, field-tested and validated instrument helps school professional staff gauge its performance on seven dimensions related to continuous school improvement, which make up the seven subscales (Meehan, Cowley, Craig, Balow, & Childers, 2002) described below.

Shared Leadership. This subscale reflects how much leadership is viewed as being shared. It assesses whether school administrators dominate decision making or if there are mechanisms for involving teachers, students, and parents. Opportunities for leadership development among the members of the school community are assessed, as are the degree to which information is shared and the extent to which school administrators listen and solicit the input of others.

Effective Teaching. This subscale ascertains the extent to which teacher practice is aligned with research on effective teaching. It assesses whether teachers actively engage students in a variety of learning tasks, pose questions that encourage reflection and higher order thinking, expect students to think critically, and use teaching strategies designed to motivate students.

School/Family/Community Connections. This subscale assesses the degree to which parents and community members are involved and feel a part of the school. It reflects the degrees to which they are kept informed, meaningful partnerships exist, communication is open, and diverse points of view are honored and respected.

Purposeful Student Assessment. This subscale reflects the extent to which student assessment data are meaningful; are used by teachers to guide instructional decisions; and are communicated to and understood by the greater school community, including teachers, parents, students, and other members of the community.

Shared Goals for Learning. This subscale assesses the extent to which the school has clear, focused goals that are understood by all members of the school community. In addition, it reflects whether shared goals affect what is taught and how teachers teach, drive decisions about resources, focus on results for students, and are developed and "owned" by many rather than a few.

Learning Culture. This subscale reflects whether the culture of the school promotes learning by all—students, staff, and administration. It reflects the extent to which the school emphasizes learning rather than passive compliance, is a safe but exciting place to be, and encourages curiosity and exploration. In addition, it indicates the extent to which teachers have opportunities and encouragement to reflect on practice, work with others, and try new ways of teaching.

Aligned and Balanced Curriculum. This scale reflects the extent to which professional staff members perceive the school's curriculum to be aligned and balanced. It assesses the principal's involvement in the monitoring of the curriculum alignment process, the lesson plans of teachers, and activities in classrooms. Also, this scale measures teachers' access to curriculum resources,

use of student achievement data in curriculum emphasis, and how subjects/courses are balanced across grades.

Source: Meehan, M. L., Cowley, K. S., Craig, J. R, Balow, N., & Childers, R. D. (2002). *AEL Continuous School Improvement Questionnaire: User manual and technical report (CSIQ)*. Charleston, WV: AEL.

APPENDIX B

Response Rates by School Type, School Name, Number of Staff, and Survey

Table B: Response Rates by School Type, School Name, Number of Staff, and Survey

Type	School Name	No. of Staff	Response Rate per Survey			Overall Teacher Response rate
			MSCI	POSC	CSIQ	
Match	Orchard View Intermediate	40	23%	23%	25%	23%
PD	Eagle Intermediate	40	60%	60%	58%	59%
PD	Martinsburg North Middle	59	49%	0%	49%	33%
Match	Moorefield Middle	38	68%	68%	66%	68%
Match	Hedgesville Middle	49	71%	65%	69%	69%
PD	Martinsburg South Middle	61	70%	67%	69%	69%
Match	Meadows Elementary	18	39%	39%	39%	39%
PD	Spring Hill Elementary	47	81%	77%	77%	78%
PD	Cammack Middle	34	50%	35%	35%	40%
Match	Enslow Middle	32	0%	0%	0%	0%
Match	Miller Elementary	19	0%	0%	0%	0%
PD	Peyton Elementary	19	84%	79%	79%	81%
PD	Collins Middle	50	74%	74%	74%	74%
Match	Mullens Middle	18	0%	0%	0%	0%
PD	Montgomery Middle	21	38%	38%	38%	38%
Match	Fayetteville Middle	18	89%	83%	83%	85%
PD	Mount Hope High	28	86%	86%	82%	85%
Match	Fayetteville High	23	61%	57%	52%	57%
PD	Charles Town Middle	68	84%	85%	82%	84%
Match	Sheperdstown Middle	42	50%	50%	50%	50%
PD	North Jefferson	33	94%	94%	94%	94%
Match	Ranson Elementary	35	57%	57%	57%	57%
Match	Sheperdstown Elementary	25	60%	60%	60%	60%
PD	Wright Denny Elementary	29	83%	83%	83%	83%
Mateh	Piedmont Elementary	24	17%	17%	17%	17%
PD	Glenwood Elementary	22	41%	41%	41%	41%
Match	J.E. Robins Elementary	21	57%	52%	52%	54%
PD	Chandler Elementary	23	43%	43%	43%	43%
PD	Stonewall Jackson Middle	52	37%	37%	35%	36%
Match	Dunbar Middle	33	70%	64%	64%	66%
Match	East Dale Elementary	34	0%	0%	0%	0%
PD	East Park Elementary	39	82%	79%	79%	80%
Match	East Fairmont Junior High	29	62%	62%	62%	62%
PD	Dunbar Middle/Miller Junior High	51	80%	80%	80%	80%

(Table B continues)

Table B (continued)

Type	School Name	No. of Staff	Response Rate per Survey			Overall Teacher Response rate
			MSCI	POSC	CSIQ	
Match	Jayenne Elementary	19	89%	89%	89%	89%
PD	Watson Elementary	33	88%	88%	88%	88%
PD	Fall River Elementary	21	62%	62%	62%	62%
Match	Montcalm Elementary	28	11%	11%	11%	11%
PD	Mount View Middle	23	0%	0%	0%	0%
Match	Summers Middle	29	24%	24%	24%	24%
PD	Sandy River Middle	23	74%	74%	74%	74%
Match	Pineville Middle	22	73%	73%	73%	73%
PD	Bluefield Intermediate	31	68%	71%	68%	69%
Match	Straley Elementary	18	61%	61%	61%	61%
PD	Bluefield Middle	54	74%	74%	74%	74%
Match	Princeton Middle	50	36%	34%	32%	34%
Match	Athens Elementary	35	71%	63%	63%	66%
PD	Lashmeet/Matoaka Elementary	36	89%	89%	89%	89%
Match	Elm Grove Elementary	29	62%	62%	62%	62%
PD	Madison Elementary	23	100%	100%	100%	100%
Match	Woodsdale Elementary	28	21%	21%	21%	21%
PD	Ritchie Elementary	24	88%	88%	88%	88%
Match	Bridge Street Middle	33	100%	100%	97%	99%
PD	Wheeling Middle	25	64%	64%	64%	64%
PD	Beckley/Stratton Middle	58	55%	50%	48%	51%
Match	Park Middle	36	44%	42%	39%	42%
PD	Stratton Elementary	23	43%	39%	39%	41%
Match	Crescent Elementary	20	0%	0%	0%	0%
PD	Beckley Elementary	26	88%	88%	81%	86%
Match	Kimball Elementary	29	0%	0%	0%	0%
All schools		**1,950**	**56%**	**54%**	**54%**	**55%**

APPENDIX C

Subscale Descriptive Statistics and Reliability Estimates for Total Sample by Survey

Table C: Subscale Descriptive Statistics and Reliability Estimates for Total Sample

Measures and Subscales	*N*	*M*	*SD*	Range		Items	Cronbach's Alpha
				Min	Max		
MSCI							
Equity in Practice	1094	5.04	0.66	2.07	6.00	15	.92
Expectations for Student Performance	1106	4.67	0.82	1.55	6.00	11	.94
Differentiated Instruction	1104	4.71	0.82	1.27	6.00	11	.94
Improvement Program Coherence	1094	4.33	0.83	1.50	6.00	9	.84
Peer-Reviewed Practice	1101	3.44	1.39	1.00	6.00	4	.87
Coordinated Curriculum	1099	4.17	1.11	1.00	6.00	4	.78
Technical Resources	1114	4.51	0.97	1.25	6.00	4	.81
POSC							
Collaborative Working Relationships	978	51.19	10.42	13	65	13	.96
Student-Centered Vision, Mission, and Policies	971	54.13	8.70	16	65	13	.96
Student Responsibility for Learning	958	43.51	9.54	13	65	13	.95
Teacher Responsibility for Learning	989	52.10	8.57	17	65	13	.95
Inviting Physical Environment	1034	19.88	4.24	6	25	5	.88
Students and Parents as Decision Makers	994	16.31	3.96	5	25	5	.86
CSIQ							
Learning Culture	1050	4.77	0.84	1.20	6.00	10	.92
School/Family/Community Connections	1041	4.54	0.97	1.20	6.00	10	.94
Shared Leadership	1047	4.66	1.06	1.00	6.00	10	.96
Shared Goals for Learning	1038	4.80	0.85	1.30	6.00	10	.93
Purposeful Student Assessment	1043	4.79	0.87	1.30	6.00	10	.93
Effective Teaching	1049	4.78	0.83	1.50	6.00	10	.95
Aligned and Balanced Curriculum	1015	4.51	0.94	1.20	6.00	10	.91

Note: Internal consistency reliability estimates and descriptive statistics were calculated for each subscale of the Measure of School Capacity for Improvement (MSCI, 7 subscales), the Perceptions Of School Culture (POSC, 6 subscales), and the Continuous School Improvement Questionnaire (CSIQ, 7 subscales).

APPENDIX D

Descriptive Statistics for MSCI Subscales by School Type and Correlations Among MSCI Subscales

Table D1: Descriptive Statistics for Subscales of the MSCI by School Type

MSCI Subscale Name	PD Schools (*n*=688)		Matching Schools (*n*=369)	
	M	*SD*	*M*	*SD*
Equity in Practice	4.98	0.67	5.14	0.63
Expectations for Student Performance	4.55	0.85	4.88	0.76
Differentiated Instruction	4.64	0.85	4.85	0.77
Improvement Program Coherence	4.29	0.83	4.39	0.83
Peer Reviewed Practice	3.42	1.38	3.51	1.41
Coordinated Curriculum	4.13	1.13	4.30	1.05
Technical Resources	4.42	0.10	4.68	0.90

Table D2: Correlations Between Subscales of the MSCI

MSCI Subscale Name	1	2	3	4	5	6	7
1. Equity in Practice		.62	.72	.63	.34	.45	.40
2. Expectations for Student Performance			.80	.62	.36	.52	.52
3. Differentiated Instruction				.69	.39	.56	.52
4. Improvement Program Coherence					.53	.67	.58
5. Peer Reviewed Practice						.51	.30
6. Coordinated Curriculum							.53
7. Technical Resources							

Note: p <.001. Sample sizes range from 1,072-1,107 respondents across subscales.

APPENDIX E

Descriptive Statistics for POSC Subscales by School Type and Correlations Among POSC Subscales

Table E1: Descriptive Statistics for Subscales of the POSC by School Type

POSC Subscale Name	PD Schools [a]		Matching Schools [b]	
	M	*SD*	*M*	*SD*
Collaborative Working Relationships	50.24	10.39	52.26	10.29
Student-Centered Vision, Mission, and Policies	53.00	8.96	55.24	8.21
Student Responsibility for Learning	41.60	9.33	46.05	9.26
Teacher Responsibility for Learning	50.91	8.75	53.47	8.07
Inviting Physical Environment	19.41	4.56	20.64	3.67
Students and Parents as Decision Makers	15.67	3.86	17.09	4.07

Note: [a] Sample sizes range from 669-680 respondents. [b] Sample sizes range from 363-368 respondents.

Table E2: Correlations Between Subscales of the POSC

POSC Subscale Name	1	2	3	4	5	6
1. Collaborative Working Relationships		.81	.68	.79	.55	.69
2. Student-Centered Vision, Mission, and Policies			.70	.83	.64	.67
3. Student Responsibility for Learning				.75	60	.79
4. Teacher Responsibility for Learning					.58	.75
5. Inviting Physical Environment						.53
6. Students and Parents as Decision Makers						

Note: p <.001. Sample sizes range from 1,066-1,071 respondents across subscales.

APPENDIX F

Descriptive Statistics for CSIQ Subscales by School Type and Correlations Among CSIQ Subscales

Table F1: Descriptive Statistics for Subscales of the CSIQ by School Type

CSIQ Subscale Name	PD Schools (*n*=651)		Matching Schools (*n*=353)	
	M	*SD*	*M*	*SD*
Learning Culture	4.68	0.84	4.90	0.81
School/Family/Community Connections	4.40	0.98	4.78	0.93
Shared Leadership	4.62	1.05	4.23	1.08
Shared Goals for Learning	4.72	0.87	4.93	0.83
Purposeful Student Assessment	4.68	0.89	4.95	0.82
Effective Teaching	4.71	0.84	4.90	0.79
Aligned and Balanced Curriculum	4.42	0.95	4.67	0.90

Table F2: Correlations Between Subscales of the CSIQ

CSIQ Subscale Name	1	2	3	4	5	6	7
1. Learning Culture		.84	.77	.87	.85	.90	.85
2. School/Family/Community Connections			.81	.86	.87	.82	.84
3. Shared Leadership				.80	.76	.70	.79
4. Shared Goals for Learning					.90	.84	.88
5. Purposeful Student Assessment						.84	.87
6. Effective Teaching							.83
7. Aligned and Balanced Curriculum							

Note: p <.001. Sample sizes range from 1,010-1,049 respondents across subscales.

APPENDIX G

Student Achievement Summary Tables

Table G1: Numbers of Students by County, School Name, School Type, and NCLB Reporting Subgroups

School Name by County	School Type	Male	Female	Asian / Pacific Islander	Black	Hispanic	American Indian / Alaska Native	White	Low-Socioeconomic Status	Migrant	Limited English Proficiency	Students With Disabilities
Berkeley												
Eagle Intermediate	PD	203	212	5	106	21	2	281	249	3	15	105
Orchard View Intermediate	Match	259	245	6	83	21	1	393	231	3	10	99
Martinsburg North Middle	PD	327	330	11	134	27	1	484	325	3	17	148
Moorefield Middle	Match	241	219	2	16	8	0	434	247	30	9	71
Martinsburg South Middle	PD	412	386	4	115	49	0	630	351	0	23	131
Hedgesville High	Match	138	136	1	1	2	0	270	45	1	3	27
Cabell												
Peyton Elementary	PD	43	43	1	27	0	0	58	73	0	0	19
Miller Elementary	Match	29	39	1	21	1	0	45	39	0	2	16
Spring Hill Elementary	PD	112	96	0	92	2	3	111	181	0	2	55
Meadows Elementary	Match	68	49	0	29	1	0	87	44	0	1	15
Cammack Middle	PD	173	175	6	102	1	0	239	158	0	6	46
Enslow Middle	Match	155	140	1	32	2	3	257	248	0	2	59
Fayette												
Collins Middle	PD	383	378	0	74	5	0	682	494	0	4	86
Mullens Middle	Match	106	89	0	7	1	0	187	133	0	0	55
Montgomery Middle	PD	129	110	0	21	0	0	218	151	0	0	34
Fayetteville Middle	Match	158	108	0	16	0	0	250	166	0	0	42
Mount Hope High	PD	129	113	0	49	2	0	191	192	0	0	40
Fayetteville High	Match	46	40	0	2	0	0	84	41	0	0	14
Jefferson												
North Jefferson Elementary	PD	77	78	0	18	8	1	128	90	0	8	33
Ranson Elementary	Match	90	72	2	37	15	0	108	93	0	12	34
Wright Denny Elementary	PD	212	191	0	56	28	1	318	117	0	22	48
Shepherdstown Elementary	Match	80	78	1	17	3	2	135	29	0	3	19
Charles Town Middle	PD	451	432	10	124	63	2	684	323	0	47	128
Shepherdstown Middle	Match	239	222	5	35	10	3	408	135	0	14	73
Kanawha												
Chandler Elementary	PD	47	28	1	45	0	0	29	73	0	1	29
J E Robins Elementary	Match	62	53	0	48	1	1	65	87	0	1	10

School Name by County	School Type	Male	Female	Asian / Pacific Islander	Black	Hispanic	American Indian / Alaska Native	White	Low-Socioeconomic Status	Migrant	Limited English Proficiency	Students With Disabilities
Glenwood Elementary	PD	46	48	0	78	0	0	16	84	0	0	24
Piedmont Year-Round Education	Match	52	56	3	75	1	0	29	88	0	4	22
Stonewall Jackson Middle	PD	315	270	2	234	1	0	348	384	0	5	104
Dunbar Middle	Match	231	218	5	150	3	0	291	275	0	6	82
Marion												
East Park Elementary	PD	117	119	1	9	0	0	226	164	0	0	54
East Dale Elementary	Match	150	159	1	4	2	0	302	135	0	0	36
Watson Elementary	PD	80	73	0	48	5	2	98	101	0	0	25
Jayenne Elementary	Match	50	49	1	10	1	0	87	43	0	0	17
Dunbar Middle	PD	168	127	0	45	0	1	249	122	0	0	40
Miller Junior High	PD	176	174	4	57	2	0	287	144	0	0	49
East Fairmont Junior High	Match	185	208	0	11	2	3	377	192	0	0	50
Mercer												
Lashmeet/Matoaka	PD	164	111	0	1	2	0	272	203	0	0	43
Athens	Match	160	157	0	3	2	0	312	127	0	2	34
Bluefield Intermediate	PD	176	161	0	123	1	2	211	247	0	0	68
Straley Elementary	Match	87	81	1	15	0	0	152	102	0	0	22
Bluefield Middle	PD	302	258	1	134	1	0	424	387	0	1	105
Princeton Middle	Match	309	285	10	39	0	1	544	309	0	2	101
McDowell												
Fall River Elementary	PD	62	45	0	1	0	0	106	99	0	0	28
Montcalm Elementary	Match	130	80	0	3	0	0	207	154	0	0	32
Sandy River Middle	PD	188	149	0	0	2	1	334	283	0	0	61
Pineville Middle	Match	145	123	0	6	0	2	260	177	0	0	43
Mount View Middle	PD	146	130	0	63	0	0	213	235	0	0	75
Summers Middle	Match	206	183	1	18	4	0	366	270	0	0	82
Ohio												
Madison Elementary	PD	77	64	1	35	3	0	102	114	0	0	31
Elm Grove Elementary	Match	79	73	0	10	1	0	141	91	0	0	29
Ritchie Elementary	PD	63	69	0	24	0	0	108	109	0	0	34
Woodsdale Elementary	Match	80	77	2	16	0	0	139	73	0	0	36
Wheeling Middle	PD	107	108	0	31	0	0	184	127	0	0	32

School Name by County	School Type	Male	Female	Asian / Pacific Islander	Black	Hispanic	American Indian / Alaska Native	White	Low-Socioeconomic Status	Migrant	Limited English Proficiency	Students With Disabilities
Bridge Street Middle	Match	194	165	2	21	0	0	336	195	0	2	39
Raleigh												
Stratton Elementary	PD	41	59	0	67	1	0	32	88	0	0	27
Crescent Elementary	Match	75	62	2	23	3	0	109	80	0	0	23
Beckley Elementary	PD	81	70	0	100	3	0	48	124	0	4	24
Kimball Elementary	Match	123	76	0	63	0	0	136	186	0	0	66
Beckley-Stratton Middle	PD	369	361	14	168	3	2	543	443	0	0	105
Park Middle	Match	220	182	0	106	5	0	291	236	0	3	55

Grade 3

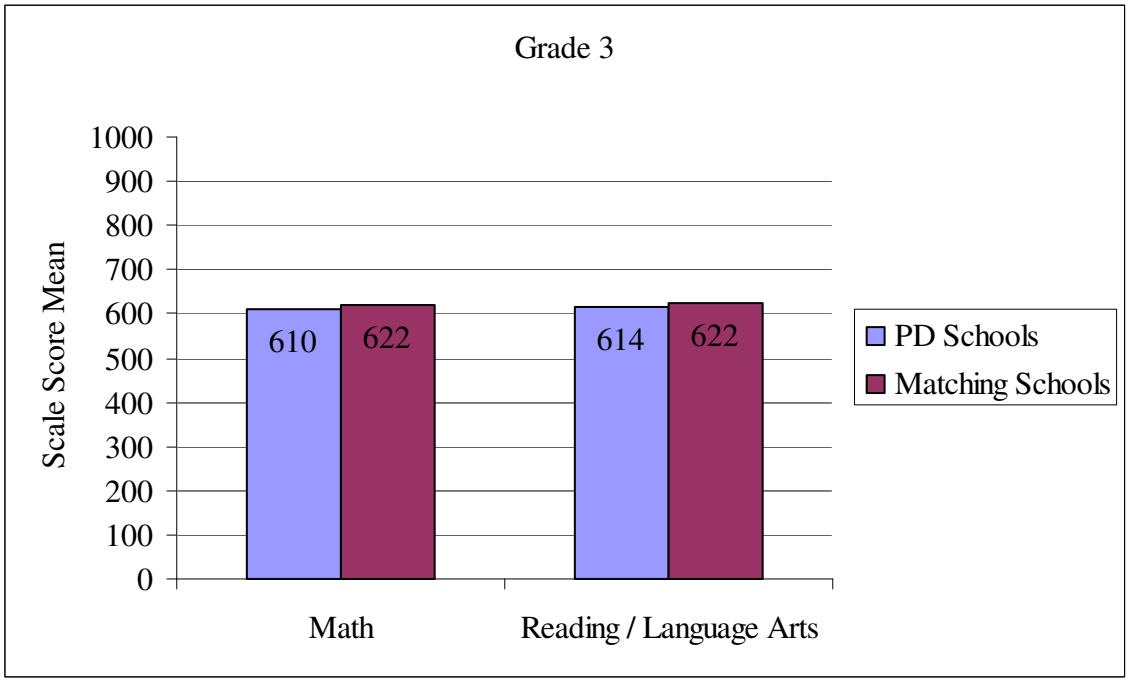

Figure G1: Grade 3 Overall PD and Matching School Mean WESTEST Subscale Scores

Table G2: Grade 3 PD and Matching School Mean WESTEST Subscale Scores by County

Subject	Type of School		Cabell	Jefferson	Kanawha	Marion	Mercer	McDowell	Ohio	Raleigh
Math	PD	*M*	604.8	616.4	588.9	624.0	607.9	609.9	612.5	595.7
		SD	73.9	72.6	84.0	30.8	87.0	25.5	98.3	96.5
		n	97	192	59	104	152	26	85	85
	Match	*M*	620.2	630.4	607.0	623.9	625.3	614.8	645.1	600.2
		SD	81.4	34.2	34.4	68.2	73.8	30.7	39.2	146.7
		n	64	101	80	125	95	48	101	94
Reading/ English Language	PD	*M*	611.7	620.0	596.7	629.0	616.3	610.8	615.8	596.3
		SD	72.8	75.2	89.6	30.8	91.8	45.2	96.6	98.5
		n	97	192	59	104	152	26	85	85
	Match	*M*	631.1	630.4	614.1	630.0	614.6	627.7	638.0	592.4
		SD	84.3	34.6	41.5	70.9	117.4	39.7	38.9	143.7
		n	64	101	80	125	95	48	101	94

MANOVA. Results from the MANOVA indicate that there is a statistically significant multivariate effect of school type (PD vs. matching schools) on the combined set of Math and

Reading/Language Arts subscale scores for third graders (Pillai's Trace=.007, $F(2, 1491)=5.067$, $p=.006$), with a strength of association size of $\eta_p^2=.006$. There is also a statistically significant multivariate effect of comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.029, $F(14, 2984)=3.145$, $p=.000$), with a strength of association size of $\eta_p^2=.019$. However, there is not a statistically significant multivariate effect of the interaction between school type and comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.013, $F(14, 2984)=1.372$, $p=.158$). Therefore, it was appropriate to compute follow-up test statistics to see which school type and counties had higher subscale means for each subscale.

Follow-Up. Separate follow-up *t* tests were computed to compare PD and matching schools on the Math and Reading/Language Arts subscales for third graders. PD and matching schools did not differ significantly on the Reading/Language Arts subscale, $t(1506)=1.852$, $p=.064$. However, matching schools scored higher than PD schools on the Math subscales, $t(1506)=3.153$, $p=.002$, with a small effect size of $d=0.09$.

A follow-up one-way ANOVA was computed to analyze the differences among county means on the Math and Reading/Language Arts subscales. The results for the ANOVA indicate that at least one county mean was significantly different than the others for both Math ($F(1,7)=4.022$, $p=.002$) and Reading/Language Arts ($F(7,1500)=3.964$, $p=.000$) subscales.

Mean difference comparisons among counties reveal which county means are significantly different from other county means on the third-grade Math subscale (see Table G3). The Raleigh County mean on the Math subscale was significantly lower than the means in Jefferson, Marion, and Ohio counties, but was not significantly different than any of the other counties. The Kanawha County mean on the Math subscale was significantly lower than the Ohio County mean, but not significantly different than any of the other counties.

Table G3: Indicator of Significantly Different Third-Grade Math Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A													
B													
C													
D													
E											E>K		
F										F<J			
G											G>K		
H													
I													
J						F<J				J>K	J>K		
K					E>K		G>K						
L													
M													

Note: p<.05; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Mean difference comparisons among counties reveal which county means are significantly different from other county means on the third-grade Reading/Language Arts subscale (see Table G4). The Raleigh County mean on the Reading/Language Arts subscale was significantly lower than the means in Jefferson, Marion, and Ohio Counties, but was not significantly different than any of the other counties.

Table G4: Indicator of Significantly Different Third-Grade Reading/Language Arts Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A													
B													
C													
D													
E											E>K		
F													
G											G>K		
H													
I													
J										J>K	J>K		
K					E>K		G>K						
L													
M													

Note: p<.05; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Grade 4

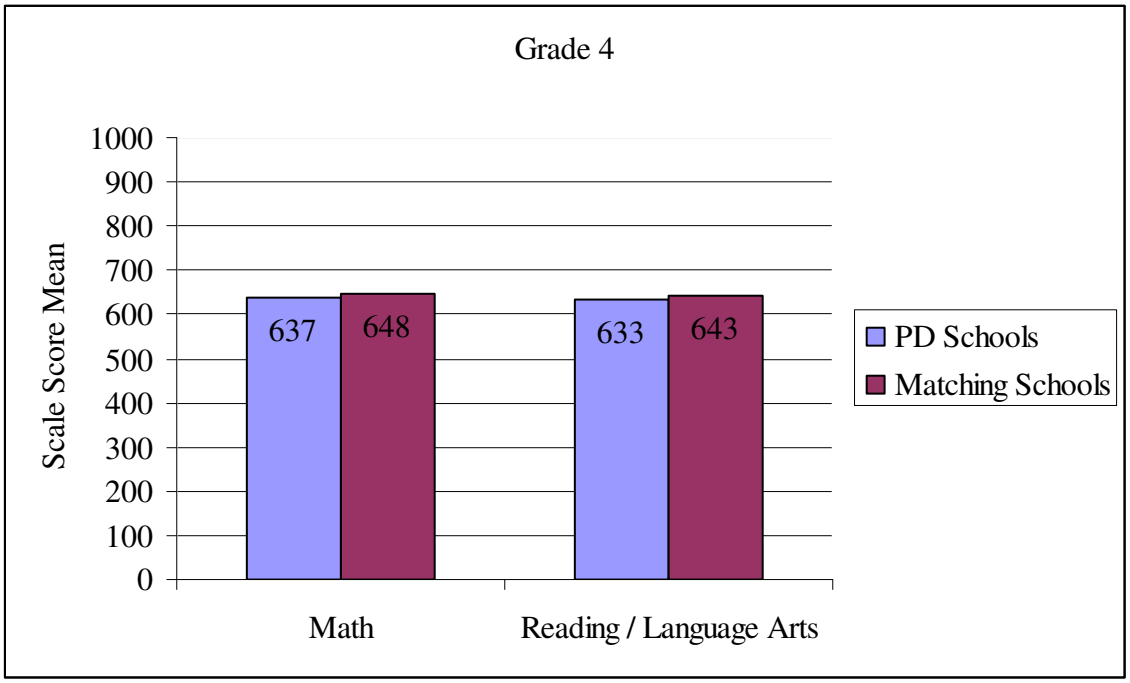

Figure G2: Grade 4 Overall PD and Matching School Mean WESTEST Subscale Scores

Table G5: Grade 4 PD and Matching School Mean WESTEST Subscale Scores by County

Subject	Type of school		Berkeley	Cabell	Fayette	Jefferson	Kanawha	Marion	Mercer	McDowell	Ohio	Raleigh
Math	PD	*M*	622.4	637.2	.	645.1	615.2	642.5	642.8	633.5	651.7	632.5
		SD	119.3	34.7	.	25.5	89.4	30.9	59.4	25.5	30.9	101.2
		n	194	91	.	178	54	147	150	28	82	81
	Match	*M*	650.3	629.2	589.0	660.3	638.9	648.1	629.3	657.5	655.7	652.8
		SD	78.0	86.4	.	27.8	31.0	83.8	95.5	35.2	98.6	36.6
		n	249	58	1	113	68	130	101	51	98	96
Reading /English Language	PD	*M*	615.6	639.7	.	647.7	617.0	643.1	635.6	622.4	636.0	620.8
		SD	120.0	35.9	.	35.9	90.7	38.0	80.6	26.2	37.0	100.7
		n	194	91	.	178	54	147	150	28	82	81
	Match	*M*	647.3	636.2	533.0	654.2	628.6	640.3	644.3	652.6	640.6	635.1
		SD	75.8	89.2	.	28.6	40.9	85.2	73.0	27.7	96.7	45.1
		n	249	58	1	113	68	130	101	51	98	96

MANOVA. Results from the MANOVA indicate a statistically significant multivariate effect of school type (PD vs. matching schools) on the combined set of Math and Reading/Language Arts subscale scores for fourth graders (Pillai's Trace=.005, $F(2, 1950)=4.683$, $p=.009$), with a strength of association size of $\eta_p^2=.005$. There is also a statistically significant multivariate effect of comparison county on the the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.044, $F(16, 3902)=5.429$, $p=.000$), with a strength of association size ($\eta_p^2=.022$). Furthermore, there is a statistically significant multivariate effect of the interaction between school type and comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.031, $F(16,3902)=3.814$, $p=.000$), with a strength of association size of $\eta_p^2=.015$. Therefore, it was appropriate to compute follow-up test statistics to see which school type and counties had higher subscale means for each subscale.

Follow-Up. Separate follow-up *t* tests were computed to compare PD and matching schools on the Math and Reading/Language Arts subscales for fourth graders. Matching schools scored higher than PD schools on the Math subscales, $t(1967)=3.453$, $p=.001$, with a small effect size of $d=0.14$. Matching schools also scored significantly higher on the Reading/Language Arts subscale, $t(1967)=3.164$, $p=.002$, with a small effect size of $d=0.15$.

A follow-up one-way ANOVA was computed to analyze the differences among fourth-grade county means on the Math and Reading/Language Arts subscales. The results for the ANOVA indicate that at least one county mean was significantly than the others for both Math ($F(8,1960)=2.405$, $p=.014$) and Reading/Language Arts ($F(8,1960)=2.333$, $p=.017$) subscales.

Mean difference comparison among counties reveals that no county means are significantly different from other county means on the fourth-grade Math subscale. However, mean difference comparisons among counties reveal which county means are significantly different from other county means on the fourth-grade Reading/Language Arts subscale (see Table G6). The Kanawha County mean on the Reading/Language Arts subscale was significantly lower than the Jefferson County mean, but was not significantly different than any of the other counties.

Table G6: Indicator of Significantly Different Fourth-Grade Reading/Language Arts Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A													
B													
C													
D													
E						E>F							
F					E>F								
G													
H													
I													
J													
K													
L													
M													

Note: p<.05; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Grade 5

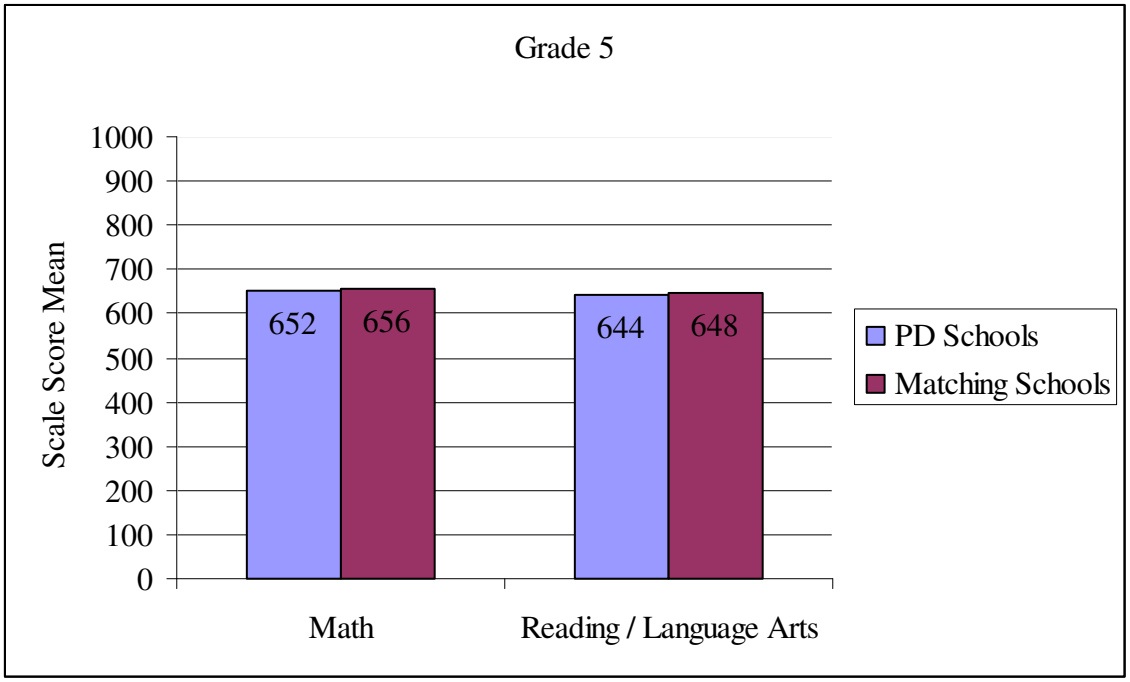

Figure G3: Grade 5 Overall PD and Matching School Mean WESTEST Subscale Scores

Table G7: Grade 5 PD and Matching School Mean WESTEST Subscale Scores by County

Subject	Type of school		Berkeley	Cabell	Fayette	Jefferson	Kanawha	Marion	Mercer	McDowell	Ohio	Raleigh
Math	PD	*M*	652.9	646.6	647.5	660.1	600.0	655.0	650.1	660.2	670.0	654.1
		SD	69.2	71.9	82.9	77.2	167.9	74.7	79.3	26.9	25.7	73.4
		n	221	106	266	188	56	199	149	22	106	85
	Match	*M*	657.8	662.7	645.6	676.5	643.4	667.5	669.8	645.1	653.1	633.9
		SD	73.0	120.7	38.2	28.1	81.3	28.2	80.4	95.0	110.2	145.4
		n	369	63	39	106	75	77	85	114	110	99
Reading /English Language	PD	*M*	647.6	640.2	643.4	654.4	588.6	647.1	642.8	645.6	653.1	634.5
		SD	70.1	94.1	82.9	75.8	169.0	86.4	79.9	40.9	31.0	76.5
		n	221	106	266	188	56	199	149	22	106	85
	Match	*M*	651.0	651.2	640.8	664.4	637.3	662.5	660.0	644.5	643.4	614.9
		SD	73.0	116.8	47.1	28.1	82.2	36.0	77.6	93.5	109.0	144.9
		n	369	63	39	106	75	77	85	114	110	99

MANOVA. Results from the MANOVA indicate that there is no statistically significant multivariate effect of school type (PD vs. matching schools) on the combined set of Math and Reading/Language Arts subscale scores for fifth graders (Pillai's Trace=.001, $F(2,2514)=1.604$, $p=.201$). However, there is a statistically significant multivariate effect of comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.026, $F(18, 5030)=3.645$, $p=.000$), with a strength of association size of $\eta_p^2=.013$. Furthermore, there is a statistically significant multivariate effect of the interaction between school type and comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.012, $F(18, 5030)=1.633$, $p=.044$), with a strength of association size of $\eta_p^2=.006$. Therefore, it was appropriate to compute follow-up test statistics to see which counties had higher subscale means for each subscale.

Follow-Up. A follow-up one-way ANOVA was computed to analyze the differences among fifth-grade county means on the Math and Reading/Language Arts subscales. The results for the ANOVA indicate that at least one county mean was significantly different than the others for both Math ($F(9,2525)=3.518$, $p=.000$) and Reading/Language Arts ($F(9,2525)=4.158$, $p=.000$) subscales.

Mean difference comparisons among counties reveals which county means are significantly different from other county means on the fifth-grade Math subscale (see Table G8). The Kanawha County mean on the Math subscale was significantly lower than Berkeley Jefferson, Marion, Mercer, and Ohio County means, but was not significantly different than any of the other counties.

Table G8: Indicator of Significantly Different Fifth-Grade Math Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A						A>F							
B													
C													
D													
E						E>F							
F	A>F				E>F		G>F	H>F		J>F			
G						G>F							
H						H>F							
I													
J						J>F							
K													
L													
M													

Note: p<.05; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Mean difference comparisons among counties reveals which county means are significantly different from other county means on the fifth-grade Reading/Language Arts subscale (see Table G9). The Kanawha County mean on the Reading/Language Arts subscale was significantly lower than Berkeley, Jefferson, Marion, Mercer, and Ohio County means, but was not significantly different than any of the other counties. The Raleigh County mean was significantly lower than the Berkeley, Jefferson, and Marion County means, but was not significantly different than any of the other counties.

Table G9: Indicator of Significantly Different Fifth-Grade Reading/Language Arts Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A						A>F					A>K		
B													
C													
D													
E						E>F					E>K		
F	A>F				E>F		G>F	H>F		J>F			
G						G>F					G>K		
H						H>F							
I													
J						J>F							
K	A>K				E>K		G>K						
L													
M													

Note: p<.05; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Grade 6

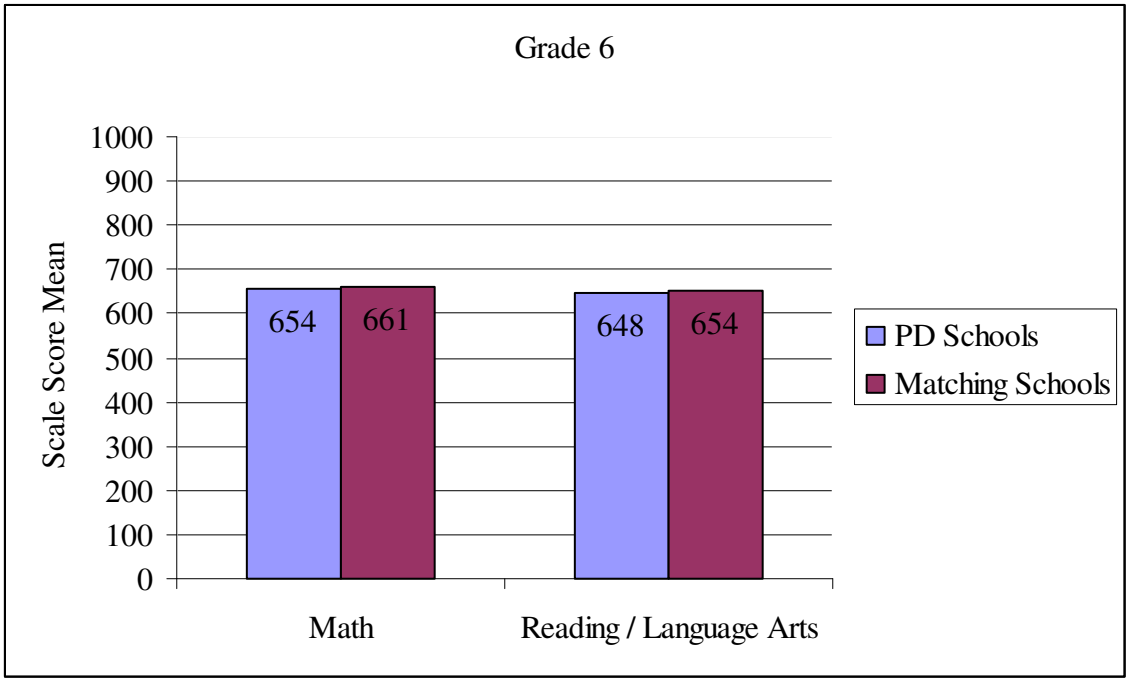

Figure G4: Grade 6 Overall PD and Matching School Mean WESTEST Subscale Scores

Table G10: Grade 6 PD and Matching school Mean WESTEST Subscale Scores by County

Subject	Type of School		Berkeley	Cabell	Fayette	Jefferson	Kanawha	Marion	Mercer	McDowell	Ohio	Raleigh
Math	PD	*M*	646.0	678.3	656.9	655.2	621.7	670.1	663.4	651.4	681.1	649.4
		SD	110.7	83.0	71.6	121.5	147.1	85.7	94.5	100.4	37.3	103.9
		n	488	103	336	301	199	234	237	134	63	239
	Match	*M*	658.9	638.3	657.6	680.1	644.1	678.5	653.8	671.9	694.3	648.4
		SD	112.3	114.3	70.3	67.4	107.8	34.3	123.8	95.0	34.6	130.5
		n	107	104	139	152	162	76	252	242	104	189
Reading /English Language	PD	*M*	648.0	667.2	655.0	650.4	616.8	645.9	653.7	636.9	671.3	649.0
		SD	103.9	74.3	71.1	112.8	144.1	117.2	82.0	101.5	27.8	91.1
		n	488	103	336	301	199	234	237	134	63	239
	Match	*M*	653.9	640.6	656.7	668.5	645.4	678.1	645.0	664.1	665.1	636.4
		SD	110.7	94.7	40.5	62.0	108.1	37.0	121.2	80.3	26.9	128.2
		n	107	104	139	152	162	76	252	242	104	189

MANOVA. Results from the MANOVA indicate that there is no statistically significant multivariate effect of school type (PD vs. matching schools) on the combined set of Math and Reading/Language Arts subscale scores for sixth graders (Pillai's Trace=.001, $F(2,3840)=1.383$, $p=.251$). However, there is a statistically significant multivariate effect of comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.018, $F(18,7682)=3.785$, $p=.000$), with a strength of association size of $\eta_p^2=.009$. Furthermore, there is a statistically significant multivariate effect of the interaction between school type and comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.016, $F(18,7682)=3.483$, $p=.000$), with a strength of association size of $\eta_p^2=.008$. Therefore, it was appropriate to compute follow-up test statistics to see which counties had higher subscale means for each subscale.

Follow-Up. A follow-up one-way ANOVA was computed to analyze the differences among sixth-grade county means on the Math and Reading/Language Arts subscales. The results for the ANOVA indicate that at least one county mean was significantly than the others for both Math ($F(9,3851)=6.227$, $p=.000$) and Reading/Language Arts ($F(9,3851)=3.066$, $p=.001$) subscales.

Mean difference comparisons among counties reveal which county means are significantly different from other county means on the sixth-grade Math subscale (see Table G11). The Kanawha County mean on the Reading/Language Arts subscale was significantly lower than Berkeley, Jefferson, Marion, Mercer, McDowell, and Ohio County means, but was not significantly different than any of the other counties. The Ohio County mean was significantly higher than the Berkeley, Fayette, Kanawha, Mercer, and Raleigh County means, but was not significantly different than any of the other counties. Finally, the Marion County

mean was significantly higher than the Berkeley County mean, but was not significantly different than any of the other counties.

Table G11: Indicator of Significantly Different Sixth-Grade Math Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A							G>A			J>A			
B													
C						C>F				J>C			
D													
E						E>F							
F			C>F		E>F		G>F	H>F	I>F	J>F			
G	G>A					G>F							
H						H>F				J>H			
I						I>F							
J	J>A		J>C			J>F		J>H			J>K		
K										J>K			
L													
M													

Note: p<.05; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Mean difference comparisons among counties reveal which county means are significantly different from other county means on the sixth-grade Reading/Language Arts subscale (see Table G12). The Kanawha County mean on the Reading/Language Arts subscale was significantly lower than Fayette, Jefferson, McDowell, and Ohio County means, but was not significantly different than any of the other counties.

Table G12: Indicator of Significantly Different Sixth-Grade Reading/Language Arts Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A													
B													
C						C>F							
D													
E						E>F							
F			C>F		E>F				I>F	J>F			
G													
H													
I						I>F							
J						J>F							
K													
L													
M													

Note: p<.05; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Grade 7

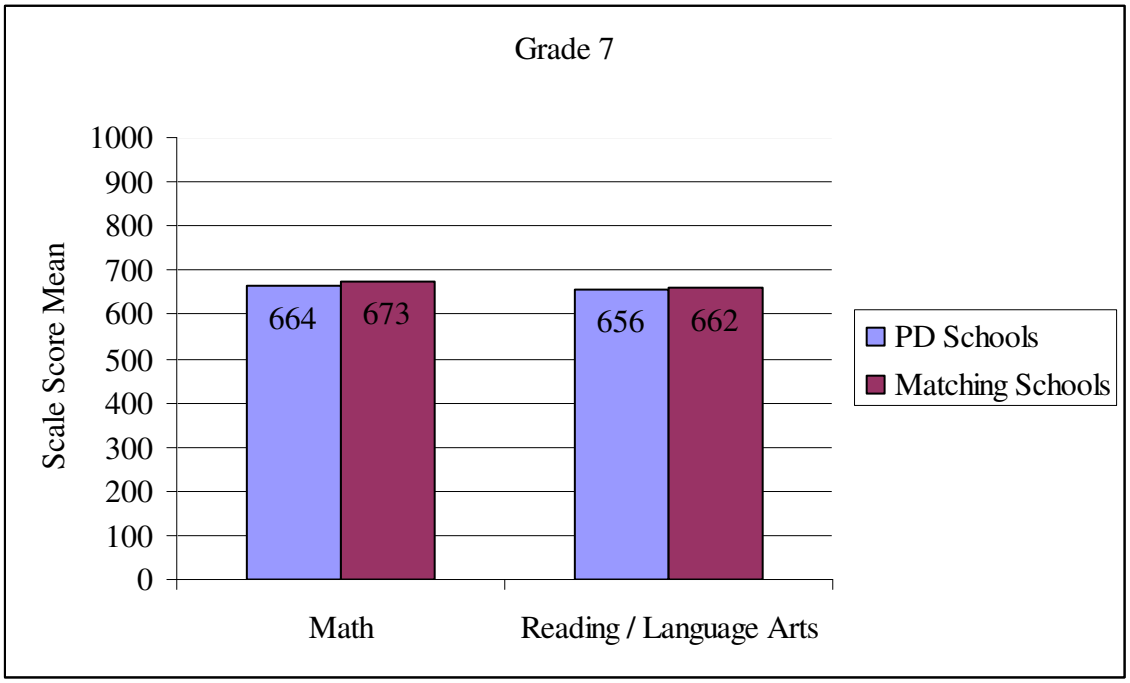

Figure G5: Grade 7 Overall PD and Matching School Mean WESTEST Subscale Scores

Table G13: Grade 7 PD and Matching School Mean WESTEST Subscale Scores by County

Subject	Type of school		Berkeley	Cabell	Fayette	Jefferson	Kanawha	Marion	Mercer	McDowell	Ohio	Raleigh
Math	PD	*M*	662.2	684.5	662.3	675.7	652.1	690.1	655.2	636.1	673.1	670.9
		SD	126.7	94.3	109.6	91.1	126.1	64.4	111.7	129.3	85.3	112.1
		n	492	128	289	272	182	170	268	259	74	238
	Match	*M*	656.5	656.5	663.1	691.2	655.5	678.3	670.6	678.9	693.0	674.5
		SD	127.8	114.9	105.5	66.8	139.2	90.3	88.2	77.2	71.7	87.4
		n	114	106	135	149	143	203	259	210	116	137
Reading/ English Language	PD	*M*	652.9	678.3	655.7	671.8	649.4	679.7	653.2	624.2	659.0	656.0
		SD	128.9	93.1	108.6	68.7	124.5	61.6	112.0	125.2	86.8	120.8
		n	492	128	289	272	182	170	268	259	74	238
	Match	*M*	650.3	649.8	644.5	685.6	640.7	661.2	665.5	674.2	669.3	669.9
		SD	126.0	110.7	120.5	34.8	136.4	102.6	87.9	78.3	70.0	72.4
		n	114	106	135	149	143	203	259	210	116	137

MANOVA. Results from the MANOVA indicate that there is no statistically significant multivariate effect of school type (PD vs. matching schools) on the combined set of Math and Reading/Language Arts subscale scores for seventh graders (Pillai's Trace=.001, $F(2,3923)=1.990$, $p=.137$). However, there is a statistically significant multivariate effect of comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.016, $F(18,7848)=3.518$, $p=.000$), with a strength of association size of $\eta_p^2=.008$. Furthermore, there is a statistically significant multivariate effect of the interaction between school type and comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.014, $F(18,7848)=3.169$, $p=.000$), with a strength of association size of $\eta_p^2=.007$. Therefore, it was appropriate to compute follow-up test statistics to see which counties had higher subscale means for each subscale.

Follow-Up. A follow-up one-way ANOVA was computed to analyze the differences among seventh-grade county means on the Math and Reading/Language Arts subscales. The results for the ANOVA indicate that at least one county mean was significantly different than the others for both Math ($F(9,3934)=4.215$, $p=.000$) and Reading/Language Arts ($F(9,3934)=3.686$, $p=.000$) subscales.

Mean difference comparisons among counties reveals which county means are significantly different from other county means on the seventh-grade Math subscale (see Table G14). The McDowell County mean was significantly lower than the Jefferson, Marion, and Ohio County means, but was not significantly different than any of the other counties. The Kanawha County mean was significantly lower than the Marion and Jefferson County means, but was not significantly different than any of the other counties.

Table G14: Indicator of Significantly Different Seventh-Grade Math Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A													
B													
C													
D													
E						E>F			E>I				
F					E>F		G>F						
G						G>F			G>I				
H													
I					E>I		G>I			J>I			
J									J>I				
K													
L													
M													

Note: p<.05; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Mean difference comparisons among counties reveals which county means are significantly different from other county means on the seventh-grade Reading/Language Arts subscale (see Table G15). The Jefferson County mean was significantly higher than the Berkeley, Fayette, Kanawha, and McDowell County means.

Table G15: Indicator of Significantly Different Seventh-Grade Reading/Language Arts Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A					E>A								
B													
C					E>C								
D													
E	E>A		E>C			E>F			E>I				
F					E>F								
G													
H													
I					E>I								
J													
K													
L													
M													

Note: p<.05; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Grade 8

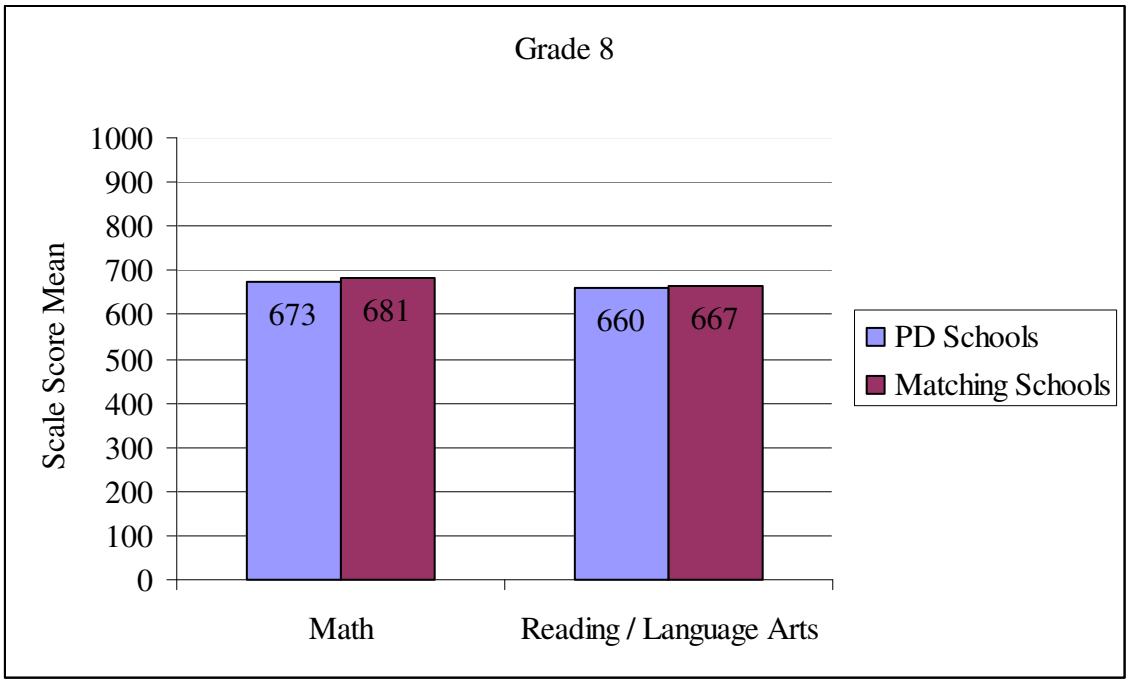

Figure G6: Grade 8 Overall PD and Matching School Mean WESTEST Subscale Scores

Table G16: Grade 8 PD and Matching School Mean WESTEST Subscale Scores by County

Subject	Type of school		Berkeley	Cabell	Fayette	Jefferson	Kanawha	Marion	Mercer	McDowell	Ohio	Raleigh
Math	PD	*M*	666.7	679.8	671.1	664.5	643.5	712.7	676.5	664.2	697.2	683.4
		SD	135.6	119.8	111.1	139.2	158.9	71.7	92.2	73.7	92.3	112.8
		n	475	117	304	310	204	180	216	251	78	253
	Match	*M*	653.5	661.1	674.3	702.9	664.3	705.1	680.8	683.7	686.8	674.4
		SD	159.4	110.6	115.9	37.5	155.0	41.4	107.4	91.8	89.0	115.3
		n	125	85	147	160	144	190	287	202	139	123
Reading /English Language	PD	*M*	655.8	665.5	658.3	657.2	636.9	680.9	666.7	655.5	681.0	669.1
		SD	135.4	144.6	113.6	129.6	155.8	94.1	87.5	69.4	85.4	101.3
		n	475	117	304	310	204	180	216	251	78	253
	Match	*M*	649.4	655.6	662.7	680.7	642.5	688.8	673.9	665.8	669.4	658.7
		SD	157.7	106.9	113.6	62.8	158.8	30.9	94.4	101.0	84.6	126.2
		n	125	85	147	160	144	190	287	202	139	123

MANOVA. Results from the MANOVA indicate that there is no statistically significant multivariate effect of school type (PD vs. matching schools) on the combined set of Math and Reading/Language Arts subscale scores for eighth graders (Pillai's Trace=.000, $F(2,3969)=.251$, $p=.778$). However, there is a statistically significant multivariate effect of comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.018, $F(18,7940)=4.001$, $p=.000$), with a strength of association size of $\eta_p^2=.009$. Furthermore, there is a statistically significant multivariate effect of the interaction between school type and comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.010, $F(18,7940)=2.129$, $p=.004$), with a strength of association size of $\eta_p^2=.005$. Therefore, it was appropriate to compute follow-up test statistics to see which counties had higher subscale means for each subscale.

Follow-Up. A follow-up one-way ANOVA was computed to analyze the differences among sixth-grade county means on the Math and Reading/Language Arts subscales. The results for the ANOVA indicate that at least one county mean was significantly different than the others for both Math ($F(9,3980)=6.526$, $p=.000$) and Reading/Language Arts ($F(9,3980)=4.261$, $p=.000$) subscales.

Mean difference comparisons among counties reveals which county means are significantly different from other county means on the eighth-grade Math subscale (see Table G17). The Marion County mean was significantly higher than the Berkeley, Cabell, Fayette, Jefferson, Kanawha, Mercer, McDowell, and Raleigh County means. The Kanawha County mean was significantly lower than the Marion, Mercer, Ohio, and Raleigh County means.

Table G17: Indicator of Significantly Different Eighth-Grade Math Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A							G>A						
B							G>B						
C							G>C						
D													
E							G>E						
F							G>F	H>F		J>F	K>F		
G	G>A	G>B	G>C		G>E	G>F		G>H	G>I				
H						H>F	G>H						
I							G>I						
J						J>F							
K						K>F							
L													
M													

Note: $p<.05$; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Mean difference comparisons among counties reveal which county means are significantly different from other county means on the eighth-grade Reading/Language Arts subscale (see Table G18). The Kanawha County mean was significantly lower than the Marion, Mercer, and Ohio County means, but was not significantly different than any of the other counties. The Marion County mean was significantly higher than the Kanawha and Berkeley County means, but was not significantly different than any of the other counties.

Table G18: Indicator of Significantly Different Eighth-Grade Reading/Language Arts Subscale County Means by County

County	A	B	C	D	E	F	G	H	I	J	K	L	M
A							G>A						
B													
C													
D													
E													
F							G>F	H>F		J>F			
G	G>A					G>F							
H						H>F							
I													
J						J>F							
K													
L													
M													

Note: p<.05; A=Berkeley; B=Cabell; C=Fayette; D=Hardy; E=Jefferson; F=Kanawha; G=Marion; H=Mercer; I=McDowell; J=Ohio; K=Raleigh; L=Summers; M=Wyoming; expressions are intentionally redundant

Grade 10

Table G19: Grade 10 PD and Matching School Mean WESTEST Subscale Scores by County

Subject	Type of school		Berekeley	Fayette
Math	PD	*M*	.	676.1
		SD	.	148.5
		n	.	47
	Match	*M*	700.5	700.1
		SD	141.1	88.3
		n	274	86
Reading /English Language	PD	*M*	.	653.7
		SD	.	146.0
		n	.	47
	Match	*M*	674.3	681.8
		SD	133.7	52.0
		n	274	86

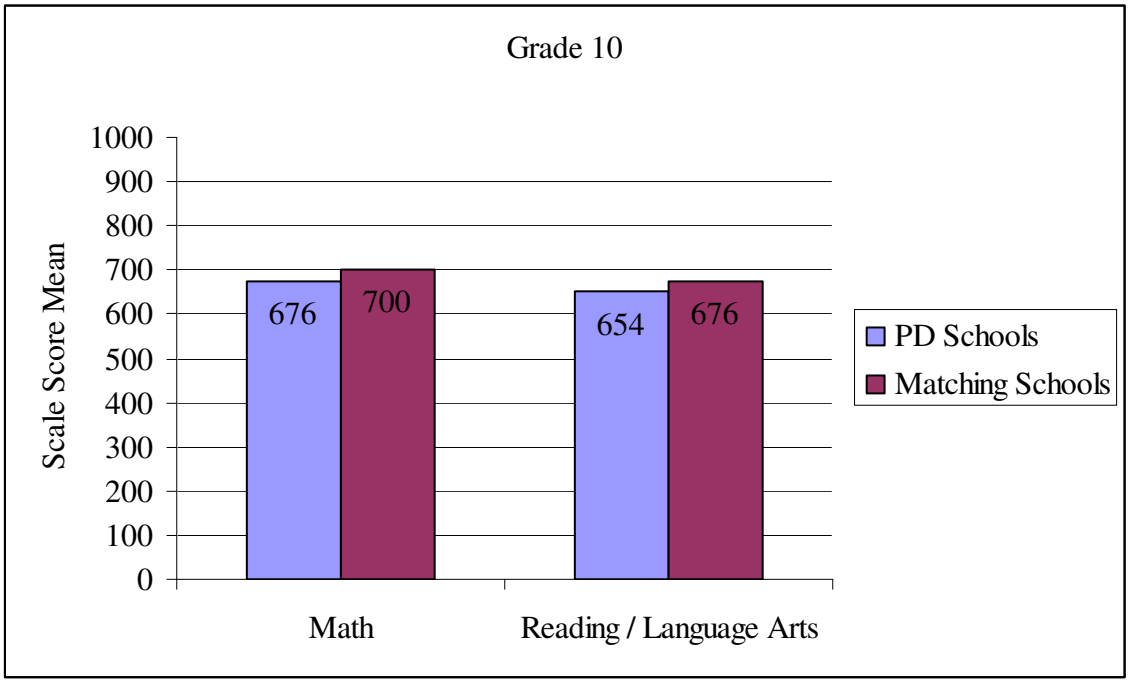

Figure G7: Grade 10 Overall PD and Matching School Mean WESTEST Subscale Scores

MANOVA. Results from the MANOVA indicate that there is no statistically significant multivariate effect of school type (PD vs. matching schools) on the combined set of Math and Reading/Language Arts subscale scores for eighth graders (Pillai's Trace=.005, $F(2,403)=1.020$, $p=.361$).There is no statistically significant multivariate effect of comparison county on the combined set of Math and Reading/Language Arts subscale scores (Pillai's Trace=.007, $F(2,403)=1.450$, $p=.236$). Therefore, there was also no statistically significant multivariate effect of interaction effect of school type by comparison county effects. Such results warrant no follow-up analyses.