

PA Leadership Standards

Core Standards

- I. Knowledge and skills to think and plan strategically creating an organizational vision around personalized student success.
- II. An understanding of standards-based systems theory and design and the ability to transfer that knowledge to the leader's job as an architect of standards based reform in the school.
- III. The ability to access and use appropriate data to inform decision-making at all levels of the system.

Corollary Standards

- I. Creating a culture of teaching and learning with an emphasis on learning.
- II. Managing resources for effective results.
- III. Collaborating, communicating, engaging and empowering others inside and outside the organization to pursue excellence in learning.
- IV. Operating in a fair and equitable manner with personal and professional integrity.
- V. Advocating for children and public education in the larger political, social, economic, legal and cultural context.
- VI. Supporting professional growth of self and others through practice and inquiry.

Alignment of Meaningful Interventions					
Characteristics of High Performing Schools	Theory of Action	Turnaround Principles	School Level Guiding Questions	Alignment to the Danielson Framework for Teaching	PA Inspired Leaders
Clear and Shared Focus	Strategy: Commit to a comprehensive, cohesive, agency-wide vision, mission, and set of aligned activities to turn around the state’s lowest-performing schools.		SLGQ12: Is there a system within the school that fully ensures the school’s resources effectively address instructional priorities that are aligned with the school’s vision and mission. SLGQ13: Is there a system within the school that fully ensures that each member of the school community promotes, enhances and sustains a shared vision of positive school climate?	Planning and Preparation: 1c: Setting instructional outcomes	Strategic/Cultural Leadership: 1a: Creates an organizational vision, mission, and strategic goals Systems Leadership: 2a: Leverages human and financial resources
High Standards and Expectations	Accountability: Develop an accountability system that sets clear standards and performance targets for schools. LEAs and the SEA, monitors and reports on progress, and incentivizes dramatic reform.		SLGQ6: Is there a system within the school that fully ensures at least 95% of the students who are required to participate in PSSA/Keystone testing do so for both Math and Reading?	Planning and Preparation: 1b: Demonstrating knowledge of students Classroom Environment: 2b: Establishing a culture for learning Instruction: 3a: Communicating with students 3b: Using questioning and discussion techniques	Systems Leadership: 2d: Establishes and implements expectations for students and staff Leadership for Learning: 3d: Sets high expectations for all students
Effective Leadership	Human Capital: Invest in highly effective teachers and leaders to drive turnaround at the district and school levels	Providing strong leadership by (1) reviewing the performance of the current principal; (2) either replacing the principal	SLGQ10: Is there a system within the school that fully ensures teachers and administrators receive timely, effective support	Professional Responsibilities 4f: Showing professionalism	Strategic/Cultural Leadership: 1d: Leads change efforts for continuous improvement 1e: Celebrates accomplishments

Alignment of Meaningful Interventions					
Characteristics of High Performing Schools	Theory of Action	Turnaround Principles	School Level Guiding Questions	Alignment to the Danielson Framework for Teaching	PA Inspired Leaders
		if such a change is necessary to ensure strong and effective leadership, or demonstrating to the SEA that the current principal has a track record in improving achievement and has the ability to lead the turnaround effort; (3) providing the principal with operational flexibility in the areas of scheduling, staff, curriculum, and budget	and intervention as needed? SLGQ11: Is there a system within the school that fully ensures classrooms are staffed with highly qualified teachers and that student needs drive decisions about teacher placement?		and acknowledges failures Systems Leadership: 2c: Complies with Federal, State, and LEA mandates Leadership for Learning: 3a: Leads school improvement initiatives Professionalism and Community Leadership: 4b: Shows professionalism
High Levels of Collaboration and Communication	Communication: Design a coherent agency-wide turnaround message and communicate it consistently.	(v) using data to inform instruction and for continuous improvement, including by providing time for collaboration for on the use of data	SLGQ3: Is there a system within the school that fully ensures students who are academically at risk are identified early and are supported by a process that provides interventions based upon student needs and includes procedures for monitoring effectiveness? SLGQ9: Is there a system within the school that fully ensures teachers and administrators meet on a regular basis to use multiple data sources to reflect on the progress of student learning as it relates to the effectiveness of professional practice?	Professional Responsibilities 4d: Participating in a professional community	Strategic/Cultural Leadership: 1c: Builds a collaborative and empowering work environment Systems Leadership: 2e: Communicates effectively and strategically

Alignment of Meaningful Interventions					
Characteristics of High Performing Schools	Theory of Action	Turnaround Principles	School Level Guiding Questions	Alignment to the Danielson Framework for Teaching	PA Inspired Leaders
Curriculum, Instruction and Assessment Aligned to Standards		(iv) strengthening the school’s instructional program based on student needs and ensuring that the instructional program is research-based, rigorous, and aligned with State academic content standards	SLGQ1: Is there a system within the school that fully ensures consistent implementation of standards-aligned curricula across all classrooms for all students? SLGQ3: Is there a system within the school that fully ensures school staff members use standards-aligned assessments to monitor student achievement and adjust instructional practices?	Planning and Preparation 1a: Demonstrating knowledge of content and pedagogy 1d: Demonstrating knowledge of resources 1e: Designing coherent instruction Instruction 3e: Demonstrating flexibility and responsiveness	Leadership for Learning: 3b: Aligns curricula, instruction, and assessments 3c: Implements high quality instruction
Frequent Monitoring of Teaching and Learning		(ii) ensuring that teachers are effective and able to improve instruction by: (1) reviewing the quality of all staff and retaining only those who are determined to be effective and have the ability to be successful in the turnaround effort; (2) preventing ineffective teachers from transferring to these schools;	SLGQ2: Is there a system within the school that fully ensures the consistent implementation of effective instructional practices across all classrooms?	Planning and Preparation: 1f: Designing student assessments Instruction 3d: Using assessment in instruction Professional Responsibilities: 4a: Reflecting on teaching 4b: Maintaining accurate records	Strategic/Cultural Leadership: 1b: Uses data for informed decision making
Focused Professional Development			SLGQ8: Is there a system within the school that fully ensures professional development is focused, strategic and implemented with fidelity?	Professional Responsibilities 4e: Growing and developing professionally	Professionalism and Community Leadership: 4c: Supports professional growth
Supportive Learning Environment		(iii) redesigning the school day, week, or year to include	SLGQ5: Is there a system within the school that fully ensures	Classroom Environment 2a: Creating an environment of	Systems Leadership: Ensures School Safety

Alignment of Meaningful Interventions					
Characteristics of High Performing Schools	Theory of Action	Turnaround Principles	School Level Guiding Questions	Alignment to the Danielson Framework for Teaching	PA Inspired Leaders
		additional time for student learning and teacher collaboration (vi) establishing a school environment that improves school safety and discipline and addressing other non-academic factors that impact student achievement, such as students' social, emotional, and health needs	school staff members monitor attendance and student participation in the learning process and respond with classroom and school-wide interventions when students are chronically absent or disengaged? SLGQ7: (High Schools Only – Graduation Rate) Is there a system within the school that fully ensures students who enter the school as 9 th graders will complete the academic program and graduate in four years?	respect and rapport 2c: Managing classroom procedures 2d: Managing student behavior 2e: Organizing physical space Instruction 3c: Engaging students in learning	Systems Leadership: 2f: Manages conflict constructively Leadership for Learning: 3e: Maximizes instructional time
High Levels of Community and Parent Involvement		(vii) providing ongoing mechanisms for family and community engagement	SLGQ14: Is there a system within the school that fully ensures the school has partnered with families and the community to support student participation in the learning process?	Professional Responsibilities 4c: Communicating with families	Professional and Community Leadership 4a: Maximizes parent and community involvement and outreach

PUBLIC SCHOOL CODE OF 1949 - OMNIBUS AMENDMENTS
Act of Jun. 30, 2012, P.L. 684, No. 82 **Cl. 24**
Session of 2012
No. 2012-82

HB 1901

AN ACT

Amending the act of March 10, 1949 (P.L.30, No.14), entitled "An act relating to the public school system, including certain provisions applicable as well to private and parochial schools; amending, revising, consolidating and changing the laws relating thereto," in preliminary provisions, further providing for definitions, for background checks of prospective employees and conviction of employees of certain offenses and for collection of identifying information of students attending institutions of higher education; providing for adjustments based on Consumer Price Index and for Keystone Exams; in school districts, further providing for moratorium on certain data collection systems and data sets; in school finances, providing for reopening of 2012-2013 budget and for intergovernmental agreements for school security and safety; in grounds and buildings, providing for limitation on new applications for the Department of Education's approval of public school building projects and further providing for work to be done under contract let on bids and exception; in books, furniture and supplies, further providing for purchase of supplies; in district superintendents and assistant district superintendents, further providing for eligibility and for manner of election or approval, providing for performance review and further providing for election of assistant district superintendents, for term and salary of assistants, for commissions and for removal; in professional employees, further providing for rating system and for causes for suspension; in pupils and attendance, further providing for liability for tuition and enforcement of payment and for school lunch and breakfast reimbursement; in safe schools, further providing for regulations; adding a requirement relating to cardiopulmonary resuscitation; providing for open campus initiatives; in high schools, further providing for attendance in other districts; providing for disclosure by school entities of certain interscholastic athletic opportunity information; reenacting and amending provisions relating to school boards and educational empowerment; in community colleges, further providing for financial program, reimbursement of payments; in Thaddeus Stevens College of Technology, further providing for contracts for construction, repair, renovation or maintenance; in State System of Higher Education, further providing for project contracts and for powers and duties of institution presidents; in school districts of the first class, further providing for superintendents of schools or buildings and of supplies; in funding for public libraries, providing for State and for fiscal year 2012-2013; in reimbursements by Commonwealth and between school districts, providing for basic education funding for 2011-2012 school year, further providing for payments to intermediate units and for special education payments to school districts, providing for assistance to school districts certified as education empowerment districts, further providing for Pennsylvania accountability grants and providing for targeted industry cluster certificate scholarship program; and making editorial changes.

The General Assembly of the Commonwealth of Pennsylvania hereby enacts as follows:

Section 1. Section 102 of the act of March 10, 1949 (P.L.30, No.14), known as the Public School Code of 1949, is amended by adding a definition to read:

Section 102. Definitions.--When used in this act the following words and phrases shall have the following meanings:

* * *

"Keystone Exam." An assessment developed or caused to be developed by the Department of Education pursuant to 22 Pa. Code § 4.51(f) (relating to State assessment system).

* * *

Section 2. Section 111 heading, (a.1), (e), (f.1), (f.2), (h), (i) and (j) of the act, amended or added July 11, 2006 (P.L.1092, No.114) and June 30, 2011 (P.L.112, No.24), are amended to read:

Section 111. [Background Checks of] **Criminal History of Employes and Prospective Employes; Conviction of [Employes of] Certain Offenses.**--* * *

(a.1) Beginning April 1, 2007, this section shall apply to all **current and** prospective employes of public and private schools, intermediate units and area vocational-technical schools, including, but not limited to, teachers, substitutes, janitors, cafeteria workers, independent contractors and their employes, except those employes and independent contractors and their employes who have no direct contact with children.

(1) Beginning April 1, 2007, this section shall apply to bus drivers **employed or** offered employment by a school district, private school, nonpublic school, intermediate unit or area vocational-technical school or by an independent contractor.

(2) Beginning April 1, 2007, this section shall apply to **student teachers and** student teacher candidates assigned to all public and private schools, intermediate units and area vocational-technical schools.

(3) For purposes of this section, **"student teacher" or "student teacher candidate"** shall mean an individual participating in a classroom teaching, internship, clinical or field experience who, as part of a program for the initial or advanced preparation of professional educators, performs classroom teaching or assists in the education program in a public or private school, intermediate unit or area vocational-technical school under the supervision of educator preparation program faculty.

(4) Prior to a **student teacher or** student teacher candidate's participation in any classroom teaching, internship, clinical or field experience, [that candidate] **the student teacher or student teacher candidate** shall provide to the administrator of his or her educator preparation program all criminal history record information required of an employe or prospective employe who is subject to this section.

(5) [The] **A student teacher or** student teacher candidate may not participate in any classroom teaching, internship, clinical or field experience if this section would prohibit an employe or prospective employe subject to this section from being employed under those circumstances.

(6) During the course of a **student teacher or** student teacher candidate's participation in an educator preparation program, the administrator of the **student teacher or** student teacher candidate's educator preparation program shall maintain a copy of the criminal history record information that was provided by the **student teacher or** student teacher candidate. The penalty provisions of subsection (g) shall be applicable to the administrator of a **student teacher or** student teacher candidate's educator preparation program.

(7) If a **student teacher or** student teacher candidate is continuously enrolled in an educator preparation program, the criminal history record information initially submitted by [that] **the student teacher or student teacher candidate** to that program shall remain valid during that period of enrollment, **subject to the requirements of subsection (j).** If a **student teacher or** student teacher candidate's enrollment in an educator preparation

program is interrupted or if [that] **the student teacher or student teacher candidate** transfers to another educator preparation program, the **student teacher or student teacher candidate** shall provide to the administrator of his or her educator preparation program all criminal history record information required of an employe who is subject to this section.

* * *

(e) No person subject to this act shall be employed **or remain employed** in a public or private school, intermediate unit or area vocational-technical school where [the] **a report of criminal history record information or a form submitted by an employe under subsection (j)** indicates the [applicant] **person** has been convicted of any of the following offenses:

(1) An offense under one or more of the following provisions of Title 18 of the Pennsylvania Consolidated Statutes:

Chapter 25 (relating to criminal homicide).

Section 2702 (relating to aggravated assault).

Section 2709.1 (relating to stalking).

Section 2901 (relating to kidnapping).

Section 2902 (relating to unlawful restraint).

Section 2910 (relating to luring a child into a motor vehicle or structure).

Section 3121 (relating to rape).

Section 3122.1 (relating to statutory sexual assault).

Section 3123 (relating to involuntary deviate sexual intercourse).

Section 3124.1 (relating to sexual assault).

Section 3124.2 (relating to institutional sexual assault).

Section 3125 (relating to aggravated indecent assault).

Section 3126 (relating to indecent assault).

Section 3127 (relating to indecent exposure).

Section 3129 (relating to sexual intercourse with animal).

Section 4302 (relating to incest).

Section 4303 (relating to concealing death of child).

Section 4304 (relating to endangering welfare of children).

Section 4305 (relating to dealing in infant children).

A felony offense under section 5902(b) (relating to prostitution and related offenses).

Section 5903(c) or (d) (relating to obscene and other sexual materials and performances).

Section 6301(a)(1) (relating to corruption of minors).

Section 6312 (relating to sexual abuse of children).

Section 6318 (relating to unlawful contact with minor).

Section 6319 (relating to solicitation of minors to traffic drugs).

Section 6320 (relating to sexual exploitation of children).

(2) An offense designated as a felony under the act of April 14, 1972 (P.L.233, No.64), known as "The Controlled Substance, Drug, Device and Cosmetic Act."

(3) An offense similar in nature to those crimes listed in clauses (1) and (2) under the laws or former laws of the United States or one of its territories or possessions, another state, the District of Columbia, the Commonwealth of Puerto Rico or a foreign nation, or under a former law of this Commonwealth.

(f.1) (1) If [the] **a report of criminal history record information or a form submitted by an employe under subsection (j)** indicates the person has been convicted of **an offense graded as a felony offense of the first, second or third degree other than [those] one of the offenses** enumerated under subsection (e), the person shall be eligible for **continued or** prospective employment only if a period of ten years has elapsed from the date of expiration of the sentence for the offense.

(2) If [the conviction is for] **a report of criminal history record information or a form submitted by an employe under subsection (j)** indicates

the person has been convicted of an offense graded as a misdemeanor of the first degree, other than one of the offenses enumerated in subsection (e), the person shall be eligible for **continued or** prospective employment only if a period of five years has elapsed from the date of expiration of the sentence for the offense.

(3) If the report of criminal history record information **or a form submitted by an employe under subsection (j)** indicates the person has been convicted more than once for an offense under 75 Pa.C.S. § 3802(a), (b), (c) or (d) (relating to driving under influence of alcohol or controlled substance) and the offense is graded as a misdemeanor of the first degree under 75 Pa.C.S. § 3803 (relating to grading), the person shall be eligible for **current or** prospective employment only if a period of three years has elapsed from the date of expiration of the sentence for the most recent offense.

(f.2) Nothing in this section shall be construed to interfere with the ability of a public or private school, intermediate unit or area vocational-technical school to make employment, discipline or termination decisions, **provided that this subsection shall not be construed to conflict with subsection (e), (f.1) or (j) (6).**

* * *

(h) [Any] **Subject to the requirements of subsection (j), any** person who has once obtained the information required under [this section] **subsections (b), (c) and (c.1)** may transfer to or provide services to another school in the same district, diocese or religious judicatory or established and supervised by the same organization and shall not be required to obtain additional reports before making such transfer.

(i) Notwithstanding subsections (b), (c) and (c.1), **and subject to the requirements of subsection (j),** administrators, before April 1, 2007, may employ in-State applicants on a provisional basis for a single period not to exceed thirty (30) days and may employ out-of-State applicants on a provisional basis for a single period not to exceed ninety (90) days and, after March 31, 2007, may employ any applicants on a provisional basis for a single period not to exceed ninety (90) days, except during a lawful strike proceeding under the provisions of the act of July 23, 1970 (P.L.563, No.195), known as the "Public Employe Relations Act," provided that all of the following conditions are met:

(1) the applicant has applied for the information required under subsection (b) and, where applicable, under subsection (c) or (c.1) and the applicant provides a copy of the appropriate completed request forms to the administrator;

(2) the administrator has no knowledge of information pertaining to the applicant which would disqualify him from employment pursuant to subsection (e) **or (f.1);**

(3) the applicant swears or affirms in writing that he is not disqualified from employment pursuant to subsection (e) **or (f.1);**

(4) if the information obtained pursuant to subsection (b), (c) or (c.1) reveals that the applicant is disqualified from employment pursuant to subsection (e) **or (f.1),** the applicant shall be suspended and subject to termination proceedings as provided for by law; and

(5) the administrator requires that the applicant not be permitted to work alone with children and that the applicant work in the immediate vicinity of a permanent employe.

(j) (1) The department shall develop a standardized form to be used by current and prospective employes of public and private schools, intermediate units and area vocational-technical schools for the written reporting by current and prospective employes of any arrest or conviction for an offense enumerated under [subsection (e)] **subsections (e) and (f.1).** The form shall provide a space in which a current or prospective employe who has not been

convicted of or arrested for any such offense will respond "no conviction" and "no arrest." The form also shall provide that failure to accurately report any arrest or conviction for an offense enumerated under subsection (e) **or (f.1)** shall subject the current or prospective employe to criminal prosecution under 18 Pa.C.S. § 4904 (relating to unsworn falsification to authorities). The department shall publish the form on its publicly accessible Internet website and in the Pennsylvania Bulletin.

(2) [Within ninety (90) days of the effective date of this subsection, all current] **All current and prospective** employes of a public or private school, intermediate unit or area vocational-technical school shall complete the form described in clause (1), indicating whether or not they have been **arrested for or** convicted of an offense enumerated under [subsection (e)] **subsections (e) and (f.1), provided that any current employe who completed the form on or before December 27, 2011, in compliance with clauses (1) and (2) on that date and who has not been arrested for or convicted of an offense enumerated under subsections (e) and (f.1) shall not be required to complete an additional form under this subsection.**

(3) If, as required in clause (2), [an] **a current or prospective** employe refuses to submit the form described in clause (1), the administrator or other person responsible for employment decisions in a school or other institution shall immediately require the **current or prospective** employe to submit to the administrator a current report of criminal history record information as required under subsections (a.1), (b) and (c.1).

(4) If the arrest or conviction for an offense enumerated under subsection (e) **or (f.1)** occurs after the effective date of this subsection, the **current or prospective** employe shall provide the administrator or designee with written notice utilizing the form provided for in clause (1) not later than seventy-two (72) hours after an arrest or conviction.

(5) If an administrator or other person responsible for employment decisions in a school or other institution has a reasonable belief that [an] **a current or prospective** employe was arrested or has a conviction for an offense required to be reported under clause (2) or (4) and the employe or prospective employe has not notified the administrator as required under this section, the administrator or other person responsible for employment decisions in a school or other institution shall immediately require the **current or prospective** employe to submit to the administrator a current report of criminal history record information as required under subsections (a.1), (b) and (c.1). The cost of the criminal background check shall be borne by the employing entity.

(6) [(i) An] **A current or prospective** employe who willfully fails to disclose a conviction or an arrest for an offense enumerated under [subsection (e)(1)] **this section** shall be subject to discipline up to and including termination or denial of employment and may be subject to criminal prosecution under 18 Pa.C.S. § 4904 (relating to unsworn falsification to authorities).

[(ii) An employe who willfully fails to disclose a conviction of any other offense required to be reported by this section may be subject to discipline and may be subject to criminal prosecution under 18 Pa.C.S. § 4904.]

Section 3. Section 118(a)(1) of the act, added June 30, 2011 (P.L.112, No.24), is amended to read:

Section 118. Collection of Identifying Information of Students Attending Institutions of Higher Education.--(a) The following provisions shall apply to the Department of Education's collection of identifying information of students:

(1) The department may collect identifying information of students only if:

(i) the department is specifically required to do so under Federal statute or regulation or under another provision of this act; or

(ii) the information is voluntarily provided by an institution of higher education.

* * *

Section 4. Section 119 of the act, added November 3, 2011 (P.L.400, No.97), is repealed:

[Section 119. Adjustments Based on Consumer Price Index.--Adjustments to the base amounts shall be made as follows:

(1) The Department of Labor and Industry shall determine the percentage change in the Consumer Price Index for All Urban Consumers: All Items (CPI-U) for the United States City Average as published by the United States Department of Labor, Bureau of Labor Statistics, for the twelve-month period ending September 30, 2012, and for each successive twelve-month period thereafter.

(2) If the department determines that there is no positive percentage change, then no adjustment to the base amounts shall occur for the relevant time period.

(3) (i) If the department determines that there is a positive percentage change in the first year that the determination is made under paragraph (1), the positive percentage change shall be multiplied by each base amount, and the products shall be added to the base amounts, respectively, and the sums shall be preliminary adjusted amounts.

(ii) The preliminary adjusted amounts shall be rounded to the nearest one hundred dollars (\$100) to determine the final adjusted base amounts.

(4) In each successive year in which there is a positive percentage change in the CPI-U for the United States City Average, the positive percentage change shall be multiplied by the most recent preliminary adjusted amounts, and the products shall be added to the preliminary adjusted amount of the prior year to calculate the preliminary adjusted amounts for the current year. The sums thereof shall be rounded to the nearest one hundred dollars (\$100) to determine the new final adjusted base amounts.

(5) The determinations and adjustments required under this section shall be made in the period between October 1 and November 15 of the year following the effective date of this section and annually between October 1 and November 15 of each year thereafter.

(6) The final adjusted base amounts and new final adjusted base amounts obtained under paragraphs (3) and (4) shall become effective January 1 for the calendar year following the year in which the determination required under paragraph (1) is made.

(7) The department shall publish notice in the Pennsylvania Bulletin prior to January 1 of each calendar year of the annual percentage change determined under paragraph (1) and the unadjusted or final adjusted base amounts determined under paragraphs (3) and (4) at which competitive bidding is required and written or telephonic price quotations are required, respectively, for the calendar year beginning the first day of January after publication of the notice. The notice shall include a written and illustrative explanation of the calculations performed by the department in establishing the unadjusted or final adjusted base amounts under this section for the ensuing calendar year.

(8) The annual increase in the preliminary adjusted base amounts obtained under paragraphs (3) and (4) shall not exceed three percent (3%).]

Section 5. The act is amended by adding sections to read:

Section 120. Adjustments Based on Consumer Price Index.--Adjustments to the base amounts shall be made as follows:

(1) The Department of Labor and Industry shall determine the percentage change in the Consumer Price Index for All Urban Consumers: All Items (CPI-U) for the United States City Average as published by the United States

Department of Labor, Bureau of Labor Statistics, for the twelve-month period ending September 30, 2012, and for each successive twelve-month period thereafter.

(2) If the Department of Labor and Industry determines that there is no positive percentage change, then no adjustment to the base amounts shall occur for the relevant time period.

(3) (i) If the Department of Labor and Industry determines that there is a positive percentage change in the first year that the determination is made under paragraph (1), the positive percentage change shall be multiplied by each base amount, and the products shall be added to the base amounts, respectively, and the sums shall be preliminary adjusted amounts.

(ii) The preliminary adjusted amounts shall be rounded to the nearest one hundred dollars (\$100) to determine the final adjusted base amounts.

(4) In each successive year in which there is a positive percentage change in the CPI-U for the United States City Average, the positive percentage change shall be multiplied by the most recent preliminary adjusted amounts, and the products shall be added to the preliminary adjusted amount of the prior year to calculate the preliminary adjusted amounts for the current year. The sums thereof shall be rounded to the nearest one hundred dollars (\$100) to determine the new final adjusted base amounts.

(5) The determinations and adjustments required under this section shall be made in the period between October 1 and November 15, 2012, and annually between October 1 and November 15 of each year thereafter.

(6) The final adjusted base amounts and new final adjusted base amounts obtained under paragraphs (3) and (4) shall become effective January 1 for the calendar year following the year in which the determination required under paragraph (1) is made.

(7) The Department of Labor and Industry shall publish notice in the Pennsylvania Bulletin prior to January 1 of each calendar year of the annual percentage change determined under paragraph (1) and the unadjusted or final adjusted base amounts determined under paragraphs (3) and (4) at which competitive bidding is required and written or telephonic price quotations are required, respectively, for the calendar year beginning the first day of January after publication of the notice. The notice shall include a written and illustrative explanation of the calculations performed by the Department of Labor and Industry in establishing the unadjusted or final adjusted base amounts under this section for the ensuing calendar year.

(8) The annual increase in the preliminary adjusted base amounts obtained under paragraphs (3) and (4) shall not exceed three percent (3%).

Section 121. Keystone Exams.--Subject to annual appropriation, not later than the 2020-2021 school year, the Department of Education shall develop and implement Keystone Exams in the following subjects: algebra I, literature, biology, English composition, algebra II, geometry, United States history, chemistry, civics and government and world history. The State Board of Education shall promulgate regulations, subject to the act of June 25, 1982 (P.L.633, No.181), known as the "Regulatory Review Act," necessary to implement this section.

Section 6. Section 221.1(a) of the act, added June 30, 2011 (P.L.112, No.24), is amended to read:

Section 221.1. Moratorium on Certain Data Collection Systems and Data Sets.--(a) For the school years 2011-2012 and 2012-2013, the Department of Education and the Department of Public Welfare shall suspend the collection of data through Pennsylvania's Enterprise to Link Information for Children Across Network (PELICAN) and the Pennsylvania Information Management System (PIMS) except as follows:

(1) Information required to meet Federal mandates in the following:

(i) The Elementary and Secondary Education Act of 1965 (Public Law 89-10, 20 U.S.C. § 6301 et seq.).

(ii) The Individuals with Disabilities Education Act (Public Law 91-230, 20 U.S.C. § 1400 et seq.).

(iii) The Educational Technical Assistance Act of 2002 (Public Law 107-279, 116 Stat. 1975).

(iv) Title VI of the America COMPETES Act or the America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science Act (Public Law 110-69, 121 Stat. 572).

(v) The American Recovery and Reinvestment Act of 2009 (Public Law 111-5, 123 Stat. 115).

(vi) The Head Start Act (Public Law 97-35, 42 U.S.C. § 9831 et seq.).

(vii) The Child Care and Development Block Grant Act of 1990 (Public Law 101-508, 42 U.S.C. § 9858 et seq.).

(viii) The Social Security Act (49 Stat. 620, 42 U.S.C. § 301 et seq.).

(ix) Any data pursuant to other Federal requirements and to meet eligibility requirements for Federal funds.

(2) Pennsylvania Value-Added Assessment System (PVAAS), including any revisions or improvements made to the system.

(3) Information required by the Department of Public Welfare to supervise, license or register a child-care provider under Articles IX and X of the act of June 13, 1967 (P.L.31, No.21), known as the "Public Welfare Code."

(4) Information relating to background checks required in section 111 and in 23 Pa.C.S. §§ 6344 (relating to information relating to prospective child-care personnel) and 6344.1 (relating to information relating to family day-care home residents).

(5) Information necessary for all payments or reimbursement by the Commonwealth.

(6) Information required to be reported pursuant to Article XIII-A of this act.

(7) Information which is voluntarily provided by an institution of higher education.

* * *

Section 7. The act is amended by adding sections to read:

Section 616. Reopening of 2012-2013 Budget.--Notwithstanding any other provisions of law, a board of school directors of a school district may reopen its 2012-2013 budget to reflect the following:

(1) Federal and State allocations for fiscal years 2011-2012 and 2012-2013 provided by the act of June 30, 2012 (P.L. , No.9A), known as the General Appropriation Act of 2012; and

(2) any increase in local revenue allocations that result from other legislation enacted by the General Assembly during the 2011 regular session.

Section 617. Intergovernmental Agreements for School Security and Safety.--The board of school directors of a school district may enter into agreements with other political subdivisions to provide for the safety and security of the school. The board of school directors may use school funds to share costs with municipalities and counties for such expenses as benefits and salaries of school resource officers and probation officers. Such officers are not required to be employees of the school district and may be employees of other political subdivisions.

Section 732.1. Limitation on New Applications for Department of Education Approval of Public School Building Projects.--(a) For the 2012-2013 fiscal year, the Department of Education shall not accept or approve new school building construction or reconstruction project applications. Completed school building construction or reconstruction project applications received by the Department of Education by October 1, 2012, are not subject to this provision.

(b) The Department of Education shall, in consultation with school district officials and the General Assembly, conduct a review of the Department of Education's current process through which public school building projects are reviewed and approved for Commonwealth reimbursement. The review shall incorporate an analysis of impacting local factors, including, but not limited to, tax effort and building requirements, and shall make recommendations to the chair and minority chair of the Appropriations Committee of the Senate, the chair and minority chair of the Education Committee of the Senate, the chair and minority chair of the Appropriations Committee of the House of Representatives and the chair and minority chair of the Education Committee of the House of Representatives by May 1, 2013.

Section 8. Sections 751(a), (a.1), (b) and (f) and 807.1 of the act, amended November 3, 2011 (P.L.400, No.97), are amended and the sections are amended by adding subsections to read:

Section 751. Work to be Done Under Contract Let on Bids; Exception.--[(a) All construction, reconstruction, repairs, maintenance or work of any nature, including the introduction of plumbing, heating and ventilating, or lighting systems, upon any school building or upon any school property, or upon any building or portion of a building leased under the provisions of section 703.1, made by any school district, where the entire cost, value, or amount of such construction, reconstruction, repairs, maintenance or work, including labor and material, shall exceed a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 119, shall be done under separate contracts to be entered into by such school district with the lowest responsible bidder, upon proper terms, after due public notice has been given asking for competitive bids. Whenever a board of school directors shall approve the use of a prefabricated unit, complete in itself, for a school building or other proper structure to be erected upon school property, the board of school directors may have prepared appropriate specifications detailing the size and material desired in a particular prefabricated unit, including all utilities such as plumbing, heating and ventilating, and electrical work, and may advertise for a single bid on all the work and award the contract therefor to the lowest responsible bidder: Provided, That if due to an emergency a school plant or any part thereof becomes unusable competitive bids for repairs or replacement may be solicited from at least three responsible bidders, and upon the approval of any of these bids by the Secretary of Education, the board of school directors may proceed at once to make the necessary repairs or replacements in accordance with the terms of said approved bid or bids.]

(a.1) Written or telephonic price quotations from at least three qualified and responsible contractors shall be requested by the board of school directors for all contracts that exceed a base amount of ten thousand dollars (\$10,000), subject to adjustment under section 119, but are less than the amount requiring advertisement and competitive bidding, or, in lieu of price quotations, a memorandum shall be kept on file showing that fewer than three qualified contractors exist in the market area within which it is practicable to obtain quotations. A written record of telephonic price quotations shall be made and shall contain at least the date of the quotation, the name of the contractor and the contractor's representative, the construction, reconstruction, repair, maintenance or work which was the subject of the quotation and the price. Written price quotations, written records of telephonic price quotations and memoranda shall be retained for a period of three years.]

(a.2) All construction, reconstruction, repairs, maintenance or work of any nature, including the introduction of plumbing, heating and ventilating, or lighting systems, upon any school building or upon any school property, or upon any building or portion of a building leased under the provisions of section 703.1, made by any school district where the entire cost, value or

amount of such construction, reconstruction, repairs, maintenance or work, including labor and material, shall exceed a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 120, shall be done under separate contracts to be entered into by such school district with the lowest responsible bidder, upon proper terms, after due public notice has been given asking for competitive bids. Whenever a board of school directors shall approve the use of a prefabricated unit, complete in itself, for a school building or other proper structure to be erected upon school property, the board of school directors may have prepared appropriate specifications detailing the size and material desired in a particular prefabricated unit, including all utilities such as plumbing, heating and ventilating, and electrical work, and may advertise for a single bid on all the work and award the contract therefor to the lowest responsible bidder: Provided, That, if due to an emergency a school plant or any part thereof becomes unusable, competitive bids for repairs or replacement may be solicited from at least three responsible bidders, and, upon the approval of any of these bids by the board of school directors, the school district may proceed at once to make the necessary repairs or replacements in accordance with the terms of said approved bid or bids; and Provided further, That the school district shall notify the Secretary of Education in a form and manner determined by the Secretary of Education that an emergency has occurred and a bid has been selected under the emergency process provided for in this section.

(a.3) Written or telephonic price quotations from at least three qualified and responsible contractors shall be requested by the board of school directors for all contracts that exceed a base amount of ten thousand dollars (\$10,000), subject to adjustment under section 120, but are less than the amount requiring advertisement and competitive bidding, or, in lieu of price quotations, a memorandum shall be kept on file showing that fewer than three qualified contractors exist in the market area within which it is practicable to obtain quotations. A written record of telephonic price quotations shall be made and shall contain at least the date of the quotation, the name of the contractor and the contractor's representative, the construction, reconstruction, repair, maintenance or work which was the subject of the quotation and the price. Written price quotations, written records of telephonic price quotations and memoranda shall be retained for a period of three years.

[(b) The board of school directors in any school district may perform any construction, reconstruction, repairs, or work of any nature, where the entire cost or value, including labor and material, is less than a base amount of ten thousand dollars (\$10,000), subject to adjustment under section 119, by its own maintenance personnel. The board of school directors in any school district may authorize the secretary of the board or other executive to award contracts for construction, reconstruction, repairs, or work of any nature, where the entire cost or value, including labor and material, subject to adjustment under section 119, is a base amount of eighteen thousand five hundred dollars (\$18,500) or less, without soliciting competitive bids, subject, however, to the provisions of subsection (a.1).]

(b.1) The board of school directors in any school district may perform any construction, reconstruction, repairs, or work of any nature where the entire cost or value, including labor and material, is less than a base amount of ten thousand dollars (\$10,000), subject to adjustment under section 120, by its own maintenance personnel. The board of school directors in any school district may authorize the secretary of the board or other executive to award contracts for construction, reconstruction, repairs, or work of any nature, where the entire cost or value, including labor and material, subject to adjustment under section 120, is a base amount of eighteen thousand five hundred dollars (\$18,500) or less, without soliciting competitive bids, subject, however, to the provisions of subsection (a.3).

* * *

[(f) No board of school directors shall evade the provisions of this section as to advertising for bids or purchasing materials or contracting for services piecemeal for the purpose of obtaining prices under a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 119, upon transactions which should, in the exercise of reasonable discretion and prudence, be conducted as one transaction amounting to more than a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 119. This provision is intended to make unlawful the practice of evading advertising requirements by making a series of purchases or contracts each for less than the advertising requirement price, or by making several simultaneous purchases or contracts each below said price, when in either case the transaction involved should have been made as one transaction for one price.]

(g) No board of school directors shall evade the provisions of this section as to advertising for bids or purchasing materials or contracting for services piecemeal for the purpose of obtaining prices under a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 120, upon transactions which should, in the exercise of reasonable discretion and prudence, be conducted as one transaction amounting to more than a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 120. This provision is intended to make unlawful the practice of evading advertising requirements by making a series of purchases or contracts each for less than the advertising requirement price, or by making several simultaneous purchases or contracts each below said price, when in either case the transaction involved should have been made as one transaction for one price.

Section 807.1. Purchase of Supplies.--[(a) All furniture, equipment, textbooks, school supplies and other appliances for the use of the public schools, costing, subject to adjustment under section 119, a base amount of eighteen thousand five hundred dollars (\$18,500) or more shall be purchased by the board of school directors only after due advertisement as hereinafter provided. Supplies costing, subject to adjustment under section 119, a base amount of eighteen thousand five hundred dollars (\$18,500) or more shall be purchased by the board of school directors only after public notice has been given by advertisement once a week for three (3) weeks in not less than two (2) newspapers of general circulation. In any district where no newspaper is published, said notice may, in lieu of such publication, be posted in at least five (5) public places.

(a.1) Written or telephonic price quotations from at least three qualified and responsible vendors shall be requested by the board of school directors for all purchases of supplies that exceed a base amount of ten thousand dollars (\$10,000), subject to adjustment under section 119, but are less than the amount requiring advertisement and competitive bidding, or, in lieu of price quotations, a memorandum shall be kept on file showing that fewer than three qualified vendors exist in the market area within which it is practicable to obtain quotations. A written record of telephonic price quotations shall be made and shall contain at least the date of the quotation, the name of the vendor and the vendor's representative, the supplies which were the subject of the quotation and the price of the supplies. Written price quotations, written records of telephonic price quotations and memoranda shall be retained for a period of three years.]

(a.2) All furniture, equipment, textbooks, school supplies and other appliances for the use of the public schools costing, subject to adjustment under section 120, a base amount of eighteen thousand five hundred dollars (\$18,500) or more shall be purchased by the board of school directors only after due advertisement as hereinafter provided. Supplies costing, subject to adjustment under section 120, a base amount of eighteen thousand five hundred dollars (\$18,500) or more shall be purchased by the board of school directors

only after public notice has been given by advertisement once a week for three (3) weeks in not less than two (2) newspapers of general circulation. In any district where no newspaper is published, said notice may, in lieu of such publication, be posted in at least five (5) public places.

(a.3) Written or telephonic price quotations from at least three (3) qualified and responsible vendors shall be requested by the board of school directors for all purchases of supplies that exceed a base amount of ten thousand dollars (\$10,000), subject to adjustment under section 120, but are less than the amount requiring advertisement and competitive bidding, or, in lieu of price quotations, a memorandum shall be kept on file showing that fewer than three (3) qualified vendors exist in the market area within which it is practicable to obtain quotations. A written record of telephonic price quotations shall be made and shall contain at least the date of the quotation, the name of the vendor and the vendor's representative, the supplies which were the subject of the quotation and the price of the supplies. Written price quotations, written records of telephonic price quotations and memoranda shall be retained for a period of three years.

[(b) The board of school directors shall accept the bid of the lowest responsible bidder, kind, quality, and material being equal, but shall have the right to reject any and all bids, or select a single item from any bid. The board of school directors in any district may authorize or appoint the secretary of the board or other executive as purchasing agent for the district, with authority to purchase supplies that cost a base amount of less than eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 119.]

(b.1) The board of school directors shall accept the bid of the lowest responsible bidder, kind, quality, and material being equal, but shall have the right to reject any and all bids or select a single item from any bid. The board of school directors in any district may authorize or appoint the secretary of the board or other executive as purchasing agent for the district, with authority to purchase supplies that cost a base amount of less than eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 120.

(c) The following shall be exempt from the above provisions: maps, music, globes, charts, educational films, filmstrips, prepared transparencies and slides, pre-recorded magnetic tapes and disc recordings, textbooks, games, toys, prepared kits, flannel board materials, flash cards, models, projectuals and teacher demonstration devices necessary for school use.

[(d) No board of school directors shall evade the provisions of this section as to advertising for bids or purchasing materials piecemeal for the purpose of obtaining prices under the base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 119, upon transactions which should, in the exercise of reasonable discretion and prudence, be conducted as one transaction amounting to more than a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 119. This provision is intended to make unlawful the practice of evading advertising requirements by making a series of purchases or contracts each for less than the advertising requirement price, or by making several simultaneous purchases or contracts each below said price, when in either case the transaction involved should have been made as one transaction for one price.]

(e) No board of school directors shall evade the provisions of this section as to advertising for bids or purchasing materials piecemeal for the purpose of obtaining prices under the base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 120, upon transactions which should, in the exercise of reasonable discretion and prudence, be conducted as one transaction amounting to more than a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 120. This provision is intended to make unlawful the

practice of evading advertising requirements by making a series of purchases or contracts each for less than the advertising requirement price, or by making several simultaneous purchases or contracts each below said price, when in either case the transaction involved should have been made as one transaction for one price.

Section 9. Section 1003 of the act, amended June 30, 2011 (P.L.112, No.24), is amended to read:

Section 1003. Eligibility.--(a) Except as otherwise provided in [subsection (b)] **subsections (b) and (b.1)**, no person shall receive a letter of eligibility or be elected or appointed as a district superintendent or assistant district superintendent, unless--

(1) He holds a diploma from a college or other institution approved by the Department of Education;

(2) He has had six (6) years' successful teaching experience, not less than three of which shall have been in a supervisory or administrative capacity;

(3) He has completed in a college or university a graduate program in education approved by the Department of Education that includes the Pennsylvania school leadership standards under section 1217. Completion of the program shall not be subject to waiver under section 1714-B unless the candidate provides to the Secretary of Education evidence that the candidate has successfully completed an equivalent leadership development program that addresses the school leadership standards under section 1217.

(4) Provided that in school districts of the first class, five (5) years of administrative experience at the level of assistant, associate or deputy superintendent, may be substituted for prescribed graduate administrative courses, and which shall be the responsibility of the Secretary of Education to review these equivalences to conform with State board regulations.

(b) Notwithstanding the requirements of subsection (a), a person shall be eligible for election or appointment as a district superintendent or assistant district superintendent if he holds a graduate degree from an accredited higher education institution in business [or finance], **finance or management** and has at least four (4) years of relevant experience in business, finance or management.

(b.1) Notwithstanding the requirements of subsection (a), a person shall be eligible for election or appointment as a district superintendent or assistant district superintendent if he holds a juris doctorate degree from an accredited law school and has at least four (4) years of relevant experience in law. This subsection shall expire three (3) years from the effective date of this subsection. A person who is issued a commission by the department based on satisfaction of the requirements of this subsection may retain his commission after the expiration of this subsection.

(b.2) The department shall, upon request in a form and manner as prescribed by the department and made available on the department's publicly accessible Internet website, confirm that an individual satisfies the requirements of subsection (b) or (b.1) and that the individual is eligible for election or appointment as a district superintendent or assistant district superintendent. Upon a school district's hiring of an individual who satisfies the requirements of subsection (b) or (b.1), the department shall issue the individual a commission.

(c) Notwithstanding the provisions of sections 1205.1(f), 1205.2(n.1) and 1205.5(h), a person elected or appointed as a district superintendent or assistant district superintendent for the first time in this Commonwealth under subsection (b) **or (b.1)** shall successfully complete a leadership development program that meets the Pennsylvania school leadership standards under section 1217.

Section 10. Section 1073 of the act, amended January 14, 1970 (1969 P.L.468, No.192) and January 16, 1974 (P.L.1, No.1), is amended to read:

Section 1073. Manner of Election or Approval.--(a) The board of school directors of each school district[, except in school districts of the first class,] shall meet at its regular place of meeting, during the last year of the term of the district superintendent or at any other time when a vacancy shall occur in the office of district superintendent, at an hour previously fixed by the board. The secretary of each board of school directors shall mail to each member thereof at least five days beforehand, a notice of the time, place and purpose of such meeting. At such meeting the board shall elect or approve a properly qualified district superintendent **to enter into a contract** to serve a term of [from] three to five years from the first day of July next following his election or from a time mutually agreed upon by the duly elected district superintendent and the board of school directors. **The contract shall be subject to the act of February 14, 2008 (P.L.6, No.3), known as the "Right-to-Know Law."**

(b) At a regular meeting of the board of school directors occurring at least one hundred fifty (150) days prior to the expiration date of the term of office of the district superintendent, the agenda shall include an item requiring affirmative action by five or more members of the board of school directors to notify the district superintendent that the board intends to retain him for a further term of [from] three (3) to five (5) years or that another or other candidates will be considered for the office. In the event that the board fails to take such action at a regular meeting of the board of school directors occurring at least one hundred fifty (150) days prior to the expiration date of the term of office of the district superintendent, he shall continue in office for a further term of similar length to that which he is serving.

(d) The term of office or commission of a district superintendent or assistant district superintendent shall not be shortened by reason of the fact that the district in which he serves shall [be come] **become** part of a joint school, or by reason of the fact that the district in which he serves shall become a part of a new school district established as the result of reorganization of school districts pursuant to Article II., subdivision (i) or section 224 of this act. Any district superintendent, assistant district superintendent or supervising principal not selected as the district superintendent of the joint school or newly established school district in which the district he serves becomes a part shall be assigned to a position or office for which he is eligible: Provided, however, That in a new school district reorganized under Article II., subdivision (i) or section 224 of this act, he shall be assigned to a position or office which is administrative or supervisory in nature only, but there shall be no reduction in salary until the expiration of his commission. Thereafter, unless elected to an office requiring a commission he shall have the status of a professional employe: Provided, That the board of school directors may adjust the salary according to the classification of the position to which he may be assigned, and that the period of service as a commissioned district superintendent, assistant district superintendent or associate superintendent shall be counted as time served as a professional employe in determining his seniority rights.

(e) The following shall apply:

(1) Notwithstanding any other provision of law, no individual shall be employed as a district superintendent or assistant district superintendent by a school district except pursuant to a written contract of employment expressly stating the terms and conditions of employment.

(2) A contract for the employment of a district superintendent or assistant district superintendent shall do all of the following:

(i) Contain the mutual and complete agreement between the district superintendent or assistant district superintendent and the board of school directors with respect to the terms and conditions of employment.

(ii) Consistent with State Board of Education certification requirements, specify the duties, responsibilities, job description and performance expectations, including performance standards and assessments provided for under section 1073.1.

(iii) Incorporate all provisions relating to compensation and benefits to be paid to or on behalf of the district superintendent or assistant district superintendent.

(iv) Specify the term of employment and state that the contract shall terminate immediately, except as otherwise provided under this section, upon the expiration of the term unless the contract is allowed to renew automatically under subsection (b).

(v) Specify the termination, buyout and severance provisions, including all postemployment compensation and the period of time in which the compensation shall be provided. Termination, buyout and severance provisions may not be modified during the course of the contract or in the event a contract is terminated prematurely.

(vi) Contain provisions relating to outside work that may be performed, if any.

(vii) State that any modification to the contract must be in writing.

(viii) State that the contract shall be governed by the laws of this Commonwealth.

(ix) Limit compensation for unused sick leave in new employment contracts entered into after the effective date of this subsection for district superintendents or assistant district superintendents who have no prior experience as a district superintendent or assistant district superintendent to the maximum compensation for unused sick leave under the school district's administrator compensation plan under section 1164 in effect at the time of the contract.

(x) Limit transferred sick leave from previous employment to not more than thirty (30) days in new employment contracts after the effective date of this subsection for district superintendents or assistant district superintendents who have no prior experience as a district superintendent or assistant district superintendent.

(xi) Specify postretirement benefits and the period of time in which the benefits shall be provided.

(3) No agreement between the board of school directors and a district superintendent or assistant district superintendent for a negotiated severance of employment prior to the end of the specified contract term shall provide for severance compensation to the district superintendent or assistant district superintendent, including the reasonable value of any noncash severance benefits or postemployment benefits not otherwise accruing under the contract or pursuant to law, that:

(i) If the agreement takes effect two (2) years or more prior to the end of the specified contract term, exceeds the equivalent of one (1) year's compensation and benefits otherwise due under the contract.

(ii) If the agreement takes effect less than two (2) years prior to the end of the specified contract term, exceeds the equivalent of one-half of the total compensation and benefits due under the contract for the remainder of the term.

Section 11. The act is amended by adding a section to read:

Section 1073.1. Performance Review.--(a) In addition to any other requirements provided for under this act, the employment contract for a district superintendent or assistant district superintendent shall include objective performance standards mutually agreed to in writing by the board of school directors and the district superintendent or assistant district superintendent. The objective performance standards may be based upon the following:

- (1) achievement of annual measurable objectives established by the school district;
- (2) achievement on Pennsylvania System of School Assessment (PSSA) tests;
- (3) achievement on Keystone Exams;
- (4) student growth as measured by the Pennsylvania Value-Added Assessment System;
- (5) attrition rates or graduation rates;
- (6) financial management standards;
- (7) standards of operational excellence; or
- (8) any additional criteria deemed relevant and mutually agreed to by the board of school directors and the district superintendent or assistant district superintendent.

(b) The board of school directors shall conduct a formal written performance assessment of the district superintendent and assistant district superintendent annually. A time frame for the assessment shall be included in the contract.

(b.1) The board of school directors shall post the mutually agreed to objective performance standards contained in the contract on the school district's publicly accessible Internet website. Upon completion of the annual performance assessment, the board of school directors shall post the date of the assessment and whether or not the district superintendent and assistant district superintendent have met the agreed-to objective performance standards on the school district's publicly accessible Internet website.

(c) The State Board of Education may promulgate regulations pursuant to the act of June 25, 1982 (P.L.633, No.181), known as the "Regulatory Review Act," in order to implement this section.

Section 12. Sections 1076 and 1077 of the act, amended January 16, 1974 (P.L.1, No.1), are amended to read:

Section 1076. Election of Assistant District Superintendents[, Except in Districts First Class].--[Except in districts of the first class, assistant] **Assistant** district superintendents shall be chosen by a majority vote of all the members of the board of school directors of the district, for a term of [from] three to five years upon the nomination by the district superintendent.

Section 1077. Term and Salary of Assistants.--(a) Assistant district superintendents may serve through the term of the district superintendent, or enter a contract for a term of [from] three to five years at salaries paid by the district, and fixed by a majority vote of the whole board of school directors prior to their election. **The contract shall be subject to the act of February 14, 2008 (P.L.6, No.3), known as the "Right-to-Know Law."**

(b) At a regular meeting of the board of school directors occurring at least one hundred fifty (150) days prior to the expiration date of the term of office of the assistant district superintendent, the agenda shall include an item requiring affirmative action by five (5) or more members of the board of school directors to notify the assistant district superintendent that the board intends to retain him for a further term of [from] three (3) to five (5) years or that another or other candidates will be considered for the office. In the event that the board fails to take such action at a regular meeting of the board of school directors occurring at least one hundred fifty (150) days prior to the expiration date of the term of office of the assistant district superintendent, he shall continue in office for a further term of similar length to that which he is serving.

Section 13. Section 1078 of the act, amended January 14, 1970 (1969 P.L.468, No.192), is amended to read:

Section 1078. Commissions.--District superintendents and assistant district superintendents shall be commissioned by the [Superintendent of Public Instruction] **Secretary of Education.**

Section 14. Section 1080 of the act is amended to read:

Section 1080. Removal.-- **(a)** District superintendents and assistant district superintendents may be removed from office **and have their contracts terminated**, after hearing, by a majority vote of the board of school directors of the district, for neglect of duty, incompetency, intemperance, or immorality, of which hearing notice of at least one week has been sent by mail to the accused, as well as to each member of the board of school directors.

(b) The board of school directors shall publicly disclose at the next regularly scheduled monthly meeting the removal of a district superintendent or assistant district superintendent from office under subsection (a).

(c) Proceedings under this section shall be held under 2 Pa.C.S. Ch. 5 Subch. B (relating to practice and procedure of local agencies).

Section 14.1. Section 1123 of the act, amended March 29, 1996 (P.L.47, No.16), is amended to read:

Section 1123. Rating System.-- **(a)** In determining whether a professional employe shall be dismissed for incompetency or unsatisfactory teaching performance as provided for in section 1122(a) of this act, and in rating [the services of a temporary professional employe, the professional employe or temporary professional employe shall be rated by an approved rating system which shall give due consideration to personality, preparation, technique, and pupil reaction, in accordance with standards and regulations for such scoring as defined by rating cards to be prepared by the Department of Education, and to be revised, from time to time, by the Department of Education with the cooperation and advice of a committee appointed by the Secretary of Education, including representation from district superintendents of schools, classroom teachers, school directors, school supervisors, parents of school-age children enrolled in a public school, a representative from a college or department of education within a higher education institution located within this Commonwealth, and such other groups or interests as the Secretary of Education may deem appropriate. Rating shall be done by or under the supervision of the superintendent of schools or, if so directed by him, the same may be done by an assistant superintendent, a supervisor, or a principal, who has supervision over the work of the professional employe or temporary professional employe who is being rated: Provided, That no unsatisfactory rating shall be valid unless approved by the district superintendent.] **professional employes and temporary professional employes, all professional employes and temporary professional employes shall be rated through the use of an approved rating tool developed by the Secretary of Education in consultation with education experts, parents of school-age children enrolled in a public school, teachers and administrators, including research and collaboration conducted by the department.**

(b) For professional employes and temporary professional employes who serve as classroom teachers, the following shall apply:

(1) Beginning in the 2013-2014 school year, the evaluation of the effectiveness of professional employes and temporary professional employes serving as classroom teachers shall give due consideration to the following:

(i) Classroom observation and practice models that are related to student achievement in each of the following areas:

- (A)** Planning and preparation.
- (B)** Classroom environment.
- (C)** Instruction.
- (D)** Professional responsibilities.

(ii) Student performance, which shall comprise fifty per centum (50%) of the overall rating of the professional employe or temporary professional employe serving as a classroom teacher and shall be based upon multiple measures of student achievement. The fifty per centum (50%) shall be comprised of the following:

(A) Fifteen per centum (15%) building-level data, including, but not limited to, all of the following:

(I) Student performance on assessments.

(II) Value-added assessment system data made available by the department under section 221.

(III) Graduation rate as reported to the department under section 222.

(IV) Promotion rate.

(V) Attendance rate as reported to the department under section 2512.

(VI) Advanced placement course participation.

(VII) Scholastic aptitude test and preliminary scholastic aptitude test data.

(B) Fifteen per centum (15%) teacher-specific data, including, but not limited to, student achievement attributable to a specific teacher as measured by all of the following:

(I) Student performance on assessments.

(II) Value-added assessment system data made available by the department under section 221.

(III) Progress in meeting the goals of student individualized education plans required under the Individuals With Disabilities Education Act (Public Law 91-230, 20 U.S.C. § 1400 et seq.).

(IV) Locally developed school district rubrics.

(C) Twenty per centum (20%) elective data, including measures of student achievement that are locally developed and selected by the school district from a list approved by the department and published in the Pennsylvania Bulletin by June 30 of each year, including, but not limited to, the following:

(I) District-designed measures and examinations.

(II) Nationally recognized standardized tests.

(III) Industry certification examinations.

(IV) Student projects pursuant to local requirements.

(V) Student portfolios pursuant to local requirements.

(2) (i) No later than June 30, 2013, the department shall develop, issue and publish in the Pennsylvania Bulletin a rating tool for professional employes and temporary professional employes serving as classroom teachers that is consistent with this subsection and includes the weights given to the multiple measures of student performance contained in clause (1) (ii).

(ii) Following publication, the rating tool developed under this subsection shall be used in the rating of all professional employes and temporary professional employes serving as classroom teachers.

(iii) After June 30, 2013, any changes to the rating tool developed under this subsection shall be made by the State Board of Education through regulations promulgated under the act of June 25, 1982 (P.L.633, No.181), known as the "Regulatory Review Act."

(c) For professional employes and temporary professional employes serving as principals, the following shall apply:

(1) Beginning in the 2014-2015 school year, principal effectiveness shall be measured using a rating tool designed specifically for professional employes and temporary professional employes serving as principals which shall give due consideration to the following:

(i) Planning and preparation.

(ii) School environment.

(iii) Delivery of service.

(iv) Professional development.

(v) Student performance, pursuant to clause (2).

(2) Student performance shall be measured as provided in subsection (b) (1) (ii) for professional employes and temporary professional employes

supervised by the principal and shall comprise fifty per centum (50%) of the principal's overall rating. The fifty per centum (50%) shall be comprised of the following:

(A) Fifteen per centum (15%) building-level data, including, but not limited to, all of the following:

(I) Student performance on assessments.

(II) Value-added assessment system data made available by the department under section 221.

(III) Graduation rate as reported to the department under section 222.

(IV) Promotion rate.

(V) Attendance rate as reported to the department under section 2512.

(VI) Advanced placement course participation.

(VII) Scholastic aptitude test and preliminary scholastic aptitude test data.

(B) Fifteen per centum (15%) correlation data based on teacher-level measures.

(C) Twenty per centum (20%) elective data, including measures of student achievement that are locally developed and selected by the school district from a list approved by the department and published in the Pennsylvania Bulletin by June 30 each year, which shall include, but not be limited to, the following:

(I) District-designed measures and examinations.

(II) Nationally recognized standardized tests.

(III) Industry certification examinations.

(IV) Student projects pursuant to local requirements.

(V) Student portfolios pursuant to local requirements.

(3) (i) No later than June 30, 2014, the department shall develop, issue and publish in the Pennsylvania Bulletin a rating tool for professional employes and temporary professional employes serving as principals that is consistent with this subsection and includes the weights given to the multiple measures of student performance contained in clause (2).

(ii) Following publication, the rating tool developed under this subsection shall be used in the rating of all principals superseding all other rating cards and forms used previously.

(iii) After June 30, 2014, any changes to the rating tool developed under this subsection shall be made by the State Board of Education through regulations promulgated under the "Regulatory Review Act."

(d) For nonteaching professional employes, the following shall apply:

(1) Beginning in the 2014-2015 school year, nonteaching professional employes shall be evaluated using a rating tool designed specifically for nonteaching professional employes which shall give due consideration to the following:

(i) Planning and preparation.

(ii) Educational environment.

(iii) Delivery of service.

(iv) Professional development.

(v) Student performance of all students in the school building in which the nonteaching professional employe is employed which shall comprise twenty per centum (20%) of the overall rating of nonteaching professional employes and temporary professional employes.

(2) (i) No later than June 30, 2014, the department shall develop, issue and publish in the Pennsylvania Bulletin a rating tool for nonteaching professional employes that is consistent with this subsection and includes the weights given to the multiple measures of student performance contained in clause

(1) (v).

(ii) Following publication, the rating tool developed under this subsection shall be used in the rating of all nonteaching professional employes.

(iii) After June 30, 2014, any subsequent changes to the rating tool developed under this paragraph shall be made by the State Board of Education through regulations promulgated under the "Regulatory Review Act."

(e) Notwithstanding subsections (b), (c) and (d), professional employes and temporary professional employes serving as classroom teachers, principals and nonteaching professional employes may be evaluated through the use of a rating tool developed by an individual school district, intermediate unit or area vocational-technical school that the department has approved as meeting or exceeding the measures of effectiveness established under this section.

(f) (1) Each rating tool developed or approved under this section shall identify the overall performance rating of the professional employes and temporary professional employes serving as classroom teachers, principals and nonteaching professional employes as one of the following:

- (i) Distinguished.
- (ii) Proficient.
- (iii) Needs improvement.
- (iv) Failing.

(2) An overall performance rating of either "distinguished" or "proficient" shall be considered satisfactory.

(3) An overall performance rating of "needs improvement" shall be considered satisfactory, except that any subsequent overall rating of "needs improvement" issued by the same employer within ten (10) years of the first overall performance rating of "needs improvement" where the employe is in the same certification shall be considered unsatisfactory.

(4) An overall performance rating of "failing" shall be considered unsatisfactory.

(5) An overall performance rating of "needs improvement" or "failing" shall require the employe to participate in a performance improvement plan. No employe shall be rated "needs improvement" or "failing" based solely upon student test scores.

(6) The department shall develop a rating scale to reflect student performance measures and employe observation results and establish overall score ranges for each of the four rating categories contained in clause (1).

(g) Upon publication in the Pennsylvania Bulletin of a rating tool developed under this section, the rating cards set forth in 22 Pa. Code § 351.21 (relating to rating form) and any alternative rating forms approved pursuant to 22 Pa. Code Ch. 351 (relating to teacher tenure hearings) prior to the implementation of this section shall be discontinued for use in the evaluation of professional and temporary professional employes.

(h) The following shall apply to the ratings of all professional employes and temporary professional employes:

(1) All ratings shall be completed using the rating tools developed or approved under this section.

(2) Professional employes shall be rated at least annually and temporary professional employes shall be rated at least twice annually.

(3) Ratings shall be performed by or under the supervision of the chief school administrator or, if so directed by the chief school administrator, by an assistant administrator, a supervisor or a principal who has supervision over the work of the professional employe or temporary professional employe being rated, provided that no unsatisfactory rating shall be valid unless approved by the chief school administrator.

(4) No employe shall be dismissed under section 1122 unless the employe has been provided a completed rating tool provided for under this section, which includes a description based upon classroom observations of

deficiencies in practice supported by detailed anecdotal records that justify the unsatisfactory rating.

(i) All school districts, intermediate units and area vocational-technical schools shall provide to the department the aggregate results of all professional employe and temporary professional employe, principal and nonteaching professional employe evaluations.

(j) (1) Any rating tool developed by the Department of Education under this section shall be exempt from:

(i) Sections 201, 202, 203, 204 and 205 of the act of July 31, 1968 (P.L.769, No.240), referred to as the "Commonwealth Documents Law."

(ii) Section 204(b) of the act of October 15, 1980 (P.L.950, No.164), known as the "Commonwealth Attorneys Act."

(iii) The "Regulatory Review Act."

(2) This subsection shall not apply to any changes made to a rating tool or new rating tool developed by the State Board of Education pursuant to subsections (b) (2) (iii), (c) (3) (iii) and (d) (2) (iii).

(k) The State Board of Education may develop standards or regulations consistent with this section.

(1) (1) The department's duty to develop a rating tool under subsection (b) (2) shall expire on June 30, 2013.

(2) The department's duty to develop rating tools under subsections (c) (3) and (d) (2) shall expire on June 30, 2014.

(m) No collective bargaining agreement negotiated by a school district and an exclusive representative of the employes in accordance with the act of July 23, 1970 (P.L.563, No.195), known as the "Public Employe Relations Act," after the effective date of this subsection shall provide for a rating system other than as provided for in this section. A provision in any agreement or contract in effect on the effective date of this subsection that provides for a rating system in conflict with this section shall be discontinued in any new or renewed agreement or contract or during the period of status quo following an expired contract.

(n) The requirements of this section shall apply to all school districts, intermediate units and area vocational-technical schools.

(o) For purposes of this section:

(1) The term "assessment" shall mean the Pennsylvania System of School Assessment test, the Keystone Exam, an equivalent local assessment or another test established by the State Board of Education to meet the requirements of section

2603-B(d) (10) (i) and required under the No Child Left Behind Act of 2001 (Public Law 107-110, 115 Stat. 1425) or its successor statute or required to achieve other standards established by the department for the school or school district under 22 Pa. Code § 403.3 (relating to single accountability system).

(2) The term "chief school administrator" shall include individuals who are employed as a school district superintendent, an executive director of an intermediate unit or a chief school administrator of an area vocational-technical school.

(3) The term "classroom teacher" shall mean a professional employe or temporary professional employe who provides direct instruction to students related to a specific subject or grade level.

(4) The term "department" shall mean the Department of Education of the Commonwealth.

(5) The term "education specialist" shall have the meaning given to it under the act of December 12, 1973 (P.L.397, No.141), known as the "Professional Educator Discipline Act."

(6) The term "nonteaching professional employe" shall mean an education specialist or a professional employe or temporary professional employe who provides services other than classroom instruction.

(7) The term "performance improvement plan" shall mean a plan, designed by a district with input of the employe, that may include mentoring, coaching, recommendations for professional development and intensive supervision based on the contents of the rating tool provided for under this section.

(8) The term "principal" shall include a building principal, an assistant principal, a vice principal or a director of vocational education.

(p) An employe's individual rating form shall not be subject to disclosure under the act of February 14, 2008 (P.L.6, No.3), known as the "Right-to-Know Law."

Section 15. Section 1124 of the act, amended August 8, 1963 (P.L.564, No.299), is amended to read:

Section 1124. Causes for Suspension.--(a) Any board of school directors may suspend the necessary number of professional employes, for any of the causes hereinafter enumerated:

(1) [Substantial] **substantial** decrease in pupil enrollment in the school district;

(2) [Curtailment] **curtailment** or alteration of the educational program on recommendation of the superintendent[, concurred in] **and on concurrence** by the board of school directors, [approved by the Department of Public Instruction,] as a result of substantial decline in class or course enrollments or to conform with standards of organization or educational activities required by law or recommended by the Department of Public Instruction;

(3) [Consolidation] **consolidation** of schools, whether within a single district, through a merger of districts, or as a result of joint board agreements, when such consolidation makes it unnecessary to retain the full staff of professional employes[.]; **or**

(4) [When] **when** new school districts are established as the result of reorganization of school districts pursuant to Article II., subdivision (i) of this act, and when such reorganization makes it unnecessary to retain the full staff of professional employes.

(b) **Notwithstanding an existing or future provision in a collective bargaining agreement or other similar employment contract to the contrary, suspension of a professional employe due to the curtailment or alteration of the educational program as set forth in subsection (a) (2) may be effectuated without the approval of the curtailment or alteration of the educational program by the Department of Education, provided that, where an educational program is altered or curtailed as set forth in subsection (a) (2), the school district shall notify the Department of Education of the actions taken pursuant to subsection (a) (2). The Department of Education shall post all notifications received from a school district pursuant to this subsection on the Department of Education's publicly accessible Internet website.**

Section 16. Section 1205.1(c) of the act, amended June 22, 2001 (P.L.530, No.35), is amended to read:

Section 1205.1. Continuing Professional Development.--* * *

(c) The professional education plan of each school entity shall be designed to meet the educational needs of that school entity and its professional employes. A school entity shall annually review its plan to determine whether or not it continues to reflect the needs of the school entity [and its strategic plan] and the needs of its professional employes, students and the community. The plan shall be amended as necessary to ensure that the plan meets the requirements of this subsection. The plan shall specify the continuing professional educational courses, programs, activities and other learning experiences approved to meet continuing professional

development requirements under section 1205.2(c), including efforts designed to improve teacher knowledge in subject areas covering the academic standards listed in 22 Pa. Code Ch. 4 (relating to academic standards and assessment).

* * *

Section 16.1. Section 1308 of the act, amended June 7, 1993 (P.L.49, No.16), is amended to read:

Section 1308. Liability for Tuition and Enforcement of Payment.--(a) In all cases not covered by the preceding section if a charge is made by any school district for tuition for the inmates of any such institution, the officers of the institution shall submit to the board of school directors a sworn statement, setting forth the names, ages, and school districts liable for tuition of all children who are inmates thereof, and desire to attend public school in the district. The district in which the institution is located shall obtain a blank acknowledging or disclaiming residence, signed by the secretary of the school district in which the institution declares the legal residence of the child to be. If said district shall fail to file said blank within fifteen (15) days from the date it is sent to the district by registered mail, the district in which the institution is located shall again notify the district of its failure to comply with the provisions of this act. If the district shall fail to comply within fifteen (15) days following the second notice, said failures to return the blank shall be construed as an acknowledgement of said child's residence. The tuition of such inmates as are included in the sworn statement to the board of school directors shall be paid by the district of residence of the inmates upon receipt of a bill from the district in which the institution is located setting forth the names, ages and tuition charges of the inmates. The district so charged with tuition may file an appeal with the Secretary of Education, in which it shall be the complainant and the district in which the institution is located the respondent. The decision of the Secretary of Education, as to which of said parties is responsible for tuition, shall be final.

(b) In the event that the district in which the institution is located contracts with a third party to provide educational services to children who are inmates of the institution, the third party may seek payment of tuition directly from the district of residence. The third party shall notify the district in which the institution is located of its payment request to the district of residence, and, if the district of residence makes payment to the third party, the third party shall notify the district in which the institution is located. Such payment to the third party shall satisfy and extinguish the contractual payment obligation of the district in which the institution is located. The district so charged with tuition by the third party may file an appeal with the secretary as set forth in subsection (a).

(c) If any inmates have been received from outside of Pennsylvania, or if the institution cannot certify as to their residence, their tuition shall be paid by the institution having the care or custody of said children, except in the case of medically indigent children hospitalized in exclusively charitable children's hospitals exempt under section 501(c) (3) of the Internal Revenue Code which make no charges to any of its patients nor accepts any third-party payments for services provided to any of its patients. In such cases their tuition shall be paid by the Commonwealth out of moneys appropriated by the General Assembly for the purposes of this act. Enrollment of any out-of-state student in a school district or intermediate unit program shall be conditioned upon a guarantee, or actual advance receipt, of tuition and transportation payment from the institution, from the student's home state or out-of-state school district, or from the out-of-state party or agency which placed the student in the institution, except in the case of medically indigent children hospitalized in exclusively charitable children's hospitals exempt under section 501(c) (3) of the Internal Revenue Code which make no charges to any of its patients nor accepts any third-party payments for services provided to any of its patients

where the Commonwealth is paying the tuition as otherwise provided for in this paragraph. If the Secretary of Education decides that the legal residence of any of said inmates is in Pennsylvania, but cannot be fixed in a particular district, the Commonwealth shall pay the tuition of such inmate out of moneys appropriated to the Department of Education by the General Assembly for the maintenance and support of the public schools of the Commonwealth.

Section 17. Section 1337.1 of the act, amended or added May 10, 2000 (P.L.44, No.16) and July 20, 2007 (P.L.278, No.45), is amended to read:

Section 1337.1. School Lunch and Breakfast Reimbursement.--(a) Schools that participate in the school lunch program shall be reimbursed in the following manner:

(1) Subject to future adjustments under clause (2), each school which offers the school lunch program shall receive a reimbursement of no less than ten cents (10¢) per lunch served, exclusive of any reimbursements under subsection (c).

(2) For the 2000-2001 school year and each school year thereafter, reimbursements for the school lunch program shall be fixed by regulation of the Department of Education: Provided, That such reimbursements shall be no less than the amounts per lunch served established by clause (1).

(b) Schools that participate in the school breakfast program shall be reimbursed in the following manner:

(1) Subject to future adjustments under clause (2), each school which offers the school breakfast program shall receive a reimbursement of no less than ten cents (10¢) per breakfast served.

(2) For the 2000-2001 school year and each school year thereafter, reimbursements for the school breakfast program shall be fixed by regulation of the Department of Education: Provided, That such reimbursements shall be no less than the amounts per breakfast served established by clause (1).

(c) Schools that participate in both the school lunch program and the school breakfast program shall be provided with the following incentive reimbursements:

(1) Subject to future adjustments under clause (3), each school which offers both a school lunch program under subsection (a) and a school breakfast program under subsection (b) which serves less than or equal to twenty per centum (20%) of its student enrollment shall receive an additional reimbursement of two cents (2¢) per lunch served.

(2) Subject to future adjustments under clause (3), each school which offers a school lunch program under subsection (a) and a school breakfast program under subsection (b) which serves more than twenty per centum (20%) of its student enrollment shall receive an additional reimbursement of four cents (4¢) per lunch served.

(3) For the 2000-2001 school year and each school year thereafter, reimbursements for the school breakfast incentive program shall be fixed by regulation of the Department of Education: Provided, That such reimbursement shall be no less than the amounts per lunch served established by clauses (1) and (2).

[(c.1) (1) In order to promote initiatives regarding child health and nutrition, the department shall establish a School Nutrition Incentive Program. The program shall provide a supplemental school lunch and breakfast reimbursement to any school in a local education agency that has adopted and implemented the nutritional guidelines for food and beverages available on each school campus published by the department pursuant to section 1422.3(5).

(2) To qualify, the local wellness policy adopted by the local education agency pursuant to section 1422.1 must indicate adoption of such guidelines.

(3) For the 2007-2008 school year and each school year thereafter, supplemental reimbursement shall be provided to schools in qualifying local education agencies as follows:

(i) Each school that offers the school lunch program under subsection (a) shall receive an additional reimbursement of one cent (1¢) per lunch served, exclusive of any additional supplemental reimbursement under subclause (iii) or (iv).

(ii) Each school that offers the school breakfast program under subsection (b) shall receive an additional reimbursement of one cent (1¢) per breakfast served, exclusive of any additional supplemental reimbursement under subclause (iii) or (iv).

(iii) Each school that offers both a school lunch program under subsection (a) and a school breakfast program under subsection (b) that serves breakfast to less than or equal to twenty per centum (20%) of its student enrollment shall receive an additional reimbursement of two cents (2¢) per lunch served.

(iv) Each school that offers both a school lunch program under subsection (a) and a school breakfast program under subsection (b) that serves breakfast to more than twenty per centum (20%) of its student enrollment shall receive an additional reimbursement of three cents (3¢) per lunch served.]

(d) For the purposes of this section, the following terms shall have the following meanings:

"School" shall have the same meaning as given to that term in 7 CFR 210.2 (relating to definitions).

"School lunch program" shall have the same meaning as given to the term "National School Lunch Program" in 7 CFR 210.2 (relating to definitions).

"School breakfast program" shall have the same meaning as given to that term in 7 CFR Pt. 220 (relating to School Breakfast Program).

Section 18. Section 1302.1-A(a) of the act, added November 17, 2010 (P.L.996, No.104), is amended to read:

Section 1302.1-A. Regulations.--(a) Within one year of the effective date of this section, the State Board of Education shall promulgate final-omitted regulations pursuant to the act of June 25, 1982 (P.L.633, No.181), known as the "Regulatory Review Act," necessary to implement this article. The regulations shall include the following:

(1) A model memorandum of understanding between school entities and local police departments. The model memorandum of understanding shall be reviewed on a biennial basis and revised where necessary. **The State Board of Education may revise the model memorandum of understanding by publishing a notice in the Pennsylvania Bulletin that contains the complete revised model memorandum of understanding. The revised model memorandum of understanding shall be incorporated into the Pennsylvania Code in place of the existing model memorandum of understanding.**

(2) Protocol for the notification of the police department when an offense listed under section 1303-A(b) (4.1) occurs on school property, which shall include a requirement that the local police department be notified immediately when such an offense occurs.

(3) Protocol for the notification of the police department at the discretion of the chief school administrator regarding an offense listed under section 1303-A(b) (4.2) or any other offense that occurs on school property.

(4) Protocol for emergency and nonemergency response by the police department, which shall include a requirement that the school district shall supply the police department with a copy of the comprehensive disaster response and emergency preparedness plan as required by 35 Pa.C.S. § 7701(g) (relating to duties concerning disaster prevention).

(5) Procedures and protocols for the response and handling of students with a disability, including procedures related to student behavior as required by 22 Pa. Code §§ 14.104 (relating to special education plans) and 14.133 (relating to positive behavior support).

* * *

Section 19. Section 1422.1 of the act, amended November 17, 2010 (P.L.996, No.104), is amended to read:

Section 1422.1. Local Wellness Policy.--(a) Not later than the first day of the school year beginning after June 30, 2006, each local education agency shall, pursuant to section 204 of the Child Nutrition and WIC Reauthorization Act of 2004 (Public Law 108-265, 118 Stat. 729), establish a local wellness policy for schools within the local education agency.

[(b) A local education agency to which 22 Pa. Code § 4.13 (relating to strategic plans) applies shall include the local wellness policy as part of the strategic plan required under 22 Pa. Code § 4.13.]

(c) A local education agency may submit its local wellness policy or information on other initiatives regarding child health, nutrition, food allergy reaction management and physical education to the Department of Education for inclusion in the clearinghouse established under section 1422.3(3). [An update to the policy information may be done in concert with the scheduled submission of the school district's strategic plan as required under 22 Pa. Code § 4.13.]

Section 19.1. The act is amended by adding a section to read:

Section 1424. Cardiopulmonary Resuscitation.--(a) A school entity shall have at each school, or in the case of a cyber charter school at each location, under its jurisdiction, except in extenuating circumstances, one person certified in the use of cardiopulmonary resuscitation during regular school hours when school is in session and students are present.

(b) The provisions of 42 Pa.C.S. §§ 8332 (relating to nonmedical good Samaritan civil immunity) and 8337.1 (relating to civil immunity of school officers or employees relating to emergency care, first aid and rescue) shall apply to a person who renders cardiopulmonary resuscitation.

(c) As used in this section, "school entity" means an area vocational-technical school, a charter school, a cyber charter school, an intermediate unit, a nonpublic school or a school district.

Section 20. Section 1524(a) of the act, added December 9, 2002 (P.L.1472, No.187), is amended to read:

Section 1524. Recognition of American Sign Language Courses.--(a) A student shall receive credit for completion of a course in American Sign Language at the high school level toward the satisfaction of the foreign language requirements included in [a school district's strategic plan or] requirements for graduation established pursuant to 22 Pa. Code Ch. 4 (relating to academic standards and assessment).

* * *

Section 21. The act is amended by adding an article to read:

**ARTICLE XV-G
OPEN CAMPUS INITIATIVES**

Section 1501-G. Legislative intent.

It is the intent of the General Assembly to encourage collaborative partnerships between school districts for the purpose of providing expanded access to high-quality curricula to students in a cost-effective manner through the use of technology.

Section 1502-G. Definitions.

The following words and phrases when used in this article shall have the meaning given to them in this section unless the context clearly indicates otherwise:

"Department." The Department of Education of the Commonwealth.

"Nonparticipating school entity." A school district, which is not a party to a cooperative agreement between school districts, a charter school, cyber charter school, nonpublic school or home education program provided under section 1327.1.

"Open campus initiative." A program established under section 1503-G.

"Participating school district." A school district which is party to the cooperative agreement between school districts establishing the open campus initiative.

Section 1503-G. Open campus initiatives.

(a) Establishment.--An open campus initiative may be established between school districts through a cooperative agreement.

(b) Courses.--Courses of an open campus initiative may be delivered outside of a school building in whole or in part using technology that may include the Internet, video conferencing or other electronic means.

(c) Grades, credit, promotion and graduation.--Course grades, credit, promotion and graduation policies for students participating in open campus initiative courses shall be determined by the school district, charter school, cyber charter school, nonpublic school or evaluator of a home education program in which the student is enrolled and the cooperative agreement between the participating school districts.

(d) Student eligibility requirements.--Eligibility requirements for student participation in open campus initiative courses shall be determined by the school district, charter school, cyber charter school, nonpublic school or supervisor of a home education program in which each student is enrolled and the cooperative agreement between the participating school districts.

(e) Nonparticipating school entities.--A cooperative agreement between participating school districts may provide for students attending a nonparticipating school entity to participate in an open campus initiative course.

(f) Compulsory attendance requirements.--The time during which a student participates in open campus initiative courses shall be considered to be compliant with the compulsory attendance requirements of section 1327.

(g) Student participation.--The school district, charter school, cyber charter school, nonpublic school or home education program in which the student is enrolled shall ensure that a student participating in open campus initiative courses is offered at least 990 hours of instruction at the secondary level and 900 hours of instruction at the elementary level.

(h) Technical assistance.--The department shall provide technical assistance as needed to school districts establishing and operating an open campus initiative.

Section 1504-G. Cooperative agreements.

(a) Contents.--School districts desiring to establish and operate an open campus initiative shall develop and enter into a cooperative agreement that shall, at a minimum, include:

(1) A policy for grading, credit and promotion of students participating in open campus initiative courses.

(2) A policy for participation of students from participating school districts which shall include minimum academic and attendance criteria.

(3) A policy for participation of students from nonparticipating school entities, which includes a fee schedule for determining tuition charges for courses delivered to those students, if the participating school districts allow participation of students from nonparticipating school entities.

(4) A policy for discipline and removal of students from open campus initiative courses in compliance with State law related to student discipline.

(5) Information about the attribution of student data to the school district, charter school, cyber charter school, nonpublic school or home education program in which the student is enrolled, including student assessment data.

(6) Information about the sharing of costs between the participating school districts.

- (7) Information about the use and distribution of tuition revenue.
- (8) Processes for adding and removing open campus initiative courses.
- (9) Processes for termination of the open campus initiative.

(b) Adoption by boards of school directors.--Each open campus initiative cooperative agreement shall be adopted by majority vote of the board of school directors of each of the participating school districts.

Section 1505-G. Reimbursements by the Commonwealth.

For the purpose of making reimbursements under Article XXV, a student participating in an open campus initiative shall be considered to be enrolled in the school district, charter school, cyber charter school, nonpublic school or home education program which determines the student's eligibility for participation in the open campus initiative.

Section 1506-G. Parental and public information.

All policies related to an open campus initiative shall be made accessible to parents and posted on the participating school districts' publicly accessible Internet websites.

Section 1507-G. Students with disabilities.

Nothing in this article or in any policy or cooperative agreement developed under this article shall conflict with:

- (1) Federal or State law regarding the protections provided to a student with a disability for receiving education in the least restrictive environment.
- (2) The legal authority of an individualized education program team to make appropriate program and placement decisions for a student with a disability in accordance with the student's individualized education program.

Section 1508-G. Collective bargaining.

Nothing contained in this article shall be construed to supersede or preempt the rights, remedies and procedures afforded to school employees or labor organizations under Federal or State law, including the act of July 23, 1970 (P.L.563, No.195), known as the Public Employe Relations Act, or any provision of a collective bargaining agreement negotiated between a school entity and an exclusive representative of the employees in accordance with that act.

Section 22. Section 1607(b) of the act, amended November 23, 2010 (P.L.1350, No.123), is amended and the section is amended by adding a subsection to read:

Section 1607. Attendance in Other Districts.--* * *

(b) If a third class school district operating under a special board of control pursuant to section 692 has, with the approval of the Secretary of Education, curtailed its educational program by eliminating its high school and has not assigned its high school pupils to another school district and provided adequate transportation in a manner under subsection (a), the secretary shall have the following authority:

- (1) To designate two or more school districts, which shall accept on a tuition basis the high school students of the distressed school district, so long as a designated school district's border is no more than three (3) miles from the border of the distressed school district. The designation under this paragraph shall occur no later than thirty (30) days after receiving the approval of the secretary to curtail its educational program by the elimination of its high school, provided, however, that if any school district meets the criteria of this subsection on the effective date of this subsection, the designation of school districts shall occur no later than thirty (30) days after the effective date of this subsection. [No designated school district shall be assigned more than one hundred sixty-five (165) students from the distressed school district.]

(2) To establish a process for the distressed school district to use to reassign its high school students to the school districts designated under paragraph (1).

(3) To establish the per-pupil tuition rate that a school district designated under paragraph (1) shall receive for each reassigned student in a regular or special education program. For the 2010-2011 **and 2011-2012** school [year and each school year thereafter] **years**, the tuition rate established under this paragraph may not exceed the product of:

- (i) the tuition rate established for the 2007-2008 school year; and
- (ii) the greater of:

(A) two percent (2%); or

(B) the percentage increase in total budgeted revenues available to a distressed school district.

(4) For the 2012-2013 school year and each school year thereafter, the per pupil tuition rate that a school district designated under paragraph (1) shall receive for each reassigned student in a regular or special education program shall be the greater of ten thousand dollars (\$10,000) or the product of:

- (i) the tuition rate established for the prior school year; and
- (ii) the greater of:

(A) the percentage increase in total budgeted revenues available to a distressed school district; or

(B) the index set pursuant to the act of June 27, 2006 (1st Sp.Sess., P.L.1873, No.1), known as the "Taxpayer Relief Act," for the distressed school district.

* * *

(h.1) For the two (2) consecutive school years following the effective date of this subsection, a school district designated under subsection (b) (1) shall receive an additional per-pupil sum of five hundred (\$500) dollars for students reassigned and entering grades seven, eight and nine pursuant to this section. These additional funds shall be used for transition services to students, including student mentoring, tutoring, employe in-service programs designed to assist transition students and security expenditures.

Section 22.1. The act is amended by adding an article to read:

**ARTICLE XVI-C
DISCLOSURE OF INTERSCHOLASTIC
ATHLETICS OPPORTUNITIES**

Section 1601-C. Scope of article.

This article requires reporting by school entities of athletic opportunities afforded to male and female secondary school students.

Section 1602-C. Definitions.

The following words and phrases when used in this article shall have the meanings given to them in this section unless the context clearly indicates otherwise:

"Athletic facilities." Locker rooms, playing fields, gymnasiums, field houses, arenas, athletic training rooms, stadiums, weight rooms or any other location used by secondary school students and their coaches for sports training, practice, competition and coaching.

"Department." The Department of Education of the Commonwealth.

"Equipment and supplies." Sport-specific equipment and supplies, general equipment and supplies, instructional devices and conditioning and weight training equipment.

"Participant." A secondary school student who is:

- (1) receiving institutionally sponsored support normally provided to athletes competing at the institution involved, such as coaching, equipment, medical and training room services, on a regular basis during a sport's season;

- (2) participating in organized practice sessions and other team meetings and activities on a regular basis during a sport's season; and
- (3) listed on the team list on the day of the team's first scheduled competition, excluding preseason scrimmages.

"School entity." A school of a school district, joint school district, area vocational-technical school or charter school that provides interscholastic athletic opportunities for secondary school students.

"Secondary school student." A student who attends a school entity in grades 7 through 12.

"Travel." Transportation, housing furnished during travel and per diem dining allowances.

"Uniforms." Clothing for practice and games, such as shoes, rain gear and warm-up suits.

Section 1603-C. Duty to disclose.

(a) Information.--A school entity shall annually submit information to the department regarding interscholastic athletic opportunity and treatment for male and female secondary school students for the preceding school year.

(b) Disclosure form.--The information shall be submitted on a disclosure form and in a manner to be established by the department.

(c) Submission.--By October 15, 2013, and October 15 of each year thereafter, a school entity shall submit to the department the completed disclosure form for the immediately preceding school year.

(d) Public access.--No later than November 1 of each year, a school entity shall make a copy of the completed disclosure form available for public inspection during regular business hours, including on any publicly accessible Internet website of the school entity. The completed disclosure form shall constitute a public record subject to public inspection under the act of February 14, 2008 (P.L.6, No.3), known as the Right-to-Know Law.

(e) Notice to students and other affected individuals.--As soon as the disclosure form required by this section is completed, each school entity shall provide notice of its availability for review to students, educational personnel, student athletes and parents by posting a notice on school bulletin boards, in the school newspaper, on any electronic mailing list or list serve and by any other means reasonably likely to provide such notice.

Section 1604-C. Department duties.

(a) Duties.--The department shall establish a disclosure form for the submission of the required information for the immediately preceding school year by school entities. The department shall provide for the distribution of the disclosure form through the department's Internet website and shall provide technical assistance to school entities.

(b) Disclosure form information.--The following information shall be collected for all secondary school students in grades 7 through 12 and shall be included in the disclosure form:

- (1) The total number of students in each school entity as of October 1 of the immediately preceding school year, including:

- (i) the total number of students by gender; and
- (ii) the total number of male students by race or ethnicity and the total number of female students by race or ethnicity.

- (2) A listing by gender of each varsity, junior varsity and freshman athletic team that competed in interscholastic athletic competition.

- (3) For each team identified in paragraph (2), the following information:

- (i) The total number of team participants as of the day of the first scheduled competition for each team by gender.

- (ii) The total number of male team participants by race or ethnicity and the total number of female team participants by race or

ethnicity as of the day of the first scheduled competition for each team.

(iii) For the initial submission under this article, the school year in which each existing interscholastic athletic team was established and, for teams that the school entity sponsored in the past but no longer sponsors, the identity of the team and the year it was eliminated or demoted from interscholastic competition. For each subsequent year, a listing of interscholastic athletic teams that were newly established, reestablished, eliminated or demoted from interscholastic competition during the reported school year.

(iv) The seasons during which each interscholastic athletic team competed.

(v) The total value of contributions and purchases made on behalf of each team by booster clubs, alumni and any other nonschool sources.

(vi) The total expenditures for each team in the school year, including a separate listing of expenditures for each team in each of the following categories:

(A) The total amount of expenditures for travel.

(B) The total amount of expenditures for purchase and replacement of athletic uniforms.

(C) The total amount of expenditures for purchase and replacement of equipment and supplies.

(D) Compensation of coaches, per sport and per season.

(E) Expenditures made for construction, renovation, expansion, maintenance, repair and rental of athletic facilities. For any facilities shared by multiple teams, expenditures per team shall be calculated either by dividing expenditures by the number of teams using the facility or percentage of time used by each team.

(F) Compensation of athletic trainers per academic year.

(vii) The total number of athletic trainers, including the amount of time spent by each athletic trainer with each team.

(viii) The total number of coaches per team by employment status, full time, part time, head and assistant.

(ix) The total number of competitions scheduled and played per team.

(x) The name of the school entity's Title IX compliance officer required under 34 CFR § 106.8(a) (relating to designation of responsible employee and adoption of grievance procedures).

(c) Copies.--The department shall make copies of all submitted disclosures available for public inspection on the department's publicly accessible Internet website.

(d) Annual report.--No later than January 15 of each year, the department shall prepare and submit an annual report to the General Assembly regarding the compliance with the disclosure requirements of this article and summarizing the information submitted to it regarding interscholastic athletic opportunity for and treatment of each gender by race and ethnic group and other such information as the department deems relevant.

Section 1605-C. Regulations.

The department may promulgate rules, regulations or standards to administer this article.

Section 22.2. Section 1704-B heading and (c) of the act, amended or added July 11, 2006 (P.L.1092, No.114) and July 9, 2008 (P.L.846, No.61), are reenacted and amended to read:

Section 1704-B. Board of School Directors of Commonwealth Partnership School Districts.--* * *

(c) (1) In addition to all current rights, powers, privileges, prerogatives and duties, a board of school directors of a school district that has been placed on the empowerment list **on or after June 30, 2006**, due to the designation by the secretary as a Commonwealth partnership school district shall have the power to cancel or renegotiate any contract, other than collective bargaining agreements, for the purpose of making necessary economies in the operation of the schools within the school district; eliminate nonprofessional positions for services nonessential for the operation of the school district; or enter into agreements with individuals, for-profit or nonprofit organizations for the operation of school buildings or groups of school buildings or for the provision of educational or other types of services to or for the school district.

(2) The superintendent shall be responsible for the implementation of a system of performance review of school administrators, as approved by the board of school directors. Administrator performance shall be evaluated on the basis of abilities and effectiveness to manage the operation of the school facilities and staff, manage resources, provide instructional leadership, implement and administer the school budget and promote and maintain a positive educational learning environment.

(3) Based upon an unsatisfactory review and evaluation of a school administrator arising from the implementation of the program established in paragraph (2), a board of school directors may reassign, transfer or suspend the school administrator without regard to section 1125.1 or 1151.

(4) Based upon an unsatisfactory review and evaluation of a school administrator arising from the implementation of the program established in paragraph (2), a board of school directors may dismiss the school administrator pursuant to the procedure contained in section 1127, provided that the board of school directors shall afford the school administrator notice and an opportunity to be heard pursuant to 2 Pa.C.S. Ch. 5 Subch. B (relating to practice and procedure of local agencies).

(4.1) In addition to powers enumerated in this act, a school district designated as a Commonwealth partnership school district may dispose of unused and unnecessary lands and buildings, if such buildings are in excess of twenty-five (25) years of age, in the following manner, notwithstanding the provisions of section 707 of this act:

(i) By negotiated sale, provided the district has an affidavit of at least three (3) persons who are familiar with the value of real estate in the locality in which the lands and buildings proposed to be sold are located, who have examined the property and set forth a value for the property and who opine that the consideration for the property is equal to or better than that which could be received by sealed bid. The sale price shall not be less than the highest value set forth in the three (3) affidavits.

(ii) By entering into agreements with an urban redevelopment authority organized under the act of May 24, 1945 (P.L.991, No.385), known as the Urban Redevelopment Law, under which the district may convey property to the authority for the purpose of the authority facilitating the conveyance of the property consistent with the goals of the school district and the authority.

(5) [As used in] **For purposes of** this subsection, ["school administrator" shall have the same meaning given to it under section 1164] **the following terms shall have the following meanings:**

"Commonwealth partnership school district." A school district for which the secretary has determined, on or after July 11, 2006, and not later than September 9, 2006, all of the following:

(i) **The school district has experienced a decline of fifteen per centum (15%) or more in student enrollment during the immediately preceding five-year period.**

(ii) The school district has experienced a loss of revenue during the immediately preceding three-year period due to the statutory removal of one or more of the sources of revenue made available pursuant to section 652.1.

(iii) The school district has an equalized millage for the 2004-2005 fiscal year of greater than twenty-seven (27).

"Empowerment list." A list prepared by the Department of Education containing school districts that fall below certain academic assessments as provided in former section 1703-B.

"School administrator." As defined in section 1164.

"Secretary." The Secretary of Education of the Commonwealth.

[(6) This subsection shall expire June 30, 2012.]

Section 23. Section 1913-A(b)(1.6)(v) of the act, amended June 30, 2011 (P.L.112, No.24), is amended and the clause is amended by adding a subclause to read:

Section 1913-A. Financial Program; Reimbursement of Payments.--* * *

(b) * * *

(1.6) For the 2006-2007 fiscal year and each fiscal year thereafter, the payment for a community college shall consist of the following:

* * *

(v) Subclauses (i), (ii), (iii) and (iv) shall not apply to the 2011-2012 and 2012-2013 fiscal [year] years.

* * *

(vii) For the 2012-2013 fiscal year, each community college shall receive an amount equal to the sum of the following:

(A) A reimbursement for operating costs determined by:

(I) dividing the amount of funding that the community college received in fiscal year 2011-2012 under subclause (vi)(A) by the total amount of funding provided to community colleges in fiscal year 2011-2012 under subclause (vi)(A); and

(II) multiplying the quotient under subparagraph (I) by \$168,167,000.

(B) An economic development stipend determined by:

(I) dividing the amount of funding that the community college received in fiscal year 2011-2012 under subclause (vi)(B) by the total amount of funding provided to all community colleges in fiscal year 2011-2012 under subclause (vi)(B); and

(II) multiplying the quotient under subparagraph (I) by \$44,000,000.

* * *

Section 24. Sections 1913-B.1(c) and 2003-A.1(c) and (c.1) of the act, amended or added November 3, 2011 (P.L.400, No.97), are amended and the sections are amended by adding subsections to read:

Section 1913-B.1. Contracts for Construction, Repair, Renovation or Maintenance.--* * *

[(c) All contracts, other than contracts for the retention of architects and engineers, authorized by this section which exceed a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 119, shall be advertised in the manner provided by law and competitively bid and awarded to the lowest responsible bidder. In case of emergencies and notwithstanding any other provision of this section to the contrary, the board of trustees may make or authorize others to make an emergency procurement whenever a threat exists to public health, welfare or safety or circumstances outside the control of the college and creates an urgency of need which does not permit the delay involved in using more formal competitive methods. Whenever practical, in the case of a procurement of a supply, at least two (2) bids shall be solicited. A written determination of the basis for the emergency and for the selection of the particular contractor shall be included in the contract file.]

(c.1) All contracts, other than contracts for the retention of architects and engineers, authorized by this section which exceed a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 120, shall be advertised in the manner provided by law and competitively bid and awarded to the lowest responsible bidder. In case of emergencies and notwithstanding any other provision of this section to the contrary, the board of trustees may make or authorize others to make an emergency procurement whenever a threat exists to public health, welfare or safety or circumstances outside the control of the college and creates an urgency of need which does not permit the delay involved in using more formal competitive methods. Whenever practical, in the case of a procurement of a supply, at least two (2) bids shall be solicited. A written determination of the basis for the emergency and for the selection of the particular contractor shall be included in the contract file.

* * *

Section 2003-A.1. Project Contracts.--* * *

[(c) All contracts, other than contracts for the retention of architects and engineers, authorized by this section which exceed a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under subsection (c.1), shall be advertised in the manner provided by law and competitively bid and awarded to the lowest responsible bidder. In case of emergencies and notwithstanding any other provision of this section to the contrary, the chancellor may make or authorize others to make an emergency procurement whenever a threat exists to public health, welfare or safety or circumstances outside the control of the State system and creates an urgency of need which does not permit the delay involved in using more formal competitive methods. Whenever practical, in the case of a procurement of a supply, at least two (2) bids shall be solicited. A written determination of the basis for the emergency and for the selection of the particular contractor shall be included in the contract file.

(c.1) Adjustments shall be made as follows:

(1) The Department of Labor and Industry shall calculate the average annual percentage change in the Consumer Price Index for All Urban Consumers: All Items (CPI-U) for the United States City Average as published by the United States Department of Labor, Bureau of Labor Statistics, for the twelve-month period ending September 30, 2012, and for each successive twelve-month period thereafter.

(1.1) If the department determines that there is no positive percentage change, then no adjustment to the base amounts shall occur for the relevant time period.

(2) The positive percentage change, as determined in accordance with clause (1), shall be multiplied by the amount applicable under subsection (c) for the current period, and the product thereof shall be added to the amount applicable under subsection (c) for the current period, with the result rounded to the nearest multiple of one hundred dollars (\$100).

(3) The determination required under clause (1) and the calculation adjustments required under clause (2) shall be made in the period between October 1 and November 15 of the year following the effective date of this subsection and between October 1 and November 15 of each successive year.

(4) The adjusted amounts obtained in accordance with clause (2) shall become effective January 1 for the period following the year in which the determination required under clause (1) is made.

(5) The Department of Labor and Industry shall give notice in the Pennsylvania Bulletin prior to January 1 of each calendar year in which the percentage change is determined in accordance with clause (1) of the amounts, whether adjusted or unadjusted in accordance with clause (2), at which competitive bidding is required under subsection (c) for the period beginning the first day of January after publication of the notice.

(6) The annual increase in the preliminary adjusted base amounts obtained under clauses (3) and (4) shall not exceed three percent (3%).]

(c.2) All contracts, other than contracts for the retention of architects and engineers, authorized by this section which exceed a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under subsection (c.3), shall be advertised in the manner provided by law and competitively bid and awarded to the lowest responsible bidder. In case of emergencies and notwithstanding any other provision of this section to the contrary, the chancellor may make or authorize others to make an emergency procurement whenever a threat exists to public health, welfare or safety or circumstances outside the control of the State system and creates an urgency of need which does not permit the delay involved in using more formal competitive methods. Whenever practical, in the case of a procurement of a supply, at least two (2) bids shall be solicited. A written determination of the basis for the emergency and for the selection of the particular contractor shall be included in the contract file.

(c.3) Adjustments shall be made as follows:

(1) The Department of Labor and Industry shall calculate the average annual percentage change in the Consumer Price Index for All Urban Consumers: All Items (CPI-U) for the United States City Average as published by the United States Department of Labor, Bureau of Labor Statistics, for the twelve-month period ending September 30, 2012, and for each successive twelve-month period thereafter.

(1.1) If the Department of Labor and Industry determines that there is no positive percentage change, then no adjustment to the base amounts shall occur for the relevant time period.

(2) The positive percentage change, as determined in accordance with clause (1), shall be multiplied by the amount applicable under subsection (c.2) for the current period, and the product thereof shall be added to the amount applicable under subsection (c.2) for the current period, with the result rounded to the nearest multiple of one hundred dollars (\$100).

(3) The determination required under clause (1) and the calculation adjustments required under clause (2) shall be made in the period between October 1 and November 15, 2012, and between October 1 and November 15 of each successive year.

(4) The adjusted amounts obtained in accordance with clause (2) shall become effective January 1 for the period following the year in which the determination required under clause (1) is made.

(5) The Department of Labor and Industry shall give notice in the Pennsylvania Bulletin prior to January 1 of each calendar year in which the percentage change is determined in accordance with clause (1) of the amounts, whether adjusted or unadjusted in accordance with clause (2), at which competitive bidding is required under subsection (c.2) for the period beginning the first day of January after publication of the notice.

(6) The annual increase in the preliminary adjusted base amounts obtained under clauses (3) and (4) shall not exceed three percent (3%).

* * *

Section 25. Section 2010-A(10) of the act, amended November 3, 2011 (P.L.400, No.97), is amended and the section is amended by adding a paragraph to read:

Section 2010-A. Power and Duties of Institution Presidents.--The president of each institution shall be appointed by the board. The president shall be the chief executive officer of that institution. He shall have the right to attend all meetings of the council of that institution and shall have the right to speak on all matters before the council but not to vote. Subject to the stated authority of the board and the council, each president shall have the following powers and duties:

* * *

[(10) Within the limitations of the operating budget and other available funds in accordance with the procedures established by the board and with the approval of the local council, to negotiate and award all contracts for equipment, services and supplies in excess of a cost of a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 119, on a competitive bid basis and to purchase instructional, educational, extracurricular, technical, administrative, custodial and maintenance equipment and supplies not in excess of a cost of a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 119, without competitive bidding, except that such items shall not be bought in series to avoid the dollar ceiling.]

(10.1) Within the limitations of the operating budget and other available funds in accordance with the procedures established by the board and with the approval of the local council, to negotiate and award all contracts for equipment, services and supplies in excess of a cost of a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 120, on a competitive bid basis and to purchase instructional, educational, extracurricular, technical, administrative, custodial and maintenance equipment and supplies not in excess of a cost of a base amount of eighteen thousand five hundred dollars (\$18,500), subject to adjustment under section 120, without competitive bidding, except that such items shall not be bought in series to avoid the dollar ceiling.

* * *

Section 26. Section 2104 of the act is amended to read:

Section 2104. Superintendents of Schools or Buildings and of Supplies.-- The board of public education in each school district of the first class shall, whenever a vacancy in said office shall occur, appoint a district superintendent, who shall be designated and known as superintendent of schools[, for a term of not more than six (6) years]. The board may also appoint a superintendent of buildings and a superintendent of supplies. The board shall prescribe the terms and duties and fix the salaries of each of such employees. They shall be responsible to the board for the conduct of their respective departments, shall make annual reports to the board, and shall from time to time submit such plans and suggestions for the improvement of the schools and the school system as they shall deem expedient or as the board of public education may require.

The superintendent of buildings shall be an engineer or architect of good standing in his profession. The superintendent of buildings and the superintendent of supplies shall each give such security for the faithful performance of the duties of their respective offices as the board of public education shall prescribe.

Section 27. The act is amended by adding sections to read:

Section 2319. State aid for fiscal year 2012-2013.

Notwithstanding any other provision of law to the contrary, each library subject to the act of June 14, 1961 (P.L.324, No.188), known as The Library Code, shall be eligible for State aid for fiscal year 2012-2013, which shall consist of the following:

(1) Funds appropriated for libraries shall be distributed to each library under the following formula:

(i) Divide the amount of funding that the library received in fiscal year 2011-2012 under section 2318 by the total State-aid subsidy for fiscal year 2011-2012.

(ii) Multiply the quotient under subparagraph (i) by the total State-aid subsidy for 2012-2013.

(2) Following distribution of funds appropriated for State aid to libraries under paragraph (1), any remaining funds may be distributed at the discretion of the State Librarian.

(3) If funds appropriated for State aid to libraries in fiscal year 2012-2013 are less than funds appropriated in fiscal year 2002-2003, the State Librarian may waive standards as prescribed in section 103 of The Library Code.

(4) (i) Each library system receiving State aid under this section may distribute the local library share of that aid in a manner as determined by the board of directors of the library system.

(ii) In the case of a library system that contains a library operating in a city of the second class, changes to the distribution of State aid to the library shall be made by mutual agreement between the library and the library system.

(5) In the event of a change in district library center population prior to the effective date of this section as a result of:

(i) a city, borough, town, township, school district or county moving from one library center to another; or

(ii) a transfer of district library center status to a county library system,

funding of district library center aid shall be paid based on the population of the newly established or reconfigured district library center.

Section 2502.51. Basic Education Funding for 2011-2012 School Year.--(a) For the 2011-2012 school year, the Commonwealth shall pay to each school district a basic education funding allocation which shall consist of the following:

(1) An amount equal to the basic education funding allocation for the 2010-2011 school year pursuant to section 2502.50.

(2) An English language learner high incidence supplement calculated for qualifying school districts as follows:

(i) To qualify for the English language learner high incidence supplement, a school district's 2011-2012 market value/income aid ratio must be greater than seven thousand ten-thousandths (0.7000) and its English language learner concentration must be equal to or greater than six percent (6%).

(ii) The English language learner high incidence supplement shall be calculated for qualifying school districts as follows:

(A) Multiply the qualifying school district's 2009-2010 average daily membership by seventeen million four hundred fifty thousand dollars (\$17,450,000).

(B) Divide the product from clause (A) by the sum of the 2009-2010 average daily membership for all qualifying school districts.

(3) A charter and cyber charter school extraordinary enrollment supplement calculated for qualifying school districts as follows:

(i) To qualify for the charter and cyber charter school extraordinary enrollment supplement, a school district's 2011-2012 market value/income aid ratio must be greater than eight thousand ten-thousandths (0.8000), its 2009-2010 average daily membership must be greater than five thousand (5,000) and its charter and cyber charter school concentration must be greater than fourteen percent (14%).

(ii) The charter and cyber charter school extraordinary enrollment supplement shall be calculated for qualifying school districts as follows:

(A) For school districts with a charter and cyber charter school concentration equal to or greater than thirty percent (30%), multiply the qualifying school district's 2009-2010 average daily membership by one thousand three hundred seventeen dollars (\$1,317).

(B) For school districts with a charter and cyber charter school concentration less than thirty percent (30%), multiply the qualifying school district's 2009-2010 average daily membership by five hundred dollars (\$500).

(4) A second class county school district supplement calculated for qualifying school districts as follows:

(i) To qualify for the second class county school district supplement, a school district must be located in a county of the second class, its 2011-2012 market value/income aid ratio must be greater than six thousand ten-thousandths (0.6000), its 2009-2010 average daily membership must be greater than one thousand (1,000) and less than two thousand five hundred (2,500) and its number of students eligible for free or reduced-price meals under the National School Lunch Program during the 2010-2011 school year must be greater than one thousand fifty (1,050) and less than one thousand two hundred fifty (1,250).

(ii) The second class county school district supplement shall be calculated for qualifying school districts as follows:

(A) Multiply the qualifying school district's 2009-2010 average daily membership by one million dollars (\$1,000,000).

(B) Divide the product from clause (A) by the sum of the 2009-2010 average daily membership for all qualifying school districts.

(5) A second class school district supplement calculated for qualifying school districts as follows:

(i) To qualify for the second class school district supplement, a school district must have been classified as a second class school district during the 2000 census, it must have received State reimbursements pursuant to section 2591.1 for the 2009-2010 school year in an amount greater than three million five hundred thousand dollars (\$3,500,000), and it must have a 2009-2010 average daily membership greater than eight thousand (8,000).

(ii) The second class school district supplement shall be calculated for qualifying school districts as follows:

(A) Multiply the qualifying school district's 2009-2010 average daily membership by one million dollars (\$1,000,000).

(B) Divide the product from clause (A) by the sum of the 2009-2010 average daily membership for all qualifying school districts.

(6) An increasing aid ratio supplement calculated for qualifying school districts as follows:

(i) To qualify for the increasing aid ratio supplement, a school district's 2011-2012 market value/income aid ratio must be greater than six thousand ten-thousandths (0.6000), its 2009-2010 average daily membership must be greater than eleven thousand (11,000), its number of students eligible for free or reduced-price meals under the National School Lunch Program during the 2010-2011 school year must be greater than six thousand (6,000) and its market value/income aid ratio must have increased from the 1991-1992 school year to the 2011-2012 school year by more than sixty percent (60%).

(ii) The increasing aid ratio supplement shall be calculated for qualifying school districts as follows:

(A) Multiply the qualifying school district's 2009-2010 average daily membership by two million dollars (\$2,000,000).

(B) Divide the product from clause (A) by the sum of the 2009-2010 average daily membership for all qualifying school districts.

(7) A personal income supplement calculated for qualifying school districts as follows:

(i) To qualify for the personal income supplement, a school district's 2011-2012 market value/income aid ratio must be greater than fifty-two hundred ten-thousandths (0.5200) and less than six thousand ten-thousandths (0.6000), its 2009-2010 average daily membership must be greater than five thousand five hundred (5,500), its 2010-2011 equalized millage rate must be greater than twenty-two (22) and less than twenty-six (26), its adjusted personal income valuation for the 2008 tax year must be greater than six hundred fifty million dollars (\$650,000,000) and its number of students

eligible for free or reduced-price meals under the National School Lunch Program during the 2010-2011 school year must be greater than one thousand five hundred (1,500).

(ii) The personal income supplement shall be calculated for qualifying school districts as follows:

(A) For each school district with an adjusted personal income valuation for the 2008 tax year greater than eight hundred million dollars (\$800,000,000), the personal income supplement shall be two million dollars (\$2,000,000).

(B) For each school district with an adjusted personal income valuation for the 2008 tax year less than eight hundred million dollars (\$800,000,000), the personal income supplement shall be one million five hundred thousand dollars (\$1,500,000).

(8) A small district increasing aid ratio supplement calculated for qualifying school districts as follows:

(i) To qualify for the small district increasing aid ratio supplement, a school district's 2011-2012 market value/income aid ratio must be greater than five thousand ten-thousandths (0.5000) and less than fifty-five hundred ten-thousandths (0.5500), its market value/income aid ratio must have increased from the 1991-1992 school year to the 2011-2012 school year by more than thirty percent (30%), and its 2009-2010 average daily membership must be greater than one thousand seven hundred (1,700) and less than one thousand eight hundred (1,800).

(ii) The small district increasing aid ratio supplement shall be calculated for qualifying school districts as follows:

(A) Multiply the qualifying school district's 2009-2010 average daily membership by three hundred thousand dollars (\$300,000).

(B) Divide the product from clause (A) by the sum of the 2009-2010 average daily membership for all qualifying school districts.

(9) A small district supplement calculated for qualifying school districts as follows:

(i) To qualify for the small district supplement, a school district's 2011-2012 market value/income aid ratio must be greater than seven thousand ten-thousandths (0.7000) and less than seventy-five ten-thousandths (0.7500), its 2009-2010 average daily membership must be greater than one thousand two hundred (1,200) and less than one thousand three hundred (1,300), and its 2010-2011 equalized millage rate must be greater than nineteen (19).

(ii) The small district supplement shall be calculated for qualifying school districts as follows:

(A) Multiply the qualifying school district's 2009-2010 average daily membership by two hundred fifty thousand dollars (\$250,000).

(B) Divide the product from clause (A) by the sum of the 2009-2010 average daily membership for all qualifying school districts.

(b) The data used to calculate the provisions contained in subsection (a) (2), (3), (4), (5), (6), (7), (8) and (9) shall be based on information available to the Department of Education as of June 30, 2012.

(c) Funds received under subsection (a) (2) (ii) by a school district with a 2011-2012 market value/income aid ratio greater than eighty-five hundred ten-thousandths (0.8500) shall be withheld until such time that a spending plan proposed by the school district's board of school directors for use of the funds is approved by the Secretary of Education in consultation with the local intermediate unit.

(d) The following shall apply:

(1) (i) Funds received by a school district under subsection (a) (3) (ii) (A) shall be used to satisfy judgments and past-due accounts payable beyond ninety (90) or more days, including health care benefits, payments to charter schools, payments to approved private schools and payments to intermediate units.

(ii) If all judgments have been satisfied and past-due accounts paid, funds may be used for timely payment of health care benefits, payments to charter schools, payments to approved private schools, payments to intermediate units and for other expenses approved by the Secretary of Education to ensure the fiscal stability of the school district.

(2) Not later than August 31, 2012, the school district shall submit a report to the Department of Education detailing the use of the funds received under this subsection, including specific payment amounts, specific payment dates and the entities receiving payment.

(e) Any undistributed funds shall be deposited in the Financial Recovery School District Transitional Loan Account.

(f) For the purposes of this section:

(1) The English language learner concentration shall be determined by dividing the school district's number of enrolled students identified as limited English proficient during the 2009-2010 school year by its 2009-2010 average daily membership.

(2) The charter and cyber school concentration shall be determined by dividing the school district's 2009-2010 average daily membership enrolled in charter and cyber charter schools by its 2009-2010 average daily membership.

Section 28. Sections 2509.1(c.1) and 2509.5(aaa) of the act, added June 30, 2011 (P.L.112, No.24), are amended to read:

Section 2509.1. Payments to Intermediate Units.--* * *

(c.1) For the 2011-2012 [school year] and 2012-2013 school years, five and one-half percent (5.5%) of the State special education appropriation shall be paid to intermediate units on account of special education services. Of this five and one-half percent (5.5%), thirty-five percent (35%) shall be distributed equally among all intermediate units. The remaining sixty-five percent (65%) shall be distributed to each intermediate unit in proportion to the number of average daily membership of the component school districts of each intermediate unit as compared to the Statewide total average daily membership.

* * *

Section 2509.5. Special Education Payments to School Districts.--* * *

(aaa) During the 2009-2010 through the [2011-2012] 2012-2013 school years, each school district shall be paid the amount it received during the 2008-2009 school year under subsection (zz). If insufficient funds are appropriated, the payments shall be made on a pro rata basis.

Section 29. The act is amended by adding a section to read:

Section 2510.2. Assistance to School Districts Certified as Education Empowerment Districts.--For the 2012-2013 fiscal year, the Department of Education may utilize up to \$4,500,000 of undistributed funds not expended, encumbered or committed from appropriations for grants and subsidies made to the Department of Education to assist school districts certified on or before June 30, 2010, as an education empowerment district under section 1705-B(h) (3). The funds shall be transferred by the Secretary of the Budget to a restricted account as necessary to make payments under this section and, when transferred, are hereby appropriated to carry out the provisions of this section.

Section 30. Section 2574(a) of the act, amended September 29, 1959 (P.L.992, No.407), is amended to read:

Section 2574. Approved Reimbursable Rental for Leases Hereafter Approved and Approved Reimbursable Sinking Fund Charges on Indebtedness.--(a) For school building projects for which the general construction contract is awarded subsequent to March 22, 1956, and for approved school building projects for which the general construction contract was awarded but for which a lease was not approved by the Department of [Public Instruction] **Education** prior to March 22, 1956, the Department of [Public Instruction] **Education** shall calculate an approved reimbursable rental or approved

reimbursable sinking fund charges. Reimbursable sinking fund charges may include charges for temporary indebtedness within constitutional limitations, if the indebtedness is incurred for approved permanent improvements to the school plant including the cost of acquiring a suitable site for a school building, the cost of constructing a new school building, or the cost of providing needed additions or alterations to existing buildings for which no bond issue is provided and for which an approved obligation or obligations other than bonds have been issued and the obligation or obligations are payable within five (5) years from the date of issue of the obligation in equal annual installments.

Nothing in this section or in the Department of Education guidelines shall prohibit a school district from receiving reimbursement for approved building improvements, including the cost of acquiring a suitable site for a school building, the cost of constructing a new school building or the cost of providing needed additions or alterations to existing buildings, if a school district elects not to remove any relocatable or modular classroom utilized after the completion of a building project. The term "relocatable or modular classroom" shall mean a classroom not of a permanent nature which meets the criteria and specifications of the Department of Education.

Approved reimbursable rental or sinking fund charge shall consist of that part of the annual rental or sinking fund charge attributable to--

(1) The cost of acquiring the land upon which the school buildings are situate, the cost of necessary rough grading to permit proper placement of the building upon said land and the cost of sewage treatment plants, as required by the Department of Health, to the extent that such costs are deemed reasonable by the Department of [Public Instruction] **Education** and the interest on such costs of acquisition, grading and sewage treatment plants earned subsequent to date the construction contract is awarded, and

(2) The approved building construction cost and the interest on such construction cost.

* * *

Section 31. Section 2599.2(e)(7) of the act, added July 9, 2008 (P.L.846, No.61), is amended to read:

Section 2599.2. Pennsylvania Accountability Grants.--* * *

(e) * * *

(7) For the 2008-2009 **and 2012-2013** fiscal [year] **years**, if insufficient funds are appropriated to make Commonwealth payments pursuant to this section, such payments shall be made on a pro rata basis.

* * *

Section 32. The act is amended by adding a section to read:

Section 2599.4. Targeted Industry Cluster Certificate Scholarship Program.--(a) The Targeted Industry Cluster Certificate Scholarship Program is established within the Pennsylvania Higher Education Assistance Agency.

(b) The agency may use funds appropriated to provide grants for defraying the necessary expense of residents of this Commonwealth who are eligible students pursuing an eligible course of study at an eligible educational provider.

(c) The agency shall determine and approve student eligibility and educational provider eligibility requirements for the program.

(d) The Department of Education shall consult with the Department of Labor and Industry to identify programs of study that train individuals for areas of immediate workforce need and provide the agency with a list of eligible programs of study.

(e) Grant awards shall be established by the agency based upon available resources.

Section 33. Section 2506-A(b) of the act, added December 23, 2003 (P.L.304, No.48), is amended to read:

Section 2506-A. Review process.

* * *

(b) [Strategic plan.--Each district's review shall be scheduled one year prior to the date required for filing of the district's strategic plan under 22 Pa. Code § 4.13 (relating to strategic plans) or its successor regulations, except that all districts placed on the education empowerment list pursuant to section 1703-B or determined to be distressed pursuant to section 691 as of the effective date of this article shall be included in the initial cycle of school districts subject to a review.] **(Reserved)**.

* * *

Section 34. This act shall apply as follows:

(1) The amendment or addition of sections 1073, 1073.1, 1076, 1077, 1078 and 1080 of the act shall apply to contracts of district superintendents or assistant district superintendents entered into or renewed on or after November 1, 2012.

(2) To contracts and purchases advertised on or after July 1, 2012, or immediately, whichever is later.

Section 35. This act shall take effect as follows:

(1) The amendment of section 1308 of the act shall take effect in 60 days.

(2) The amendments of section 1439 of the act shall take effect in 90 days.

(3) The addition of Article XV-G of the act shall take effect immediately.

(4) Except as otherwise provided in paragraph (5), the addition of Article XVI-C of the act shall take effect in 60 days.

(5) The addition of section 1604-C(b)(3)(v) of the act shall take effect two years from the effective date of this section.

(6) The reenactment and amendment of section 1704-B heading and (c) of the act shall take effect immediately.

(7) Section 34 and this section shall take effect immediately.

(8) The remainder of this act shall take effect July 1, 2012, or immediately, whichever is later.

APPROVED--The 30th day of June, A.D. 2012.

TOM CORBETT

**ESEA Flexibility Proposal Consultation
(General)**

Date	Group	Location	Educators	Parents	Com- munity	Other	TOTAL
February 4, 2013	All Penn*Link Subscribers (thousands)	via eMail	unknown # of recipients; 9 respondents				
February 13, 2013	School Leaders	PDE - Harrisburg	9			4	13
February 14, 2013	PA Association of Federal Programs - Executive	Harrisburg	23			2	25
February 21, 2013	Legislative Leaders	Capitol - Harrisburg				9	9
February 22, 2013	Committee of Practitioners	via Conference Call	9	1		3	13
February 28, 2013	All visitors to PDE website	via Internet	unknown # of viewers; will track respondents				

**Pennsylvania Common Core Standards
State Board Roundtables for Public Discussion**

Date	Group	Location	Educators	Parents	Com- munity	Other	TOTAL
August 17, 2010	State Board Common Core Roundtable	Erie				26	26
August 31, 2010	State Board Common Core Roundtable	Easton				24	24
September 15, 2010	State Board Common Core Roundtable	Pittsburgh				23	23
September 27, 2010	State Board Common Core Roundtable	Harrisburg				32	32
		TOTALS				105	105

**School Performance Profile
Focus Group Summary**

Date	Group	Location	Educators	Parents	Community	Other	TOTAL
December 7, 2011	SAS Institute	Hershey	350				350
April 11, 2012	PA State System of Higher School Improvement Grant Conference	Harrisburg					*
June 28, 2012	Pennsylvania Association of Career and Technical Administrators	Camp Hill	102			7	109
July 26, 2012	State Parent Advisory Council	State College	113			4	117
July 31, 2012	Career and Technical Centers	Seven Springs	28	150		4	182
September 18, 2012	Intermediate Unit Executives	Harrisburg	16				16
October 5, 2012	Focus Group	Harrisburg	25				25
October 22, 2012	Focus Group	Pittsburgh (IU 3)	45	4	6	9	64
October 24, 2012	Focus Group	Doylestown (IU 22)	25	7		2	34
October 25, 2012	Focus Group	Harrisburg	48	8	7	1	64
October 26, 2012	Focus Group	Harrisburg	11				11
November 6, 2012	Focus Group	Reading (IU 14)	83	2	1	1	87
November 15, 2012	Intermediate Units Curriculum Coordinators	State College	62				62
November 16, 2012	Focus Group	West Decatur (IU 10)	23			1	24
November 19, 2012	Focus Group	Kingston (IU 18)	39				39
November 20, 2012	Focus Group	Morton (IU 25)	39			4	43
December 5, 2012	SAS Institute	Hershey	859			10	869
February 13, 2013	House Education Committee	Harrisburg				25	25
		TOTALS	1868	171	14	68	2121

* Attendance information is not available for this event

**Pennsylvania Educator Effectiveness Initiative
Steering Committee, Focus Groups, Consultation**

Date	Group	Location	Educators	Parents	Community	Other	TOTAL
September 16-17, 2010	Teacher Evaluation Steering Committee	Pittsburgh	20			3	23
January 11, 2011	Teacher Evaluation Steering Committee	Harrisburg	20			3	23
June 30, 2011	Teacher Evaluation Steering Committee	Harrisburg	20			3	23
September 20, 2011	Principal Evaluation Focus Group	Greensburg	28				28
January 17, 2012	Principal Evaluation Focus Group	Duncansville	25				25
January 26, 2012	Teacher Evaluation Steering Committee	Harrisburg	20			3	23
February 22, 2012	Act 82 Stakeholder Group	Harrisburg	18	3		2	23
April 5, 2012	Principal Evaluation Focus Group	Lancaster	12				12
April 10, 2012	Principal Evaluation Focus Group	Schnecksville	25				25
April 25, 2012	Principal Evaluation Focus Group	Morton	30				30
May 25, 2012	Teacher Evaluation Briefing	Philadelphia					*
June 18, 2012	Principal Evaluation Focus Group	Harrisburg	12				12
July 24, 2012	PSEA Leadership Conference	Gettysburg					*
August 7, 2012	Principal Evaluation Focus Group	Montandon	30				30
September 25, 2012	Principal Evaluation Focus Group	Archbald	28				28
September 28, 2012	PASA Board of Governors Meeting	Harrisburg					*
October 4, 2012	SEA Panel on Teacher Evaluation at Mid-Atlantic Education Research Conference	Philadelphia					*
October 17, 2012	School Leadership Conference	Hershey					*
October 18, 2012	Act 82 Stakeholder Group	Harrisburg	18	3		2	23
October 24, 2012	Learning Forward PA October Institute	Harrisburg					*

**Pennsylvania Educator Effectiveness Initiative
Steering Committee, Focus Groups, Consultation**

November 7, 2012	CLIU #21 Superintendents Advisory Council Leadership Conference	Hershey					*
November 19, 2012	PASCD	Hershey					*
November 30, 2012	PSEA House of Delegates	Philadelphia					*
December 13, 2012	Principal Evaluation Focus Group	Grove City	25				25
January 17, 2013	PASA New Superintendents' Academy	Harrisburg					*
	TOTALS:		311	6		13	330

* Attendance information is not available for this event

Pennsylvania Teacher and Principal Evaluation Pilot Final Report

Suzanne Lane
Christy Horner

University of Pittsburgh

September 2011

We extend our sincere appreciation to the teachers, principals, supervisors, and superintendents who participated in the pilot. Their contributions are critical in the improvement of the evaluation system.

Preparation of this document was supported by a contract (#707089) from the Team Pennsylvania Foundation in collaboration with the Pennsylvania Department of Education. Any opinions expressed herein are those of the authors and do not necessarily represent the views of the Team PA Foundation.

Executive Summary

During the 2010-2011 academic year, Pennsylvania undertook a pilot study for the design, implementation and evaluation of a teacher and principal evaluation system (Commonwealth of Pennsylvania, 2010). It was undertaken in response to the U.S. Department of Education Race to the Top initiative (U.S. Department of Education, 2010). Under the Race to the Top initiative, states are required to 1) measure individual student growth, 2) design and implement transparent evaluation systems for teachers and principals that can differentiate effectiveness and take into account student growth, and are designed with principal and teacher involvement, 3) conduct annual evaluations of teachers and principals, and 4) use the results of the evaluations to inform decisions.

Pilot Study Process

A Stakeholder Committee, consisting of approximately 40 participants representing school districts, teacher unions, intermediate units, and industry and policy groups, met periodically to guide and support the design and implementation of the pilot teacher and principal evaluation study. Two work groups were formed, Teacher Evaluation Committee and the Principal Evaluation Committee, to design the evaluation system based on the recommendations of the Stakeholder Committee.

The four pilot sites that participated in the pilot study were Allentown School District, Cornell School District, Mohawk Area School District, and Northwest Tri-County IU5. The sample included principals/supervisors and teachers in elementary and secondary schools. Teachers who taught mathematics, language arts (English, reading), and science were included in the sample.

The training of the pilot participants occurred in January and February of 2011 and was conducted by an external consultant group. Separate training sessions were held for the teacher evaluation process and the principal/supervisor evaluation process. Teachers and principals attended the training on the teacher evaluation process, and principals and principal evaluators attended the training on the principal evaluation process. Each training session took approximately 3 hours and several training sessions took place to accommodate the pilot sites. After the completion of the trainings, the pilot sites began the implementation of the teacher and principal evaluation system.

Evaluation of the Pilot Study

To help improve the pilot evaluation system, several methods were used to gather information about the teacher and principal evaluation pilot from participants including surveys, focus groups with teachers, and interviews with principals and principal evaluators. Across the four sites, 114 teachers responded to the online survey (71% response rate), 19 principals responded to the online principal survey (63% response rate), and 5 principal evaluators responded to their survey (83% response rate). Overall, 34 teachers participated in the focus groups, 6 principals were interviewed, and 3 principal evaluators were interviewed across three of the pilot sites.

In interpreting the results of the evaluation of the pilot study it is important to consider the time frame of the study. Participants were trained in January and February and the implementation of the study immediately followed. This resulted in less than 5 months to implement the teacher and principal evaluation system.

The results of the surveys, focus groups and interviews indicated that the training sessions could be improved to better prepare participants for the evaluation system. Participants suggested revising the training format for the teacher evaluation system so that it is tailored to teachers' grade levels and content areas. They also indicated the need to extend the training so that it is more comprehensive and focuses on

all the domains covered in the rubric and on the types of evidence that teachers can provide to support their effectiveness. Teachers and principals responded favorably to many aspects of the teacher evaluation rubric. In particular, they valued the domains and criteria reflected in the rubric – Danielson’s Framework- and its comprehensiveness, capturing the many aspects that are reflective of effective teaching. Participants provided constructive feedback on the rubric wording, including the need for greater specificity, clarity and usability of the rubrics, as well as the need to provide more information on the types of evidence required. Overall, participants responded favorably to the observations and conferences in the teacher evaluation process. Some participants suggested that more formal observations by different observers would be preferable, while others suggested the need for more walk-throughs. Teachers indicated that the conversations with their principals and the time for self-reflection were valuable components of the evaluation system. Both principals and teachers indicated that the teacher evaluation system supports good teaching.

Principal and principal evaluators indicated that the training for the principal evaluation system needs to be revised so as to better prepare them. Overall, they suggested that the training needs to be more comprehensive and focus on all aspects of the evaluation system. The principal and principal evaluators indicated that the core and corollary competencies reflected in the rubric are valuable, but some indicated that the scoring system should be revised so as not to reduce a score to the lowest competency level in which all indicators were met. Others indicated the need to reduce the redundancy within the rubric and to shorten the rubric. Participants indicated that additional information on the sources of evidence and the allocation of time to meet with colleagues to discuss evidence of practice would enhance the evaluation system. Principals valued the time to conference with their supervisor and suggested the need for additional conferences, at least a midyear conference and evaluation. Both principals and principal evaluators indicated that the principal evaluation system supports good leadership.

Recommendations

The pilot for the Pennsylvania Teacher and Principal Evaluation System succeeded in providing important information for modifying the system. The results from the surveys, focus groups and interviews provide valuable information for the revision of the Pennsylvania Teacher and Principal Evaluation System. The specific recommendations are categorized into three areas: Training, Rubric and Evaluation Process.

Recommendations for the Teacher Evaluation System

Training

- More extensive training with follow-ups during the year
- Smaller training groups formed by grade span and content area
- Align video clips of teacher practice to grade span and content area
- Spend more time on the distinction between the rubric levels (especially between Proficient and Distinguished)
- More information on the nature of evidence for each Domain Component
- More specific information about the evaluation process and steps required for the evaluation
- Train the supervisors prior to the teachers so they can be a resource

Teacher Evaluation Rubric

- Reduce the verbiage in the criteria and redundancy across the criteria
- Clarify the distinction between Proficient and Distinguished
- Create a user-friendly online version
- Provide clear examples of the type of evidence for the Domain Components at different levels

Principle 3 – Appendix C

- Ensure consistency with Principal Evaluation Rubric
- Establish the reliability of the rubric scores
- Establish the validity of the score inferences, in particular, the validity of the domain score inferences

Evaluation Process – Observations/walkthroughs/conferencing

- Start the evaluation process at the beginning of the academic year with a goal setting conference between teacher and supervisor
- Provide documentation on the evaluation process and steps required
- Maintain multiple formal observations with conferences
- Specify more than one walk-through and more condensed rubrics for their use (and how the results from the walk-throughs would be combined with the results from the formal observations)
- Use multiple observers (may allow for more observations)
- Create time for teachers to meet and discuss the rubric, criteria, evidence, etc. throughout the year

Recommendations for the Principal/Supervisor Evaluation System

Training

- More extensive training with follow-up
- More information on the nature of evidence for each Competency
- More specific information about the evaluation process and steps required for the evaluation

Principal Evaluation Rubric

- Reduce the verbiage in the Competencies, and redundancy across the Competencies and Summary of Evidence
- Change the rating process so that a principal who receives a higher rating for all but one descriptor does not receive the lower rating
- Create a friendly online version
- Provide clear examples of the type of evidence for the Competencies at different levels
- Ensure consistency with Teacher Evaluation Rubric
- Establish the reliability of the rubric scores
- Establish the validity of the score inferences, in particular the validity of the domain score inferences

Evaluation Process

- Start the evaluation process at the beginning of the year with goal setting activities
- Consider a mid-year conference between the principal and evaluator
- Provide documentation on the evaluation process and steps required
- Create time for principals to meet and discuss the rubric, criteria, evidence, etc.

Pennsylvania Teacher and Principal Evaluation Pilot

During the 2010-2011 academic year, Pennsylvania undertook a pilot study for the design, implementation and evaluation of a teacher and principal evaluation system (Commonwealth of Pennsylvania, 2010). It was undertaken in response to the U.S. Department of Education Race to the Top initiative (U.S. Department of Education, 2010). Under the Race to the Top initiative, states are required to 1) measure individual student growth, 2) design and implement transparent evaluation systems for teachers and principals that can differentiate effectiveness and take into account student growth, and are designed with principal and teacher involvement, 3) conduct annual evaluations of teachers and principals, and 4) use the results of the evaluations to inform decisions.

It is widely recognized that effective teachers have a great influence over the improvement of student achievement and learning (Darling-Hammond, 2000), and effective principals have a critical role in promoting system capacity for successful learning and teaching (Elmore, 2002). The Race to the Top initiative aims to not only provide this country with a new generation of student assessments that focus on high level thinking skills, but also with quality standards-based evaluation systems to evaluate both teacher and principal effectiveness in improving student achievement and learning. Research has demonstrated that some standards-based performance evaluation systems show a positive relationship with student achievement, suggesting that instructional practices measured by these evaluation systems may have an impact on student learning (e.g., Heneman, et al., 2006; Konold et al., 2008; Newmann, et al., 2001). As would be expected, the relationship between evaluation results and student achievement is limited given the number of factors that influence both.

A previous paper (Lane & Horner, 2010) submitted to Team PA and PDE provided a review of principal and teacher evaluation systems and a discussion on design and implementation considerations so as to ensure the validity and reliability of the systems and their assessments. It was organized into nine sections: Standards for an Educational Evaluation Process, Validation of Assessment Methods, Characteristics of Effective Teacher Evaluation Systems, Assessment Methods Used in Teacher Evaluation Systems, District and State Teacher Evaluations Systems, Characteristics of Effective Principal Evaluation Systems, Assessment Methods Used in Principal Evaluation Systems, District and State Principal Evaluation Systems, and Conclusion.

The current paper provides an overview of the Pennsylvania principal and teacher evaluation pilot and an evaluation of the pilot. It is organized into six sections: Summary of the Standards for an Educational Evaluation Process and the Validation of Assessment Methods, Stakeholder Meetings, Pilot Sites, Training, Evaluation of the Principal and Teacher Evaluation Pilot, and Recommendations.

I. Summary of the Standards for an Educational Evaluation Process and the Validation of Assessment Methods

Standards for an Educational Evaluation Process

Howard and Gullickson (2010) emphasize that sound evaluations of teachers and principals are those that meet the *Joint Committee Standards for Educational Evaluation* (JCSEE; 1988). Personnel evaluation is defined by the JCSEE (2009) as “the systematic assessment of a person’s performance and/or qualification in relation to a professional role and some specified and defensible institution practice” (p. 27). The JCSEE developed the widely recognized *Personnel Evaluation Standards* (<http://www.jcsee.org/personnel-evaluation-standards>) that address four accepted attributes of educational evaluation: Propriety, Utility, Feasibility, and Accuracy. The seven Propriety Standards concern the

legality and ethical conduct of personnel evaluations. There are six Utility Standards, designed to help ensure that evaluations are “informative, timely, and influential”. The three Feasibility Standards are designed to help ensure that evaluations are “as easy to implement as possible, efficient in their use of time and resources, adequately funded, and viable from a political standpoint”. The eleven Accuracy Standards address the completeness and soundness of the information collected during an evaluation process. While not all of these standards need apply to a particular evaluation system, it is imperative that those who are responsible for an evaluation system identify the standards that are appropriate for the intended purpose(s) of the system.

Validation of Assessment Methods

Validity is the most fundamental technical consideration in the evaluation of educational assessment systems. Any assessment method used to evaluate an examinee’s performance in a domain which is subsequently scored using a standardized procedure can be considered a test, including observation protocols, portfolios, and other standardized inventories. As stated in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), validity is the “degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests” (p. 9). Test score inferences therefore need to be verified by both empirical evidence and a logical argument. The *Standards* (AERA, APA & NCME, 1999) has proposed a meaningful way to organize the collection and examination of evidence for a validity evaluation. This includes evidence based on assessment content, evidence based on response processes, evidence based on internal structure, evidence based on relations to other variables, and evidence based on consequences of testing.

Evidence based on assessment content is obtained thorough a review of the literature on the domain to be assessed and an examination by content experts of the relationship between the content of the assessment and the target domain as well as its intended use and score interpretations. Evidence based on response processes may include an examination of the processes of those who evaluate examinee performances or products. For example, it is important to ensure that evaluators are applying the scoring criteria appropriately and consistently, and are not being influenced by irrelevant factors during the evaluation. Evidence based on internal structure requires analyses of the extent to which relationships among assessment parts are consistent with the targeted domain. As an example, if a teacher assessment provides multiple scores that are used for making evaluative judgments, evidence is needed to ensure that the scores are not highly related, and instead each assess some unique aspect of the domain.

The relationships between assessment scores and variables external to the test can also serve as important validity evidence. For example, if a teacher assessment is intended to predict student achievement scores, evidence supporting this claim is needed. Evidence based on consequences of testing requires the examination of both intended positive consequences and potential negative consequences that occur when using an assessment and making decisions about individuals based on its results. This requires the delineation of the values inherent in teacher and principal assessments and an examination of the consequences of the decisions and actions based on the assessment results.

II. Stakeholder Meetings and Committee Work

The Stakeholder Committee that was formed for the Pennsylvania Teacher and Principal Evaluation Pilot consists of approximately 40 participants representing school districts, teacher unions, intermediate units, and industry and policy groups.

The first Stakeholder meeting for the Pennsylvania Teacher and Principal Evaluation Pilot occurred on September 16 and 17, 2010. The purpose of the meeting was the development of a framework for a research-based evaluation model for teachers and principals in the Commonwealth. Representatives from

the University of Pittsburgh, Mathematica Policy Research, Inc., and the National Institute for School Leadership presented on current evaluation systems being used throughout the country and methods to measure student growth (value-added measures). After research findings were presented to the group, two smaller work group discussions (a Principal Evaluation committee and a Teacher Evaluation committee) centered on characteristics of effective teachers and principals. Representatives from the workgroups shared the committees' ideas with the larger group. These ideas and conclusions were used to form the basis of the framework for the teacher and principal evaluation model.

The second Stakeholder meeting took place in October, 2010, after each committee met separately to move forward with the development of the systems. At this meeting, the Principal Evaluation Committee and Teacher Evaluation committee reported on their progress. The Principal Evaluation Committee reported that during their second meeting in early October: 1) NISL provided a presentation on the alignment of the NISL curriculum to the PIL standards, 2) Discovery Communications provided a presentation on the VAL-ED process, and 3) the committee reviewed examples of state and district principal evaluation systems. The committee provided a set of recommendations for the Principal Evaluation Tool and Process and indicated that a draft of the evaluation tool would be completed by the first of November. The Teacher Evaluation Committee reported that during their second meeting in early October, 1) a representative from the Danielson group provided a presentation on the best practices of teacher evaluation and 2) the committee discussed the rubrics, evaluation process and required training. The committee provided a set of recommendations for the Teacher Evaluation Rubrics and indicated that a draft of the evaluation tool would be completed by the first of November. After the presentations by the committees, the Stakeholder committee provided comments and recommendations for the teacher and principal evaluation system. The efforts in the design of the evaluation systems by these committees provide content validity evidence for the system.

During the January 2011 Stakeholder committee meeting, the group reviewed and discussed the pilot training for teachers and principals and the evaluation model. The Stakeholder committee approved the evaluation tools and rubrics as well as the professional development modules. The Danielson Framework was used in the design of the Teacher Evaluation Rubric and the PIL Standards were used as the framework for the Principal Evaluation Tool.

Additional Stakeholder committee meetings to review the project results and to provide input in the revision of the evaluation system have been planned.

III. Pilot Sites

The pilot sites assisted PDE in the development and implementation of the evaluation systems for teachers and principals. The pilot sites were asked to:

1. Identify a valid sampling of participating teachers and principals to pilot the protocols/processes to be utilized throughout the 2010-2011 school year.
2. Collaborate with the stakeholder steering committee and leadership team to monitor, adjust, and provide reflections throughout the pilot year.
3. Assist in the development and utilization of the professional development modules to support the evaluation system and statewide rollout.
4. Participate in interviews and reporting throughout the process for interim progress reports and final report, "Lessons Learned."

Initially, the pilot sites were Allentown School District, Bellwood-Antis, Quaker Valley School District, Cornell School District, Mohawk Area School District, and Northwest Tri-County IU5. Due to unforeseen circumstances two sites could not participate.

The four pilot sites that participated were Allentown School District, Cornell School District, Mohawk Area School District, and Northwest Tri-County IU5. The evaluation process included elementary and secondary schools, and teachers who taught mathematics, language arts (reading, English), and science. More specific information about the sample is provided in Section IV.

IV. Training

The training of the pilot participants took place in January and early February of 2011. Separate training sessions were held for the teacher evaluation process and the principal evaluation process. It should be noted that the principal evaluation process also included other supervisors of teachers, but in this document they will be referred to as principals. Teachers and principals attended the training on the teacher evaluation process. Principals and principal evaluators attended the training on the principal evaluation process. Each training session was approximately 3 hours and several training sessions took place to accommodate the pilot sites. After the trainings were complete, the pilot sites began the implementation of the teacher and principal evaluation system.

Observers from Team PA, PDE and the University of Pittsburgh attended some of the training sessions and completed a short evaluation form. The survey asked the observers to evaluate components of the training sessions given the training session agenda and the 3 hour timeframe. Three observers of the principal training sessions completed the survey and five observers of the teacher training sessions completed the survey. In addition, all participants in the pilot were asked to provide their feedback on the training sessions via an online survey which is presented in the next section.

The results for the teacher training sessions are displayed in Table 1. Overall, the observers tended to somewhat agree or agree with the majority of the items for the teacher training sessions.

Table 1. Observer Responses to Teacher Training Sessions (frequencies)

Survey item	Strongly Agree	Agree	Somewhat Agree	Somewhat Disagree	Disagree/ Strongly Disagree
Time spent on each topic was appropriate	1	2	2		
Printed material supported the training		4	1		
Printed material was clear		5			
Evaluation tool/ procedures presented clearly	2	3			
Participants actively engaged	2	3			
Participants questions addressed		3	2		
Training achieved its goals	1	2	2		

The observers provided comments for each item, which can be found in Appendix A. These comments offer valuable information for the design of future training sessions for the evaluation system. As an example, below are comments regarding whether the training achieved its stated goals:

Principle 3 – Appendix C

Except for how the schools are to implement the evaluation system

Goals were stated several times during the training, but without much consistency or clarity....However, the last time the goals were stated, they were goals that had been achieved.

Focus is too much on “what” and not enough on “how” ... I think the principals are going to need a great deal of support as they begin to implement this

Time was our enemy and allowed for only cursory skill training and no assessment of learning

We would have profited from time for the training staff and the Leadership staff to meet and review all of the elements of the training. This was lacking again due to the amount of time available for the pilot.

The results for the principal training sessions are in Table 2. Overall, the observers tended to agree with the majority of the items for the principal training sessions.

Table 2. Observer Responses to Principal Training Sessions (frequencies)

Survey Item	Strongly Agree	Agree	Somewhat Agree	Somewhat Disagree	Disagree/ Strongly Disagree
Time spent on each topic was appropriate	1	1	1		
Printed material supported the training	1	2			
Printed material was clear	1	2			
Evaluation tool/procedures presented clearly	1	1	1		
Participants actively engaged	2	1			
Participants questions addressed		1	1	1	
Training achieved its goals		3			

The comments the observers provided on each item for the principal training sessions are presented in Appendix A. Below are comments on whether the training for the principal evaluation system achieved its stated goals:

With the exception of the vagueness surrounding the procedures, the stated goals were achieved

...some follow-up coaching may be necessary as they get into the process

Time was our enemy and allowed for only cursory skill training and no assessment of learning

We would have profited from time for the training staff and the Leadership staff to meet and review all of the elements of the training. This was lacking again due to the amount of time available for the pilot.

V. Evaluation of the Implementation of the Teacher and Principal Evaluation Pilot in the Schools

To help improve the pilot evaluation system, several methods were used to gather information about the teacher and principal evaluation pilot from the participants including online surveys, focus groups and interviews.

Surveys

Separate online surveys were designed for teachers, principals/supervisors, and principal/supervisor evaluators. The surveys asked teachers, principals, and principal evaluators to respond to questions in seven areas:

- Training Effectiveness
- Rubric/Criteria Quality
- Evaluation Process Quality (e.g., observations, conferencing, self-input)
- Feedback Quality/Amount
- Support Teaching/Leadership Goals
- Operational Ease/Worth Effort
- Process Should Remain the Same (Improvement)

Because most principals assumed dual roles in the pilot, the principal evaluation survey had two sections; one section pertained to their role as an evaluator in the teacher evaluation process and the second section pertained to their role as an evaluatee in the principal evaluation process.

Survey Sample

Participants were asked to complete the online surveys after they held their final conference with their evaluator or conducted their final conference with their evaluatee. The surveys were closed on June 22. Tables 1 through 3 provide the number of surveys that were sent, the number of surveys that were completed and the response rate.

Overall, 114 teachers responded to the online teacher survey with a 71% response rate, 19 principals responded to the online principal survey with a 63% response rate, and 5 principal evaluators responded to their survey with an 83% response rate.

Table 3. Teacher Pilot Evaluation Survey Response Counts and Response Rates by Grade and Subject

Grade	Subject	Total Sent	Total Responded	Response Rate	Incomplete Responses
1	Math	7	6	86%	
	R/E/LA	11	7	64%	
2	Math	4	2	50%	
	Science	5	5	100%	
	R/E/LA	6	3	50%	1
3	Math	9	6	67%	
	R/E/LA	8	7	88%	
4	Science	13	5	38%	
5	Math	12	8	67%	
	R/E/LA	7	4	57%	
8	Math	11	9	82%	2
	Science	9	7	78%	1
	R/E/LA	12	12	100%	1
11	Math	14	9	64%	
	Science	10	9	90%	
	R/E/LA	10	5	50%	1
Elementary ESOL and Sp. Ed.	Reading ESOL	2	0	0%	
	Reading Sp. Ed.	2	2	100%	
	Math Sp. Ed.	1	1	100%	
Middle School ESOL/Sp. Ed.	Reading ESOL	2	2	100%	1
	Math Sp. Ed.	2	1	50%	
High School ESOL/Sp. Ed.	Reading ESOL	2	2	100%	1
	Math Sp. Ed.	2	2	100%	1
Total		161	114	71%	9

Note: Incomplete responses are included in total number responded and response rate calculations.

Table 4. Principal/Supervisor Pilot Evaluation Survey Response Counts and Response Rates

Level	Total Sent	Total Responded	Response Rate	Incomplete Responses
Elementary	17	9	53%	2
Middle School	4	2	50%	
High School	5	4	80%	
Other	4	4	100%	
Total	30	19	63%	2

Note: Incomplete responses are included in total number responded and response rate calculations. Others include principals and supervisors for whom there was no indication of the grade level/s at which they performed evaluations.

Table 5. Superintendent/Principal Evaluator Pilot Evaluation Survey Response Count and Response Rate

Total Sent	Total Responded	Response Rate	Incomplete Responses
6	5	83%	1

Note: Incomplete responses are included in total number responded and response rate calculations.

Survey Results: Likert Items

When interpreting the results it is important to consider the context of the pilot evaluation system. Typically, evaluation systems commence at the beginning of the school year, but for this pilot it was only possible to implement the system much later in the school year. Training took place in January and February with the evaluation process followed.

The teacher surveys consisted of 36 Likert items and 2 constructed-response items. The principal survey consisted of 23 Likert items and 2 constructed-response items regarding the teacher evaluation process, and 33 Likert items and 2 constructed-response items regarding the principal evaluation process. The principal evaluator survey consisted of 23 Likert items and 2 constructed-response items regarding the principal evaluation process. The Likert items had a scale consisting of 6 levels:

- 6 – Strongly Agree
- 5 – Agree
- 4 – Somewhat Agree
- 3 – Somewhat Disagree
- 2 – Disagree
- 1 – Strongly Disagree

Table 6 provides a summary of the teacher responses and the principal responses to the Likert items regarding the teacher evaluation process. The range of the mean scores for the items in each of the survey areas is provided. Appendix B provides the responses to each of the items in the survey. Responses to each of the items in Appendix B will be valuable in guiding the revision of the evaluation process.

In evaluating these results it is important to consider that the average teacher score for the item “*I am satisfied with my overall evaluation*” is 4.81 which is close to the Agree level of 5.

Table 6. Teacher and Principal Mean Scores for the Likert Items on the Survey Regarding the Teacher Evaluation Process

Survey Area	Teacher Mean Score Range (n=107 – 114)	Principal Mean Score Range (n=14)
Training Effectiveness	4.10 – 4.20	3.50 – 3.86
Rubric/criteria Quality	4.04 – 4.58	4.29 – 4.79
Evaluation Process Quality (e.g., observations, conferencing)	4.08 – 4.80	4.36 – 4.71
Feedback Quality/Amount	4.36 – 4.68	NA
Support Teaching	4.02 – 4.21	4.50 – 4.64
Operational Ease/ Worth Effort	3.79 – 4.54	3.64 – 4.36
Process Should Remain the Same	3.06 – 3.68	3.14 – 3.79

Both the teacher and principal mean scores for the items typically were between Somewhat Disagree (3) and Agree (5). Principals were slightly less positive than teachers in the training for the teacher evaluation process, with the mean scores for principals between Somewhat Disagree (3) and Somewhat Agree (4). Both teachers and principals responded, on average, between Somewhat Agree (4) and Agree (5) on items pertaining to the rubrics and criteria. For example, the teacher mean score for “*The training provided me with sufficient understanding of the Teacher Evaluation Process*” is 4.17 (Somewhat Agree), while the principal mean score is 3.50. The teacher mean score for the item “*I am satisfied with the (teacher) rubrics*” is 4.04 (Slightly Agree), while the principal mean score is 3.50. The mean score for the item “*The 11 Priority Components and the 4 Performance Domains adequately define good teaching*” is 4.33.

Both teachers and principals responded, on average, between Somewhat Agree (4) and Agree (5) with statements suggesting that the evaluation process supports good teaching. Both teachers and principals responded, on average, between Somewhat Disagree (3) and Somewhat Agree (4) with items suggesting that the process should remain the same. As an example, for the item “*The Teacher Evaluation System should be continued in its current form,*” the mean score for teachers is 3.21 (Somewhat Disagree) and 3.29 for principals.

Table 7 provides a summary of the principal responses and the principal evaluator responses to the Likert items regarding the principal evaluation system. The range of the mean scores for the items in each of the survey areas is provided. Appendix B provides the responses to each of the items in the survey. Results for each of the items will be valuable in the revision of the evaluation system. In evaluating these results it is important to consider that the average principal score for the item “*I am satisfied with my overall evaluation*” is 4.36, which is closest to the Somewhat Agree level (4).

Table 7. Principal and Principal Evaluator Mean Scores for the Likert Items on the Survey Regarding the Principal Evaluation Process

Survey Area	Principal Mean Score Range (n=14 – 17)	Principal Evaluator Mean Score Range (n=4 – 5)
Training Effectiveness	3.12 – 3.59	3.60 – 4.60
Rubric/criteria Quality	3.50 – 4.53	4.25 – 5.00
Evaluation Process Quality (e.g., observations, conferencing)	3.94 – 5.00	4.25 – 4.75
Feedback Quality/Amount	4.33 – 4.87	NA
Support Leadership	3.93 – 4.13	4.75 – 5.00
Operational Ease/ Worth Effort	3.33 – 4.64	3.75 – 4.25
Process Should Remain the Same	2.93 – 3.50	3.00 – 4.25

Both the principal and principal evaluator mean scores for the items typically were between Somewhat Disagree (3) and Agree (5). Principals responded, on average, between Somewhat Disagree (3) and Somewhat Agree (4) to items pertaining to the effectiveness of training. Principals responded, on average, between 3.50 and 4.53 for items pertaining to the rubrics and criteria. For the statement “*The training provided me with a sufficient understanding of the Principal Evaluation Process,*” the principal mean score is 3.24, while the principal evaluator mean score is 4.20. For the statement “*I am satisfied with the rubrics,*” the principal mean score is 3.50 (between Somewhat Disagree and Somewhat Agree) and the principal evaluator mean score is 4.50 (between Somewhat Agree and Agree).

Overall, both principals and principal evaluators responded, on average, between Somewhat Agree (4) and Agree (5) with statements suggesting that the evaluation process supports effective leadership. They responded, on average, between Somewhat Disagree (3) and Somewhat Agree (4) with items suggesting that the process should remain the same. As an example, the principal mean score for “*The Principal Evaluation System should be continued in its current form*” is 3.00 (Somewhat Disagree) and it is 3.75 for the principal evaluators.

Survey Results: Constructed-Response Items

Two items on the teacher and principal surveys pertaining to the Teacher Evaluation System asked respondents to comment on what should remain the same and what should change in the teacher evaluation process. Tables 8-11 summarize these comments made by teachers (Tables 8 and 9) and principals (Tables 10 and 11).

Table 8 provides a summary of the teacher comments on what should remain the same in the teacher evaluation process. Overall there were 110 distinct comments.

Table 8. Teacher Survey Comments on What Should Remain the Same for Teacher Evaluation Process (n=104)

Comment	Frequency
Rubrics and expectations (most of the rubric)	23
Meeting with and reflecting with principals/Amount of feedback	21
Danielson’s Framework and Domains	18
Multiple Observations	16
Self-Reflection/self-evaluation	9
Walkthroughs	2
Overall Process	2
Paperwork at appropriate level	1
Could lead to great staff interaction and professional development	1
Objectivity of process	1
Teacher input	1
Keep most or all of it the same	11
Keep nothing	5
Total Comments	110

The most frequently cited areas to keep mostly the same, with some modifications, were the rubrics, meetings with the principals, Danielson’s framework, the need for multiple observations, and time for self-reflection. As an example, 23 teachers indicated that the rubrics and expectations should remain the same. Some teacher comments regarding the rubrics include:

The rubrics that describe what teacher/classroom evidence demonstrates what level of performance in each domain, once shared with a wider group of our teachers, will lead to many great staff conversations and professional development opportunity to discuss what exemplary teaching looks like. I would keep the categories and descriptors mostly the same.

The rubrics are instructional and informative.

The rubric makes the goals very clear for teachers

Principle 3 – Appendix C

The rubric is clear and should be kept

Teachers also cited the importance of the conferences with the principals (n=21) and the multiple observations (n=16). Teacher comments on what should remain the same regarding conferences and multiple observations included:

*The amount of conferences and walk throughs helped the principal understand what went on in my classroom on a day-to-day basis rather than Isnapshot.
Interaction with principal is very good*

*The format of the observations and interview with the principal
Number of observations*

I think it is important to keep multiple observations. An administrator can't walk into a classroom one time and automatically get a picture of how a teacher teachers The demeanor of the class can change on any given day which may alter the success of the lesson. I think it is important that administrators are in the classrooms constantly and consistently to truly know how a teacher teaches. I think it is also important to receive feedback from an administrator about the success of your lesson. However, the feedback must be constructive. How can you improve to a 4 without knowing what to do to improve?

Teachers also commented on the benefits of self-reflection (n=7), the appropriateness of the domains (n=18), and the overall process:

I like the idea of self-reflection and getting input from the principals. I also like how it is broken down into the rubric. I just wish the rubric was more practical and specific.

I loved the time I was able to reflect and review with my principal to help better my lessons. I came up with some great ideas after my principal and myself would meet and reflect on my lesson using the rubric from the pilot. Using the rubric made me see what things I was missing and what areas I was very strong in.

Reflection process – I felt it was valuable to reflect and think about my teaching strategies, students, parent communication, and classroom environment. I also felt the conferencing with administration was valuable.

I liked the opportunity to reassess myself and have time to reflect with my principal. I think it can make a good teacher better

Teachers reflecting with the principal

I feel the domains are fair and equitable

The four domains were adequate and explained what “good teaching practices” should look like

The major objectives of the evaluation process

Overall, I am satisfied with the Teacher Evaluation Process

The overall process was good- much better than the school's prior method of evaluating teachers.

I would keep the basic structure of the program in place while tweaking some parts of it.

Table 9 provides a summary of the comments on what should change in the teacher evaluation process. There were 158 distinct comments. While there were 23 comments indicating that the rubrics should mostly remain the same, there were 60 comments indicating that the rubrics/criteria should be changed in some way. As an example, 23 comments indicated one or more of the following: rubrics need to be clearer, shorter, easier to follow, more specific, and reflect more realistic expectations. Another 14 comments indicated that there is a need for better explanation of domains, types of evidence, and/or how to get from one level to the next. Another 7 comments indicated that the levels 3 (Proficient) and 4 (Advanced) need to be better distinguished or collapsed.

The following are teacher comments regarding the rubrics:

The wording on some of the objectives is too wordy and not clear and to the point of what a 4 should be.... Each objective must be clear and understanding what is expected of the teacher to be distinguished... we weren't 100% sure what "using assessment in a sophisticated manner" meant or how that is demonstrated ... or..."each student is highly motivated and highly intellectual throughout the lesson"

The evaluation rubric is very vague in the fact that it does not list specific actions for the teacher to model. It would be nice to see specific teaching ideas to make this more practical. I would like more of the how instead of the what.

It would be nice to have more examples for each of the criteria/domains so that you know exactly what to expect or what the principals/ supervisors are looking for.

Either the four domains as stated need to be severely trimmer or the administrative staff needs to set aside many more days of observation for an accurate and meaningful evaluation

Specific evidence should be listed so teachers know what is being looked for. It's also a very cumbersome document and it took hours.

Need to have more clarifications of expectations of each level; specific examples of distinctions between levels

Too much information and no examples were provided in exactly what evidence was expected or how to fill out the rubric.

The evaluation form isn't easy to follow and some of the criteria don't lend itself to all subject areas; such as shared reading

More clear examples of criteria, some items are very open to interpretation

The extensive workload necessary for completing all the forms. Make the forms more concise and user friendly.

Table 9. Teacher Survey Comments on What Should Change for the Teacher Evaluation Process (n=104)

Comment	Frequency
<u>Rubric/Criteria/Evidence:</u>	
Rubrics need to be clearer, shorter, easier to follow, more specific, more realistic expectations	23
Better explanation of domains/examples/types of evidence/how to get from one level to the next (e.g., from 3 to 4)	14
Rubric level 3 and 4 hard to distinguish/combine or clarify differences	7
Criteria not appropriate for some grades/subjects/special education	3
Add/emphasize sections on rubric (teacher's role in student ownership for learning, teacher content knowledge, effectiveness of instruction)	3
Deemphasize sections on rubric – lesson plans	2
Principal comments redundant within rubric	2
<u>Observations/conferences:</u>	
Poor timing of observations	13
More observations/ Follow-up Observations	10
More feedback/conferences/communication/pre-observation conference	8
More walkthroughs	6
Have multiple Evaluators/less subjective	4
Every lesson cannot reflect all components/ focus on different components across observations or years/teachers cannot adequately demonstrate	4
Observe in different settings/classes	3
Have announced observations/have unannounced observations	2
Planning and Preparation/Professional Responsibility should not be part of the observation , need conferences with principals for these sections	1
Less conferences	1
More input in evaluation from teachers	1
<u>Training:</u>	
Need more/better training/in-service	10
<u>Operational Ease:</u>	
Time consuming/need to streamline process	11
Less paper work	2
<u>Other:</u>	
Overall process poor (not organized, not a random group of teachers chosen, Danielson's Framework assumed to be the best – need control group)	3
Need time to reflect/discuss with colleagues	2
Teacher evaluation on a cycle	2
Process should include student data/should not include student data	2
Did not obtain feedback or have conferences	2
Include a teacher portfolio	1
No changes/No comment	10
Many changes	1
Total Comments	158

Principle 3 – Appendix C

There were 52 comments regarding the observations and conferences. Ten comments suggested the need for more observations and 6 indicated the need for more walkthroughs. Poor timing of observations was indicated by 13 teachers. This may have been due to the timeframe of the evaluation process (late January/early February to the end of May/beginning of June). Several teachers also suggested that observations should occur in different settings/classes and 8 indicated the need for more feedback and conferences. Teacher comments regarding the need for changes in the observation and conferencing process include:

More observations – some announced, others not announced, maybe with different evaluators

Two brief visits are not enough to make an accurate evaluation of instruction and planning that takes place throughout an entire year. I would rather have a series of evaluations by different administrators throughout the year, less informally with time to discuss my reflections and adjustments made as a result. I would also like to be evaluated and discuss lessons with colleagues.

I believe that one person should not determine our effectiveness as a teacher. I would like to see several people do evaluations and there should also be several and maybe an entire week of these observations.

Principals would need to see more than one lesson to complete the eval. Perhaps 2 or 3 of the same class over a discrete period of time. This would allow assessment of all classroom systems – formative assessment, discipline, use of technology, etc.

There were 10 comments indicating the need for additional training. Example comments include:

Training – teachers and administrators need to be FULLY trained before the process begins.

I would like to see an example of a teacher at each level

Another training with better understanding of all items in the rubric

The principals/supervisors were also asked to comment on what should remain the same and what should change for the teacher evaluation process. As indicated in Table 10, principals provided 18 distinct comments when asked what should remain the same in the teacher evaluation process. There were 8 comments suggesting that the rubrics should remain the same, 3 comments on the value of requiring evidence, 2 comments regarding the value of Danielson's Framework, 2 comments regarding the value of the observation, walkthrough and conferencing process, and 2 specific comments on principals and teachers sharing ratings and reflecting together.

Table 10. Principal Survey Comments on What Should Remain the Same for Teacher Evaluation Process (n=14)

Comment	Frequency
Rubrics (levels of proficiency, key to teacher success)	8
Need to site evidence/documentation	3
Danielson’s Framework and domains	2
Format of observation, walkthrough and conferencing	2
Conference with teacher and principal comparing ratings/self-reflection	2
Keep most or all of it the same	1
Total Comments	18

Principal/supervisor comments regarding what should remain the same with the teacher evaluation process include:

The rubric was key to teacher success

The rubrics. I am familiar with the look-fors within each category and I understand the rating system.

The levels of proficiency

The teacher’s self-evaluation and following discussions between evaluator and teacher, comparing rubrics.

Comprehensive look at components of teaching. The use of reflection by the teachers for collaborative measurement of effectiveness. Rubric helps with look fors. Need to site evidence is great. The use of formal observation and walk through data for comprehensive look is a positive.

Danielson model is a positive.

The format of the observation and conferencing

As indicated in Table 11, the principals/supervisors provided 34 distinct comments when asked what should change in the teacher evaluation process. Twenty-two comments were on the rubric/criteria, 3 comments were on the observation and conferencing process, and 6 were on the training. Comments from the principals are not provided here, because they are well specified in the table.

Table 11. Principal Survey Comments on What Should Change for Teacher Evaluation Process (n=14)

Comment	Frequency
<u>Rubric/Criteria/Evidence:</u>	
Redundancy in rubric, and redundancy within rubric and Summary of Evidence	5
Provide shorter rubric for end-of year summative evaluation/reduce the length/simplify	4
Provide examples of evidence at different levels (especially 3 and 4) – drop down menu	3
Great disparity between levels 2 and 3	1
Provide area for a narrative summary of lesson observed	1
Provide walk-through evaluation rubrics or templates that build on collecting evidence throughout the year	1
Create separate rubric and shorter evaluations for mid-year	1
<u>Format for reporting:</u>	
Wider range of scores (e.g., 2.5)	1
Need area for commendation as well as recommendation	1
Consistency in Scoring is a concern (need mentors for principals)	1
Need to include teacher goals	1
Suggest attachments as evidence (student achievement data, attendance data, etc.)	1
Need to ensure administrator’s evaluation is consistent with teacher’s evaluation (terminology should be consistent)	1
<u>Observations/Conferencing:</u>	
Administrators need to visit classrooms for informal discussion of evidence with teacher prior to use of evaluation forms	1
More time for observations	1
Teachers need to observe colleagues in their areas and document their findings regarding best practices – lead to discussion and collaboration among staff	1
<u>Training:</u>	
Staff needs extensive and better training on Danielson, types of evidence, goal setting, evaluation process. Provide mock video of process of training, supply sample timelines	6
<u>Operational Ease:</u>	
Need entire year	1
Time Consuming/Amount of time is overwhelming	1
Evaluate teachers on a cycle	1
<u>Total Comments</u>	34

Principle 3 – Appendix C

Principals and principal evaluators were also asked to comment on what should remain the same and what changes are needed in the principal evaluation process. Tables 12-15 summarize these comments and frequencies.

Principals made 14 comments on what should remain the same as indicated in Table 12.

Table 12. Principal Survey Comments on What Should Remain the Same for Principal Evaluation Process (n=14)

Comment	Frequency
Rubrics (levels of proficiency, criteria)	5
Need to site evidence/documentation	2
The Standards	2
Conference with supervisor	2
Ability to demonstrate performance on multiple tasks	1
Keep just the Core Standards (maybe)	1
Comprehensive reflection	1
Total Comments	14

Some comments regarding what should remain the same are provided below:

The rubric format was effective and thorough.

The categories of the evaluation

The evaluation rubric was a huge help in allowing me to prepare my evidence for my evaluation.

My ability to demonstrate performance on multiple tasks

Collection of evidence

I think the comprehensive reflection is great. I also like the fact that one is responsible to site evidence.

There were 28 comments on what should change in the principal evaluation system as indicated in Table 13. There were suggestions for the rubric/criteria, evaluation process, training and operational ease. Comments from the principals are not provided here, because they are well specified in the table.

Table 13. Principal Survey Comments on What Should Change for Principal Evaluation Process (n=14)

Comment	Frequency
<u>Rubric/criteria/evidence:</u>	
Redundancy in rubric, redundancy within rubric and Summary of Evidence, descriptors too broad because same evidence used across descriptors, need to collapse areas of redundancy	6
Provide examples of evidence/examples of quality artifacts- use a drop menu	5
Simplify questions and shorten the document	4
Need to be able to provide evidence easily within the document throughout the year	1
Need clarity on the rating process and which ratings need evidence	1
Criteria may need to change depending on the position (principal vs. assistant principal)	1
<u>Evaluation Process:</u>	
More meetings with evaluator/mid-year meeting	2
Notion that all core standards and complementary corollaries are all indicators of high quality performance at all times	1
Evaluation should be completed by evaluatee and evaluator prior to conference for a collaborative approach	1
Provide a mid-year evaluation form/checklist for feedback from supervisor	1
Need to ensure administrator's evaluation is consistent with teacher's evaluation (terminology/expectations should be consistent)	1
<u>Training:</u>	
Comprehensive training for the evaluatee and evaluator needed	1
<u>Operational Ease:</u>	
Time Consuming	2
Need entire year	1
Total Comments	28

Principal Evaluators provided 5 comments on what should remain the same for the principal evaluation process as indicated in Table 14 and 9 comments on what should change as indicated in Table 15.

Table 14. Principal Evaluator Survey Comments on What Should Remain the Same for Principal Evaluation Process (n=4)

Comment	Frequency
Overall, process should be kept the same	2
Rubrics	1
Use core and corollary standards with some consideration for the redundancy that occurs within the descriptors	1
Set goals, ongoing discussions about the evaluation process and evidence of support	1
Total Comments	5

Table 15. Supervisor/Evaluator Survey Comments on What Should Change for Principal Evaluation Process (n=4)

Comment	Frequency
Use the year to arrive at common understanding of the standards descriptors and rubrics, discuss evidence.	2
More conferences with principal throughout the year and increased data collection	2
More examples/evidence	1
Changing rating scale – if principal misses one of the descriptor levels within a rating, give partial credit.	1
Need to ensure administrator’s evaluation is consistent with teacher’s evaluation (terminology/expectations should be consistent)	1
Cycle for principal evaluation	1
More opportunities for practice especially for new evaluators	1
Total Comments	9

Focus Groups and Interviews

In May and June 2011, 6 focus groups with teachers, 6 interviews with principals, and 3 interviews with principal evaluators were conducted across 3 district pilots. The focus groups and interviews were designed to obtain information on the training, rubrics, and the overall system for both the teacher and the principal evaluation systems. Facilitators asked teachers questions pertaining to the teacher evaluation system. Principals answered questions pertaining to both the teacher and principal evaluation system, and principal evaluators answered questions regarding the principal evaluation system.

Focus Group and Interview Sample

Table 16 shows the number of teachers who participated in the focus groups at each grade level and in each subject area. Overall, 34 teachers across 3 districts participated in the focus groups.

Table 16. Number of Teachers Participating in Focus Groups across Three Districts

Subject	Grade		
	1 – 5	8	11
Mathematics	3	4	3
Reading/English/LA	5	7	2
Science	5	3	1
Learning Support-Math			1
Total	13	14	7

Table 17 lists the number of principals/supervisors and principal evaluators who were interviewed at each school level. Overall, 9 principals/supervisors were interviewed for the study.

Table 17. Number of Principals/Supervisors Participating in Interviews across Three Districts

Position	Number
Elementary Principal	2
Middle/Jr High Principal	1
High School Principal	3
Super/Asst. Super	3

Note: 1 HS Principal was also the Jr High Principal

Focus Group/Interview Results: Teacher Evaluation System

The following is a summary of the data gathered through the focus groups and interviews regarding the Teacher Evaluation Pilot. Audio files were coded and analyzed using NVivo 8 qualitative analysis software. Several types of codes were applied to the data, including:

- predetermined content codes – denoted discussion topic (e.g., Training, Danielson Framework, Scoring Rubrics, etc).
- structural codes - Positive reaction, Negative Reaction, and Need More Information (when participants indicated that more information should be provided or asked the interviewers questions about system). These three codes allowed researchers to generate matrices and interpret the content codes meaningfully.

- emergent content codes – recorded suggestions for improvement generated by participants (e.g., More Room for Variability, More Sources of Evidence, and Preconference Form) and allowed sub-categories of discussion topics to emerge (i.e. rubric wording).

Other types of coding allowed researchers to organize data by source (e.g., Principal Interview, Teacher Focus Group) or context (e.g., Teacher Evaluation System, Principal Evaluation System). Tables were generated using compound and matrix queries. The full codebook with detailed coding definitions can be obtained by the author.

This section is divided into three parts: Training, Scoring Rubrics, and System Process. Each section first addresses the participants’ reactions to different components of the pilot system, and then addresses participants’ related suggestions for improving the system. Reference counts are provided (i.e., the number of times each code was applied to the data) as well as a discussion of the thematic content of participants’ comments, along with selectively transcribed representative quotes. Unless otherwise specified, these data include the combined reactions and suggestions of both teachers and principals pertaining to the Teacher Evaluation System Pilot.

Training. Table 18 presents teachers’ and principals’ reactions to the Teacher Evaluation System trainings for evaluators and evaluatees.

Table 18. Reactions to Teacher Evaluation System Training

	Negative Reaction	Positive Reaction	Need More Information
Training	24	7	30
Danielson Framework Training	4	3	9
Evidence Training	5	2	10
Operational Process Training	3	0	2
Scoring Rubrics Training	12	2	17

Note: Cell content represents the number of coding references (or coding occurrences) across all focus groups/interviews which fall at the intersection of the content (row) and structural (column) codes. Comments sometimes warranted the application of several different specific content codes simultaneously or only the general content code, resulting in coding occurrence counts that sum to a number other than the total number of general content coding occurrences.

The data in Table 18 suggest the indication of the need for more extensive training in regards to the Danielson Framework, the types of evidence teachers need to exhibit, and the qualifications for each level of proficiency according to the rubrics. There was considerable confusion regarding the difference between evidence that would determine a *proficient* versus a *distinguished* lesson in many of the domains, and some confusion about the difference between the *needs improvement/progressing* level and the *proficient* level. The distinction between the two lowest levels of proficiency was reportedly clearest. According to the participants’ comments, the operational process (e.g., number and timing of observations, use of evaluation tools) should also be further addressed in the training.

The following is a summary of the different types of positive and negative comments that teachers and principals made regarding the Evaluation System Pilot training.

Positive Reactions to Training

Training was well conducted/designed:

It was a nice way to outline the domains, it was well done.

It was nice that they...had us come up with extra ideas of things that they would be looking for because some of the specifics weren't written on [the rubrics]

Negative Reactions to Training

Too much information was given in too short a time:

It was a lot of information in a short amount of time, so it was a little bit hard to process everything that was presented.

Too much was thrown at us.

It was information overload.

It was real fast, it was kind of thrown at us real quick.

Not enough information was given, content was too abbreviated:

There wasn't enough time the day of the training for us to look and listen...it would have been more fruitful to have this [documented information] in advance.

It could have been in greater detail, more specific.

[The rubrics] were still confusing.

I would like to have seen specific examples of what [evidence] they were expecting—instead of putting us into small groups to figure it out on our own.

I was still confused when I left...I didn't think it was enough.

Logistics of training (e.g., too many trainees, hard to hear/see presenters) prevented effective learning:

Sometimes when you're in a big room like that, it's the fear factor—people don't want to ask questions, they don't feel comfortable.

Training materials were not as relevant or helpful as they could be:

We saw a science [video]—we teach English. It didn't really relate to what we do. The content is completely different.

Table 19 presents suggestions for the improvement of the training for the Teacher Evaluation System which go beyond the revisions that are implied by Table 18.

Table 19. Suggestions for Improving Teacher Evaluation System Training

Suggestion	Frequency
Training over longer time period	7
Train evaluators first	5
Train in-district trainers	4
More video clips	4
Video clips tailored to content area/grade level	2
Video clips of different proficiency levels	1
Provide Danielson's book	1

Focus group/interview participants suggested that the training would be more beneficial if it took place over a longer period of time and included multiple sessions, rather than being positioned as a “crash course” on the entire evaluation system. Participants reported that this would allow them to better

understand the Danielson Framework, the rubrics, and the types of evidence teachers need to exhibit in the classroom to receive high scores. One teacher said:

I think a longer amount of time would have helped because they could have done more with each individual section so that we had a better understanding.

Another participant commented on the novelty of this framework and the need for more intensive training in the following way:

If you haven't had training on what they are trying to move into this new development, then your classroom doesn't look anything like this...this isn't like it was ten years ago...coming out of college...this is nothing we talked about...this has all changed. If you haven't had the practice of it and the teaching of it, this would have been tough for someone like that.

Participants suggested that they would value the availability of trained evaluators and/or in-district personnel specifically prepared to conduct training sessions/follow-ups, provide information, and answer questions about the Evaluation System.

Some participants suggested the inclusion of more examples of evidence in action. Specifically, more video clips were desired, as well as improved relevance and value of the video clips used in trainings (pertaining to relevant content area, relevant grade level, and differentiated proficiency level). For example, one teacher commented that:

The video they showed was not an excellent example...it was just your basic classroom. If we could have seen the different levels—this would be a number one, this would be a number two, — I think we would have had a better understanding.

Rubrics. Table 20 summarizes teachers’ and principals’ reactions to the Teacher Evaluation System scoring rubrics, including content and functionality.

Table 20. Reactions to Teacher Evaluation System Scoring Rubrics

	Negative Reaction	Positive Reaction	Need More Information
Scoring Rubrics	28	16	5
Danielson Framework	7	11	9
Lesson Plan Requirements	3	0	0
Rubric Format	6	0	0
Rubric Wording	23	2	8

Note: Cell content represents the number of coding references (or coding occurrences) across all focus groups/interviews which fall at the intersection of the content (row) and structural (column) codes. Comments sometimes warranted the application of several different specific content codes simultaneously or only the general content code, resulting in coding occurrence counts that sum to a number other than the total number of general content coding occurrences.

Some teachers and principals indicated that they found the rubrics to be “wordy,” and “cumbersome,” and reported difficulty distinguishing between the different levels of proficiency based upon the descriptions in the rubrics. For example, one principal commented:

I'm having trouble picking out the difference between threes and fours.

Other participants also indicated that the wording made it difficult to decide on the appropriate proficiency level to assign, for example:

It says with the 2--even if I admit that I have "minimal communication" [with parents] but then it says...uh (reading from rubric) "insensitive communication." Well no--I might have minimal, but it's not insensitive!

When you go through 1, 2, and 3, those are all solid things I think you could check off. When you get to 4, it almost seems like it's very subjective...and I think that has a lot to do with your relationship with the Principal.

So (reading from rubric) "little knowledge...some awareness...solid knowledge" and then you have "extensive knowledge." I think all of those you could prove up to extensive, and then what do you determine extensive and what does your principal determine extensive? Does it depend on what he's already seen you do, or does it depend on what class he was in before he came into your room?

Additionally, evaluators reported difficulties with the rubric format, and several of them described how they created revised versions for personal use during the pilot. They recommended formatting modifications such as re-organizing the document to reduce redundancy, eliminating much of the text and using only keywords, underlining/bolding keywords, and adding informative headings to compliment numerical/alphabetical labels.

In addition to clarifying and simplifying the rubric wording and format, participants identified other ways to improve the scoring rubrics. These suggestions and frequencies are presented in Table 21.

Table 21. Suggestions for Improving Teacher Evaluation System Scoring Rubrics

Suggestion	Frequency
More Evidence Sources	19
Pre-Conference Form	8
Space for Recommendations and Commendations	6
Expand System to Apply to Different Roles	5
Need More Room for Variability	5
Additional Components Valuable (those listed but not included in pilot should be included in system)	2
Alignment Between Principal and Teacher Evaluation Rubrics	1
Electronic Tools	1

The most common suggestion was the addition of other sources of evidence, such as portfolios and pre-conference forms for teachers to complete. Many participants indicated that they did not think the observation process alone could capture the efficacy of a teacher. Likewise, many teachers and principals indicated that it would not be possible to complete the rubrics fully with only observational data. Most discussed was Domain 4: Professional Responsibilities; participants indicated that these facets of a teacher's role would not be apparent to an evaluator during classroom observations. Evaluators reported relying on the post-conference conversations to complete those portions of the rubrics which were not feasible to complete during observations. For example, one principal commented:

I could not do section 4 without talking to the teachers and having them bring me information.

It kind of made us discuss our evaluation process as far as--we do the walk throughs and we do the classroom observations but some of the evidence it's almost on the teachers as far as maybe creating a portfolio

Participants also suggested a structured approach that would guide teachers in the collection of evidence:

If you kinda just made a little rubric that said 'this is what you need to keep in the portfolio' ...keep it simple, so we could follow it.

System Process. Table 22 summarizes teachers' and principals' reactions to the entire Evaluation System process.

Table 22. Reactions to Teacher Evaluation System Process

	Negative Reaction	Positive Reaction	Need More Information
Time and Effort	10	10	0
System Fairness	10	9	0
Evidence	6	8	11
Observation and Conference Frequency	9	4	1
Evaluator Feedback	6	15	3
Self-Assessment and Self Reflection	2	10	2
Professional Development Support	2	11	4
Operational Process	8	0	3

Note: Cell content represents the number of coding references (or coding occurrences) across all focus groups/interviews which fall at the intersection of the content (row) and structural (column) codes.

Because the positions on many of these components were likely different for principals and teachers due to their opposite roles, several of them will now be examined by principal reaction versus teacher reaction to provide a more complete and accurate picture of the data. The positive and negative reactions to time and effort, system fairness, evidence, observation and conference frequency, and evaluator feedback will all be analyzed in this way, while participants' reactions to other components (self-assessment and self-reflection, professional development support, and operational process) were more equally represented by teachers and principals and will therefore be discussed across the two groups.

Table 23 provides a summary of the reactions to the amount of time and effort the Teacher Evaluation system required of teachers and principals. Participants responded according to the amount of time they were required to spend in the role they fulfilled during the pilot. They were asked to think about the time and effort that would be required of them if the process were to take place over an entire year.

Table 23. Reactions to Time and Effort Required by Teachers and Principals

	Positive Reaction	Negative Reaction
Teachers	6	4
Principals	4	6

The following is a summary of the different types of positive and negative comments that teachers and principals made in regards to the time and effort the system required of them.

Positive Teacher Reactions to Time and Effort:

Virtually no extra time needed to be spent by teachers in order to participate in pilot outside of the training and conference(s) with evaluators.

I don't think we had to put any more effort into it than we already do...it was easy on our part, there wasn't really anything additional.

Negative Teacher Reactions to Time and Effort:

The system demands too much time expenditure in regards to the formal written lesson plan requirements.

Too much time needed to be spent becoming oriented to the system on the teachers' own time due to lack of rigorous training.

It was a whole lot of effort...the fact that we didn't know all the specifics, and it wasn't explained really well, it took us extra time to do that.

Conferencing was more rigorous than conferencing processes currently in use.

Rubrics were cumbersome documents to sift through to understand expectations (pre-observation) and scores (post observation).

Positive Principal Reactions to Time and Effort:

The time/effort issues were due to short timeline of pilot, not due to system:

[Time and effort] would have been [reasonable] if we had more time...if it was spread out over the whole year.

Time and effort requirements were comparable to current protocols in use.

Negative Principal Reactions to Time and Effort:

Providing the rigorous and specific feedback was labor intensive.

...I felt that it was really time consuming...I guess trying to provide the feedback to each of those domains took me a long time, because I wanted to get it right, I wanted to be as specific as I could...

The volume of observations/evaluations each administrator needed to perform was too cumbersome.

Table 24 provides a summary of Teachers' and Principals' reactions to the fairness of the Teacher Evaluation system.

Table 24. Reactions to System Fairness by Teachers and Principals

	Positive Reaction	Negative Reaction
Teachers	4	10
Principals	5	0

Principle 3 – Appendix C

The following is a summary of the different types of positive and negative comments that teachers and principals made in regards to the fairness of the system.

Positive Teacher Reactions to Fairness:

Relative to other systems, there was more room for variability (versus a satisfactory/unsatisfactory system) and more information was collected about each teacher.

I'd say between this one and the one we had before, this one was 'true.'

Negative Teacher Reactions to Fairness:

It is impossible to display all evidence needed to receive a *distinguished* rating in a single lesson; insufficient opportunity to present evidence of practice.

2d, managing student behavior...now the 4 is (reading from rubric) 'standards of conduct are clear with evidence of student participation in setting them.' So, for the 4, the kids have a chance to set some of the behavior stuff, but who's going to see that? That's going to be set a long time ago.

Rubric scores are subjective, too much room for interpretation, concerns about inter-rater reliability.

How could it be fair? There are so many different subjective terms in each of the categories that are open to interpretation by the evaluators themselves...how could it possibly be consistent and fair given the vague nature of a lot of the wording in the rubric?

Positive Principal Reactions to Fairness:

Rubrics are less subjective than existing system in place; not as easy for administrators to “go after” teachers.

More evidence-based than current system.

Table 25 provides a summary of Teachers' and Principals' reactions to the evidence required by the Teacher Evaluation system.

Table 25. Reactions to Evidence by Teachers and Principals

	Positive Reaction	Negative Reaction
Teachers	4	4
Principals	2	4

The following is a summary of the different types of reactions to the evidence requested by and gathered in accordance with the system.

Positive Teacher Reactions to Evidence:

If you're doing what you're supposed to be doing, you're covered.

I think everything on the rubric was something a good teacher should be doing, and I think it was all observable—except maybe for the last...the fours—those are things I don't know that in a snapshot a principal would be able to know, like data and talking to families and things like that, but everything else I thought was important to be seen and could be seen.

Negative Teacher Reactions to Evidence

This was a two shot deal...I don't think it's enough [to provide sufficient evidence of teaching].

To hit all of these [indicators] in one lesson in one day? It's hard! Why am I getting a 2 because one day I forgot [a certain material]?

Positive Principal Reactions to Evidence:

Interviewer: Do you think that teachers were able to provide the evidence that they needed to show that they were effective teachers?

Principal: I think so...if you look at it from a unit standpoint rather than a lesson standpoint, they had more opportunity to do that.

Negative Principal Reactions to Evidence

To find evidence sometimes in one visit's tough.

How do I know if students are self-assessing during just that one period? It might have just been a brainstorming session at the beginning [of a unit].

Table 26 provides a summary of Teachers' and Principals' reactions to the frequency of the Observations and Conferences. Because the number of observations/conferences varied among participants and therefore satisfaction with the experienced number of these events is relative to the number different participants experienced, a summary of the desired frequency reported by teachers and principals will be provided following the table.

Table 26. Reactions to Observation and Conference Frequency by Teachers and Principals

	Positive Reaction	Negative Reaction
Teachers	1	7
Principals	3	2

The following is a summary of the different types of comments that teachers and principals made in regards to the number of observation and conference opportunities they received as well as the number they reported would be sufficient. Most teacher comments indicated that the number of formal observations/walk throughs/conferences they experienced (1 or 2) was not sufficient, and facilitators asked them to think about the type and amount of contact they would like to have with evaluators if the process were to take place over a year.

Most teachers reported that only one observation using the rubrics would be insufficient and would not allow them to demonstrate the evidence needed to receive a high score, or allow them to show improvement based on feedback received. Many teachers made comments similar to this:

*I think having multiple observations is getting a better picture of what’s going on.
We have to have multiple—not one or two a year—but multiple evaluations throughout the year.*

Many teachers suggested about four observations per year, with a mix regarding how many should be formal observations versus informal walk-throughs. Some teachers reported that they would like to have informal walk-throughs as often as possible, even as often as one per week or month. Generally, teachers suggested holding a conference after at least each formal observation; some teachers indicated that this may not be reasonable in terms of the principals’ time expenditure.

Principals, on the other hand, generally reported that one or two formal observations per year would be a reasonable undertaking, and had mixed opinions regarding whether this would be enough to gain a sufficient picture of teachers’ overall practice. Some thought that the number that they had done for the pilot would be sufficient for a full year, while others suggested increasing the number. For example, one principal said:

I would say more than one walk-through, but one formal observation. I would say three or four walk-throughs at least, a year.

Another principal suggested two formal observations and one walk-through.

During these discussions, both teachers and principals expressed the desire for more opportunities to present different types of evidence of practice, with the most common suggestion being the use of a portfolio system to compliment the observational data.

Table 27 provides a summary of teachers’ and principals’ reactions to the feedback from evaluators prompted by the evaluation system process. Teachers were asked about the feedback they received, and evaluators were asked about feedback the system solicited from them. Facilitators asked participants to take into account the amount and quality of feedback, as well as helpfulness and applicability to practice.

Table 27. Reactions to Evaluator Feedback by Teachers and Principals

	Positive Reaction	Negative Reaction
Teachers	9	1
Principals	6	5

As suggested by Table 27, many teachers indicated that they valued the feedback they received, saying things like:

He wrote for me like this little written report, which was helpful.

Some teachers compared the feedback prompted by this system to feedback prompted by existing systems, and reported that their current systems offered virtually no feedback, or feedback that was perfunctory and did not lead to professional development, while the pilot system prompted more thorough and helpful feedback. For example, one teacher commented that:

Principle 3 – Appendix C

With our observation now we get little one or two sentence blurbs under each thing...it was nice to see the whole, drawn out...

However, during one discussion with teachers they explained that feedback given by someone with expertise in their specific content areas would have been more helpful.

Principals, on the other hand, reported mixed opinions of the feedback prompted by the system. Some comments indicated that the feedback the system demands from principals is too detailed and cumbersome, but other comments indicated valuation of the opportunity to help teachers to self-reflect in such a rigorous way. For example, one principal said:

I think a checklist would be...easier, and more manageable, where you're just able to check the various things you've seen, or highlight the various parts, then give that as feedback rather than writing up paragraphs... where yes, I have a template I use where I am cutting and pasting some of the domains, but uh, I think a checklist might be easy--just help with the time management, help with the--even finding the evidence, alright, I saw that, I saw that, I saw that, I mean, and keep going down each domain. I think that would be easier in my opinion. Or, more manageable.

Other comments indicated that principals thought the system had prompted them to give feedback that was helpful for teachers; for example:

The very most important part of the whole process was talking with the teachers...we really had points to look at... We never before had something to guide that conversation...it was good...the discussion..

The best way to describe it would be, um, it gave us a better framework to talk and discuss the lesson, it's much better than our current observation system that really touches on things that are not really about instruction, for example, professional dress, fire drill safety issues, attendance, where this gets you right to the instruction, to the meat and potatoes, if you will, of the lesson. So I think it was a helpful framework to have a better conversation with the teachers.

Two principals indicated that there was not a clearly delineated space to provide constructive criticism balanced with specific commendations, which they suggested adding during future revisions.

Teachers and principals reported valuing the opportunities for teachers' self-assessment of and self-reflection on their own teaching practice. Participants reported that the Danielson Framework allowed for richer discussions of and reflections on specific aspects of teaching practice. Comments regarding teachers' input during conferences with evaluators were mostly positive. For example, one teacher said:

I think it's made you, as a teacher, more reflective...thinking about these domains...you just thought about your teaching... 'wow, maybe I should do this or that more'...

A few participants reported the desire for more formalized procedures to aid in the self-assessment and self-reflection process, such as preconference forms which would provide structured space for teachers to self-evaluate meaningfully, and would be brought to each conference to inform evaluative discussion and results. One participant commented:

They need a preconference from that they need to complete based on what's in here...to say 'this is what we're looking for, this is what we want you to think about.'

When participants discussed the operational process (e.g., conference/observation timing and procedures), most indicated that it was either unclear or problematic in some way; for example, one teacher said:

There were a couple times where I was observed and I didn't get a chance to talk to [my evaluator] for two weeks! And by the time we sat down to talk I said "I don't even know what my lesson was," I said "I don't know what I taught, you'll have to refresh my memory."

Participants also commented that the procedures were inconsistent among evaluators and between evaluatees, and indicated the desire for more clarity regarding logistical details and procedural expectations in future revisions.

Most comments regarding the capacity of the Pilot Evaluation System to support and encourage professional development were positive. Participants often contrasted the pilot system with existing systems in terms of the support for professional growth, making comments such as:

This helps fill the hole...like how kids have knowledge gaps—we do too, so it kind of like 'ok, I'm a four or a three here, but not so much here so I need to work on this'—like that's just self-improvement.

In the past, we'd just go in there and sign a form and leave.

However, a small number of teachers also made comments regarding the low utility of knowing at what point they fell on a four point scale without knowing specifically how to improve.

Table 28 includes the suggestions made regarding the improvement of the evaluation system process, and the number of times the suggestions were made across focus group sessions.

Table 28. Suggestions for Improving Teacher Evaluation System Process

Suggestion	Frequency
Time With Colleagues to Discuss Evaluation System Components	13
Stress Positive Formative Goals	5
Multiple Evaluators	2

As shown in Table 28, the most frequent suggestion for the improvement of the evaluation system process was the allotment of time to spend with colleagues (i.e., teachers meeting with other teachers, principals meeting with other principals) discussing the evaluation system. Different formats were suggested; some teachers wished to meet with other teachers in their own subject areas, while others wanted to meet with teachers across subjects—especially those in small districts with very few teachers in each subject. Participants explained that this would allow everyone to “get on the same page” about the meaning of rubric items and system process, encourage professional development, and help with inter-rater reliability among evaluators. One participant cautioned that while it would be beneficial to meet with peers to “get on the same page,” doing this without formal guidance or the presence of a trained expert on the system may result in groups of teachers or principals with a common but incorrect understanding of system components.

While most concerns about inter-rater reliability were focused on the training rigor and the rubric wording, the desire for multiple evaluators was also expressed. Participants who advocated this approach

thought that it would improve system fairness and allow for a diversity of feedback. Some teachers indicated the desire for an opportunity to receive feedback from an expert in their own subject areas.

Finally, multiple participants suggested that more should be done to stress the formative goals and professional development opportunities during the evaluation process. Some participants indicated the belief that the evaluation process would be primarily used for punitive purposes. Many participants indicated that achieving the “distinguished” level of proficiency across domains would be impossible. Participants suggested that making sure teachers and principals understand the implications of the evaluation results would alleviate this uncertainty. For example, one principal made the following comment, which was congruent with statements made by several others:

I think a gray area that some folks might agree with is the gray area between proficient and distinguished, where I don't want teachers to feel insulted that the lesson wasn't distinguished, because proficient is an excellent lesson. I think maybe a little more talking on hey--distinguished is phenomenal, unbelievable, is a teacher going to be distinguished every day in every lesson? I don't think so in my opinion. I think maybe a little more explanation on you know, here's what distinguished means, but if you're proficient that's fantastic too, I guess. I was just worried as I'm doing these post observations and informal observations, that I commented mostly on proficient things, I explained to them 'hey, you're moving towards the distinguished level on this, this, and this' and I hope they didn't take that as 'ah, I'm not distinguished, he thinks I'm terrible.' So, I don't know if that can be explained better to teachers. I don't know the answer to that.

Similarly, one teacher indicated the need for teachers to understand that “it’s ok not to get a four,” saying:

I think that's why this works because we are honestly admitting that like—I have threes in here because, like here it says 'did you talk with parents' ...did I call home as much as I should have? No—I told [my evaluator] I need a three because I regret not calling home more than I did.

Summary. Most comments regarding the training indicated that participants did not find that the training sessions adequately prepared them for the pilot. Participants made suggestions for improving the training in the future, including introducing a revised training format that is tailored to different teachers’ grade levels and content areas, and takes place over a longer period of time with multiple sessions. While many participant comments about the rubrics indicated that they valued the content—specifically, the Danielson Framework—many constructive comments were made about the rubric wording. Participants suggested improving the specificity, clarity, and usability of the rubrics. Participant reports regarding the system process were mixed and reports were often different for teachers and principals (e.g., regarding system fairness, observation/conference frequency, and feedback). The most common suggestion was the use of multiple forms of evidence to compliment observational data, such as portfolios, artifacts, or pre-conference forms.

Focus Group/Interview Results: Principal Evaluation System

This section addresses the results of the interviews with principals and principal evaluators in regards to the Principal Evaluation System pilot. It is divided into three parts: Training, Scoring Rubrics, and System Process. Each subsection first addresses the participants’ reactions to different components of the pilot system, and then addresses participants’ related suggestions for improving the system. Reference counts are provided (i.e., the number of times each code was applied to the data) as well as a discussion of the thematic content of participants’ comments, along with selectively transcribed representative quotes. To maintain confidentiality, these data include the combined reactions and suggestions of both

principals/supervisors (evaluatees) and superintendents/principal evaluators pertaining to the Principal Evaluation System Pilot.

Training. Table 29 presents principals’ and principal evaluators’ reactions to the Principal Evaluation System trainings for evaluators and evaluatees.

Table 29. Reactions to Principal Evaluation System Training

	Negative Reaction	Positive Reaction	Need More Information
Training	6	1	10
Core and Corollary Competency Training	2	0	4
Evidence Training	2	0	5
Operational Process Training	1	0	1
Scoring Rubrics Training	1	1	5

Note: Cell content represents the number of coding references (or coding occurrences) across all focus groups/interviews which fall at the intersection of the content (row) and structural (column) codes. Comments sometimes warranted the application of several different specific content codes simultaneously or only the general content code, resulting in coding occurrence counts that sum to a number other than the total number of general content coding occurrences.

Positive Reactions to Training:

I understood the rubric, yes, and I think there was enough time spent on how the rubrics worked, yes.

Negative Reactions to Training:

It was rushed...it was difficult, I think...maybe having training in summer at the beginning of the year would have been—but of course, that wasn’t how this was taking place.

I think maybe it needs to be a different day [from the teacher evaluation training]...I can barely remember what they talked about! I can’t even remember the second part...did we do more of Charlotte Danielson?

I didn’t know exactly what made me look like a four or look like a three.

In addition to the provision of additional information pertaining to the topics listed above, one individual suggested improving the training by extending it over a longer time period.

Rubrics. Table 30 summarizes principals’ and principal evaluators’ reactions to the Principal Evaluation System scoring rubrics, including content and functionality.

Table 30. Reactions to Principal Evaluation System Scoring Rubrics

	Negative Reaction	Positive Reaction	Need More Information
Scoring Rubrics	5	8	12
Core and Corollary Competencies	6	11	9
Rubric Wording	3	1	4

Note: Cell content represents the number of coding references (or coding occurrences) across all focus groups/interviews which fall at the intersection of the content (row) and structural (column) codes. Comments sometimes warranted the application of several different specific content codes simultaneously or only the general content code, resulting in coding occurrence counts that sum to a number other than the total number of general content coding occurrences.

Positive Reactions to Rubrics:

I think it [the rubric] hit enough areas.

When we talked about it, the points were really such that it made it easy to score...the bullet points were there. I don't think it was too much—because you do a lot, and it's nice to hear "I know you do that."

It is a pretty robust evaluation.

Negative Reactions to Rubrics:

They make it sound nice, [reading] "core standard 1: the leader has knowledge and skills to think and plan strategically creating an organizational vision around the personalized student success...[whistles]. What does that look like? How do you distinguish between what you think is good, and what I think is good?"

I think it's pretty simple, about how the scoring system is supposed to work—but I don't agree with it!

In addition to clarifying the meanings of the core and corollary competencies through more explicit wording, participants identified several other ways to improve the scoring rubrics. These suggestions and their frequencies are presented in Table 31.

Table 31. Suggestions for Improving Principal Evaluation System Scoring Rubrics

Suggestion	Frequency
More Evidence Sources	2
Electronic Tools	2
More Room for Variability	2
Alignment Between Principal and Teacher Evaluation Rubrics	1
Add item/s capturing "effective management"	1
Add item/s capturing "interpersonal skills"	1

Two individuals suggested including other sources of evidence, saying:

It might even be valuable for [my evaluator] to talk to some teachers about what they perceive...that could have all been part of it.

I think maybe a portfolio—or a log—might serve that purpose well.

Others suggested creating electronic tools (such as online rubrics) to make the system more user-friendly and improve data organization. One participant said:

Doing this void having the technology to go with it, I think is going to be very difficult...if we want this to be something that is used with fidelity we've got to tool it up.

Two participants suggested using a larger scale to provide more room for variability. For example, one individual stated:

...[Y]ou can't do it on a 1, 2, 3, and 4. There's too many—the rubric would have to be 20 wide for it to be effective, for you to see a big difference between a 1 and a 4 here.

Additionally, one participant advocated the alignment of the Principal Evaluation System and the Teacher Evaluation System:

The verbiage that is used in here should be similar to the verbiage that is used in the teacher document, and I'm hearing it's not. If we're saying to a principal, talk about "intended, enacted, taught, learned"—is that verbiage also in the [Teacher Evaluation System]? ...Is the verbiage in there about cognitive demand...about pedagogical versus content knowledge? ...If we are saying the principals are responsible for this but then they can't hold the teachers responsible, there's a disconnect.

Two participants identified areas they felt were not captured by the system: interpersonal skills and management skills:

There's not any area—maybe there shouldn't be—with personal...personality...there's stuff about working with your staff, but there's not...I don't know how to say it...there wasn't an area that [principals having interpersonal difficulties with staff] would come out. There wasn't a way to put that in here...the personality factor—personal relationship factor. But I don't know if it needs to be.

This speaks to instructional leadership, but it doesn't always capture that management piece that sometimes is the reality of a principal's life...if there's a lockdown going on in your school because there is police activity in the area, you have to deal with that...if you have say, an autistic student...who will not get off of the photographer's perch—you know what that did to the schedule of the photography for the rest of the day!? Those little things...the fire drill that has to happen, the parent who needs immediate attention...the reporter who comes into your office and wants to speak to you...all those types of management things...you know, the ability to handle that.

System Process. Table 32 summarizes principals’ and principal evaluators’ reactions to the entire Principal Evaluation System process.

Table 32. Reactions to Principal Evaluation System Process

	Negative Reaction	Positive Reaction	Need More Information
Time and Effort	4	5	0
System Fairness	4	4	1
Evidence	5	4	8
Conference Frequency	3	2	2
Evaluator Feedback	1	10	1
Self-Assessment and Self Reflection	0	2	0
Professional Development Support	2	3	0
Operational Process	1	0	0

Note: Cell content represents the number of coding references (or coding occurrences) across all focus groups/interviews which fall at the intersection of the content (row) and structural (column) codes.

Participants’ opinions were mixed regarding the amount of time and effort required by the system.

Positive Reactions to Time and Effort:

Interviewer: Do you think the Time and effort you were asked to put into the principal evaluation process was reasonable?

Principal: Yes. Yes. Um...because it would be me reflecting on each of the domains, gathering evidence on each of them, then having the meeting with [my evaluator] to review each of the artifacts...I guess I think that is time that is needed, I guess, that's a meeting that needs to occur, especially at the end of the year...

Negative Reactions to Time and Effort:

No...this is a bear...it took too much time.

Indications regarding system fairness were also mixed, with most positive comments being general (e.g., “oh yeah, I think it’s fair”) or focusing on the content of the rubrics (the core and corollary competencies) and most negative comments focused on the scoring process.

Positive Reactions to Fairness:

I think if you're not covering those areas, I think there would be reason for concern--if a principal is not able to provide evidence in one of those areas I think it would be a good indication that something's wrong.

Negative Reactions to Fairness:

Something else that bothers us—you can have done everything in 1s, 2s, and 4s, but you missed one thing in 3 and you are dropped back down to a 2—that’s not fair.

There's so many descriptors within each number, and to say that if you miss one you go back to the other one, I think is totally unfair to the person who is being rated.

The following is a summary of the different types of positive and negative comments that participants made in regards to the evidence the system prompted them to provide. Generally, participants indicated the need for more guidance regarding the types of evidence appropriate for demonstrating principals' performance in each domain.

Positive Reactions to Evidence:

I wish I'd spent more time leading up to this evaluation...if I had the whole year...and I knew what I was supposed to be providing, it wouldn't be a problem.

Negative Reactions to Evidence:

Interviewer: Do you think the system is clear enough about what you should bring with you?

Participant: Probably not, no...but again I think that goes with the conversation about 'what does it look like to be there in our district?'

Participants indicated that they valued the evaluation conferences, and some indicated that more than one meeting would be beneficial in the future. There was some indication that the conference frequency expectations had not been clear during the pilot, and that specific guidelines would also be helpful.

Positive Reactions to Conference:

You don't take time sometimes and just sit down and talk, and [because of the evaluation] we actually scheduled a meeting—which we hardly ever do, [my evaluator] and I—to just talk together.

Negative/Positive Reactions to Conference:

We should have met more. The one on one was valuable. I wouldn't want to bog someone down...I would think at least two to three times a year.

The following is a summary of the comments that participants made in regards to the evaluator feedback prompted by the system. Participants made almost exclusively positive comments regarding evaluator feedback, for example:

Positive Reactions to Feedback:

I would say it was helpful, I would say it was a good conversation that gave me the type of feedback I needed.

Areas where we both agreed I was a three, areas we both agreed I was a four, that was helpful--at least it was reassuring to me that I see the same thing! And uh, certainly a few areas where I could improve on and [my evaluator] agreed--they were good, but they could be better...

I think the conversation is certainly higher than it was before.

However, one participant commented that more attention should be given to the provision of feedback targeting professional development according to the evaluation system.

The two comments regarding self- reflection were both positive; for example, one participant said:

I'm looking forward to getting my materials together for the reflective piece, because I need a lot of that. It can't just be mental reflection, I need to write it down, journal it, keep an electronic file somewhere. So yeah, I'm looking forward to that—how can you not reflect when you're gathering your evidence? It kind of forces you to take a walk.

Participants indicated mixed opinions about the direct support of professional development prompted by the system.

Positive Reactions to Supporting Professional Development:

I think it was helpful to discuss these things...I think having this as a basis for us to really go through and have those discussions, I think it will help [the evaluate] to improve in those areas.

Negative Reactions to Supporting Professional Development:

What we need to do—any good evaluation I'm going to say “you're a 3, or you're a 2.” If you're a good educator, you're going to want to know, what do I have to do to look like a 3? Or what do I have to do to look like a 4? And I think this tool might be a little bit weak in that area...it all comes down to providing individual help.

Only one comment was made regarding the general operational process of the system. One individual indicated that in one year, a principal could not give the expected amount of attention to each area addressed by the evaluation:

Have one maybe two major focal points...then, if you look at this [pointing to guidebook], they'll have us running like nuts trying to demonstrate competencies in all of these areas...we shouldn't have to show...that we've mastered anything over and over and over again if the system is working well, because otherwise you never get to concentrate on what needs better attention.

Table 33 includes the suggestions made regarding the improvement of the Principal Evaluation System process (beyond the improvements suggested above) and the number of times the suggestions were made across focus group/interview sessions.

Table 33. Suggestions for Improving Principal Evaluation System Process

Suggestion	Frequency
Time With Colleagues to Discuss Evaluation System Components	3
Stress Positive Formative Goals	1

A few participants indicated that they would value time to work with colleagues to form a common understanding of the system components, including evidence gathering, the meaning of each competency,

the proficiency levels, etc. Also, one participant suggested emphasizing the formative goals and opportunities for professional growth and development associated with this evaluation system.

Summary. Participants generally indicated that the training did not adequately prepare them for the pilot, and that more intensive training over a longer period of time would be beneficial in the future. Most participants reported that they found the core and corollary competencies valuable, but several indicated that the scoring system could be fairer; specifically, that they did not agree with reducing a score to the lowest competency level in which all indicators were met. Participants reported that the evaluator feedback prompted by the system was valuable, as was the process of gathering evidence and self-reflecting on practice. Suggestions for the improvement of the system included: more guidance regarding evidence sources/types and time with colleagues to discuss evidence of practice, more frequent conferences with evaluators, the use of electronic tools, the addition of indicators regarding interpersonal skills and management skills, and more room for competency level variability. Finally, one participant advocated the alignment of the content of the Principal Evaluation System and the Teacher Evaluation System.

VI. Summary and Recommendations

The pilot for the Pennsylvania Teacher and Principal Evaluation System succeeded in providing important information for modifying the system. To help improve the pilot evaluation system, several methods were used to gather information about the teacher and principal evaluation pilot from participants including surveys, focus groups with teachers, and interviews with principals and principal evaluators. Across the four sites, 114 teachers responded to the online survey (71% response rate), 19 principals responded to the online principal survey (63% response rate), and 5 principal evaluators responded to their survey (83% response rate). Overall, 34 teachers participated in the focus groups, 6 principals were interviewed, and 3 principal evaluators were interviewed across three of the pilot sites.

In interpreting the results of the evaluation of the pilot study it is important to consider the time frame of the study. Participants were trained in January and February and the implementation of the study immediately followed. This resulted in less than 5 months to implement the teacher and principal evaluation system.

The results of the surveys, focus groups and interviews indicated that the training sessions could be improved to better prepare participants for the evaluation system. Participants suggested revising the training format for the teacher evaluation system so that it is tailored to teachers' grade levels and content areas. They also indicated the need to extend the training so that it is more comprehensive and focuses on all the domains covered in the rubric and on the types of evidence that teachers can provide to support their effectiveness. Teachers and principals responded favorably to many aspects of the teacher evaluation rubric. In particular, they valued the domains and criteria reflected in the rubric – Danielson's Framework- and its comprehensiveness, capturing the many aspects that are reflective of effective teaching. Participants provided constructive feedback on the rubric wording, including the need for greater specificity, clarity and usability of the rubrics, as well as the need to provide more information on the types of evidence required. Overall, participants responded favorably to the observations and conferences in the teacher evaluation process. Some participants suggested that more formal observations by different observers would be preferable, while others suggested the need for more walk-throughs. Teachers indicated that the conversations with their principals and the time for self-reflection were valuable components of the evaluation system. Overall, teachers and principals indicated that the teacher evaluation system supports good teaching.

Principal and principal evaluators indicated that the training for the principal evaluation system needs to be revised so as to better prepare them. Overall, they suggested that the training needs to be more comprehensive and focus on all aspects of the evaluation system. The principal and principal evaluators indicated that the core and corollary competencies reflected in the rubric are valuable, but some indicated that the scoring system should be revised so as not to reduce a score to the lowest competency level in which all indicators were met. Others indicated the need to reduce the redundancy within the rubric and to shorten the rubric. Participants indicated that additional information on the sources of evidence and the allocation of time to meet with colleagues to discuss evidence of practice would enhance the evaluation system. Principals valued the time to conference with their supervisor and suggested the need for additional conferences, at least a midyear conference and evaluation. Overall, principals and principal evaluators indicated that the principal evaluation system supports good leadership.

The results from the surveys, focus groups and interviews provide valuable information for the revision of the Pennsylvania Teacher and Principal Evaluation System. The following recommendations are provided based on these results. An overall recommendation is that the System adhere to the *Joint Committee Standards for Educational Evaluation* (JCSEE; 1988). The specific recommendations are categorized into three areas: Training, Rubric and Evaluation Process.

Recommendations for the Teacher Evaluation System

Training

- More extensive training with follow-ups during the year
- Smaller training groups formed by grade span and content area
- Align video clips of teacher practice to grade span and content area
- Spend more time on the distinction between the rubric levels (especially between Proficient and Distinguished)
- More information on the nature of evidence for each Domain Component
- More specific information about the evaluation process and steps required for the evaluation
- Train the supervisors prior to the teachers so they can be a resource

Teacher Evaluation Rubric

- Reduce the verbiage in the criteria and redundancy across the criteria
- Clarify the distinction between Proficient and Distinguished
- Create a user-friendly online version
- Provide clear examples of the type of evidence for the Domain Components at different levels
- Ensure consistency with Principal Evaluation Rubric
- Establish the reliability of the rubric scores
- Establish the validity of the score inferences, in particular, the validity of the domain score inferences

Evaluation Process – Observations/walkthroughs/conferencing

- Start the evaluation process at the beginning of the academic year with a goal setting conference between teacher and supervisor
- Provide documentation on the evaluation process and steps required
- Maintain multiple formal observations with conferences
- Specify more than one walk-through and more condensed rubrics for their use (and how the results from the walk-throughs would be combined with the results from the formal observations)
- Use multiple observers (may allow for more observations)
- Create time for teachers to meet and discuss the rubric, criteria, evidence, etc. throughout the year

Recommendations for the Principal/Supervisor Evaluation System

Training

- More extensive training with follow-up
- More information on the nature of evidence for each Competency
- More specific information about the evaluation process and steps required for the evaluation

Principal Evaluation Rubric

- Reduce the verbiage in the Competencies, and redundancy across the Competencies and Summary of Evidence
- Change the rating process so that a principal who receives a higher rating for all but one descriptor does not receive the lower rating
- Create a friendly online version
- Provide clear examples of the type of evidence for the Competencies at different levels
- Ensure consistency with Teacher Evaluation Rubric

Principle 3 – Appendix C

- Establish the reliability of the rubric scores
- Establish the validity of the score inferences, in particular the validity of the domain score inferences

Evaluation Process

- Start the evaluation process at the beginning of the year with goal setting activities
- Consider a mid-year conference between the principal and evaluator
- Provide documentation on the evaluation process and steps required
- Create time for principals to meet and discuss the rubric, criteria, evidence, etc.

References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Commonwealth of Pennsylvania, (2010). Technical Proposal, Pennsylvania: Ready to Go! Reaching Beyond! Harrisburg, PA.
- Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education*, 51(3), 166.
- Elmore, R. F. (2002). Hard questions about practice. *Educational Leadership*, 59(8), 22.
- Heneman, H.G., III, Milanowski, A., Kimball, S.M., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay*. Philadelphia, PA: Center for Policy Research in Education.
- Howard, B.B. & Gullickson, A.R. (2010). Setting standards for teacher evaluation. In M. Kennedy (ed.), *Teacher Assessment and the Quest for Teacher Quality* (pp. 337-353). San Francisco, CA: Jossey-Bass.
- Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Joint committee on Standards for Educational Evaluation (1988; 2009). *The Personnel Evaluation Standards*. Newbury Park, CA: Sage.
- Joint Committee on Standards for Educational Evaluation (n.d.). *Personnel Evaluation Standards*. Retrieved August 27, 2010, from <http://www.jcsee.org/personnel-evaluation-standards>
- Konold, T., Jablonski, B., Nottingham, A., Kessler, L., Byrd, S., Imig, S., Berry, R., & McNergney, R. (2008). Adding value to public schools: Investigating teacher education, teaching and pupil learning. *Journal of Teacher Education*, 59(4), 300-312.
- Lane, S. & Horner, C. (September 2010). *Teacher and principal standards-based evaluation systems*. Pittsburgh, PA: University of Pittsburgh.
- Newmann, F.M., Bryk, A.S., & Nagaoka, J.K. (2001). Authentic intellectual work and standardized tests; Conflict or coexistence? Chicago, IL: Consortium on Chicago School Research. Retrieved September 1, 2010 from, http://ccsr.uchicago.edu/content/publications.php?pub_id=38.
- U.S. Department of Education (2010). Retrieved August 28, 2010 from <http://www2.ed.gov/programs/racetothetop/index.html>.

Appendix A

Observer Comments on the Teacher Evaluation Training Sessions

Comments on the time spent on each topic include:

There was little time spent on discussing the implementation including timeline, observations, meetings, etc.

Time was managed well in regards to the planned activities. Groups were given ample time to finish assigned tasks...presenters were consistent about asking if there were any questions... presenters focused on several components, which seemed to work well given the limited time for training.. more time could have been spent on defining the components they did cover... some confusion regarding, for example, what kinds of evidence would be appropriate for “Setting Instructional Outcomes”

The material was covered in its entirety. The PowerPoint and the Trainer presentation were well planned. The trainers covered all of the material and provided time for an exercise that allowed interaction and for participants to give thought to the types of evidence that could be used for each of the priority components.

There was insufficient time to ensure that everyone was prepared to participate in the new system ... particularly concerned that we were unable to spend more time on the process and the elements of the pilot – the informal observation, formal observation, conferences, etc ... no opportunity to test inter-rater reliability on the instrument either. It was a good awareness session with little opportunity to build and assess knowledge and skills due to the time constraints.

Comments on whether the printed material provided supported the training:

The PA Teacher Evaluation Rubric was provided to the participants. ...They often referred to specific pages in the document and used complimentary slides to delve more deeply into certain aspects of the system

However, the power point slides were not distributed. At a minimum the implementation chart could have been provided.

Some of the material covered in the teacher training was in the PowerPoint but not in the rubric document. ... The process components, for example, are very important and were not given in print to the participants.

Comments on whether the printed material provided to the participants was clear:

The PA Evaluation Rubric was clear.

The addition of a table of contents, however, would be helpful

I did not observe any questions or concerns raised about clarity of material

Comments on whether the rubric and procedures were presented clearly:

Principle 3 – Appendix C

This was the focus of the training and the rubric was presented clearly and the activities provided to the teachers allowed them to work with the rubric

The presenters organized the activities and lessons in such a way as to maximize understanding and explained things in just enough depth (with time constraints in mind)

Again, the weakness was in the description of the process surrounding the rubric and the lack of time for training and testing ability to use the rubric reliably across schools and supervisors.

Comments on whether the participants were actively engaged:

Throughout the training, presenters engaged the participants through questioning and the assignment of group work

...allowed them to work with the rubric and the types of evidence to support it

Comments on whether participant questions were addressed:

Questions regarding the implementation were not well addressed.

Questions regarding the rubric were addressed.

...however, the discussion after the training indicates that not all participants understood and were comfortable

Comments on the important issues identified by the participants:

*Superintendents need the steps and materials to implement the evaluation system
...superintendent wanted to know if additional support and rubrics will be provided.... the principal wasn't clear about when to use the rubric*

...participants were asked to write questions/concerns on note cards to be discussed at a later meeting... overheard a teacher expressing concern regarding the subjectivity involved in deciding what constituted "enough" evidence

How are we going to ensure inter-rater reliability?... one person commented that the training was too basic... one person commented they needed more time to learn how to apply the evaluation system

Will this evaluation model eventually be the same for regular and special educators... concern with timeline for the project... evaluation being tied to student performances on a test is an issue...

Observer Comments on the Principal Evaluation Training Sessions

Comments on the time spent on each topic include:

Ample time was allotted for all planned activities and discussions

Most of the time was spent on group activities – which I think was appropriate... perhaps a little more time could have been spent discussing the four different performance levels. This was covered very quickly

The flow of the conversation was good and the important components were adequately addressed

The PowerPoint was covered and the participants were able to participate in a rating exercise to address how they each rated on a component and discuss why they did so. This was a better exchange than the teacher exercise in that it addressed differences and meanings. Again, the depth of this exercise was limited by the availability of time and the difficulties caused by the problems experienced. The trainers were able to bring their materials to the Cornell participants the following day, but they did not have the rubric for the training.

There was no direction given on the use of the Principal instrument or process steps. Should there be a pre-conference? What is the role of the Principal? Is there a goal setting step? The participants were told they had flexibility in the use of the rubric. I have a concern that the results of the pilot will be based on different processes in each pilot site, making it difficult to know what worked and what didn't work.

Comments on whether the printed material provided supported the training:

The materials were extensive but were not available to the group who were teleconferencing....again, the available time did not permit for effective use of the extensive material distributed during training

The trainers frequently referred to the printed material and checked for understanding

Comments on whether the printed material provided to the participants was clear:

The materials provided clear direction, but they will need more time to digest everything

Comments on whether the rubric and procedures were presented clearly:

The presentation was well organized, and explanations regarding levels of competency and standards were excellent. However, there was substantial confusion regarding details of the procedures (who writes the narratives regarding the evidence, is there only one conference or two, etc. These issues were not resolved.

We did not use the "Pennsylvania School Leadership Standards Evaluation Guidance" so I can't comment on its value to participants

Comments on whether the participants were actively engaged:

Principle 3 – Appendix C

The participants spent more than two hours engaged in the activities.... With both small and large group discussion. The facilitators did a good job of pushing the participants to be concrete in their description of evidence and to provide a good rationale for their ratings in the scenarios

Comments on whether participant questions were addressed:

The trainers were unable to provide the specific clarification that participants requested regarding the procedures. Other questions were addressed.

Comments on the important issues identified by the participants:

Some of the items rated may be outside the realm of what a particular principal has the authority to control... There was a concern that no “process” has been identified yet

Principals indicated that in many districts, it may be hard for principals and superintendents to have “courageous conversations” and overcome defensiveness

Presenters indicated that there may be problems with inter-rater reliability. This particular group had very high agreement practice activity; however, they commented on how much better this group had done than previous group. The previous groups seemed to be extremely discrepant in ratings - the vignettes almost followed the guide and rubric word for word.

Tools provide a framework for how principals should conduct their day-to-day business...provides transparency – clear directions of what is expected...process should result in highly qualified leaders in all schools... need more scenarios in the training that would require participants to discern between level 2 and 3... perceived benefits include clear expectations of leaders, changes how we supervise leaders, more of a collaborative model of evaluation, gets to real evidence,

Better understanding of the evidence required to support the evaluation...rubrics will be used to guide behavior and reflection.... Alignment with the standards is important

**Value-Added Estimates for
Phase 1 of the Pennsylvania
Teacher and Principal
Evaluation Pilot**

Final Report

January 12, 2012

Stephen Lipscomb
Hanley Chiang
Brian Gill



MATHEMATICA
Policy Research

All statistics are calculated by Mathematica unless stated otherwise

Mathematica Reference Number:
06815.300

Submitted to:
Team Pennsylvania Foundation
100 Pine Street, 9th Floor
Harrisburg, PA 17101
Project Officer: Jennifer Cleghorn

Office of the Deputy Secretary of Elementary and
Secondary Education
Pennsylvania Department of Education
333 Market Street
Harrisburg, PA 17126-0333
Project Officer: Carolyn C. Dumaresq

Submitted by:
Mathematica Policy Research
955 Massachusetts Avenue
Suite 801
Cambridge, MA 02139
Telephone: (617) 491-7900
Facsimile: (617) 491-8044
Project Director: Stephen Lipscomb

**Value-Added Estimates for
Phase 1 of the Pennsylvania
Teacher and Principal
Evaluation Pilot**

Final Report

January 12, 2012

Stephen Lipscomb
Hanley Chiang
Brian Gill



MATHEMATICA
Policy Research

All statistics are calculated by Mathematica unless stated otherwise

ACKNOWLEDGMENTS

This report would not be possible without the support of the Team Pennsylvania Foundation, the Pennsylvania Department of Education, the Bill & Melinda Gates Foundation, and members of the Stakeholder Steering Committee who volunteered their time during Phase 1 of this pilot study. We are especially grateful to Carolyn Dumaresq, Jennifer Cleghorn, Matt Zieger, Pat Hardy, Rich Maraschiello, Sharon Brumbaugh, and Theresa Barnaby for their project leadership, to Harris Zwerling for his insightful comments and suggestions, and to Dave Ream, Thomas Gillin, and Thresa Stafford for providing us with the statewide data used in this report. We also thank Bob Hughes, Suzanne Lane, and the teachers, principals, and superintendents from the four pilot school districts for their important contributions during the pilot toward the improvement of the evaluation system.

We would like to recognize several staff at Mathematica Policy Research who made substantive contributions to this report. Clare Wolfendale provided outstanding research support in estimating the value-added models and populating tables of results. We are also grateful to Serge Lukashanets, Xiaofan Sun, and Alena Davidoff-Gore for their contributions to organizing the data, to Duncan Chaplin for his comments and suggestions on an earlier version, to John Kennedy for his editorial assistance, and to Autumn Parker and Eileen Curley for their production support.

Any opinions expressed herein are those of the authors and do not necessarily represent the views of the Team Pennsylvania Foundation or the Pennsylvania Department of Education.

THIS PAGE LEFT BLANK FOR DOUBLE-SIDED PRINTING

CONTENTS

ACKNOWLEDGMENTS	iii
EXECUTIVE SUMMARY	xiii
I. INTRODUCTION	1
A. The Pennsylvania Teacher and Principal Evaluation Pilot	1
B. Description of Value-Added Models	2
II. CHARACTERISTICS OF VAMS ESTIMATED IN THIS REPORT AND THE APPLICABILITY OF EFFECTIVENESS MEASURES TO TEACHERS IN PHASE 1	7
A. Outcome Measures	7
B. Teachers with VAM Estimates from Phase 1 of the Pilot	10
C. Control Variables that Are Included in the VAMs	12
III. VALUE-ADDED RESULTS FOR TEACHERS.....	17
A. Variation in Teacher Effectiveness Based on PSSA Outcomes	17
B. Statistical Uncertainty in the Teacher Effectiveness Estimates	20
C. Sensitivity of Measured Effectiveness to Alternative VAM Specifications	24
D. Key Characteristics of Teacher Effectiveness Estimates Based on Pilot District Samples and Other Outcomes	29
IV. RELATIONSHIPS BETWEEN TEACHER PRACTICES AND VALUE ADDED	33
A. The Phase 1 Teacher Observation Rubric and Score Distribution	33
B. Observation Scores and Value-Added Scores for Phase 1 Teachers with VAM Estimates	36
C. Relationships Between Value-Added and Observation Scores	38
V. VALUE-ADDED RESULTS FOR PRINCIPALS	41
A. An Approach to Estimating Pure Principal Effects	41
B. The School VAM as the Basis for Evaluating Principals.....	44
C. Key Characteristics of School Effectiveness Estimates Based on PSSA Outcomes	46

V (continued)

D.	Key Characteristics of School Effectiveness Estimates Based on Outcomes Other than PSSA Scores	50
VI.	LOOKING AHEAD TO SUBSEQUENT PHASES OF THE PENNSYLVANIA TEACHER AND PRINCIPAL EVALUATION PILOT	57
	REFERENCES.....	59
	APPENDIX A: TECHNICAL SPECIFICATIONS OF THE VAMS	
	APPENDIX B: DATA SOURCES AND SAMPLE CHARACTERISTICS	
	APPENDIX C: TECHNICAL RESULTS FROM VALUE-ADDED ANALYSES	

TABLES

1	Number of Teachers with Effectiveness Estimates Reported Based on the Number of Cohorts in the VAM and Share of Reported Estimates that Are Statistically Different from the Average	xvi
I.1	Strengths and Limitations of Value-Added Models Relative to Other Evaluation Methods	4
II.1	Outcomes Considered in Value-Added Models for Teacher and School Evaluation in this Report.....	8
II.2	Summary of Teacher Evaluation Pilot, Phase 1	11
II.3	Number of Teachers from Phase 1 with at Least One VAM Estimate that Can Be Used for Correlating Value-Added with Teacher Practices in Chapter IV.....	12
II.4	Baseline Measures for Value-Added Models Estimated in this Report, by Outcome.....	13
II.5	Student and Classroom Control Variables Included in VAMs Estimated for this Report	15
III.1	Distribution of Teacher VAM Estimates for Selected PSSA Outcomes	17
III.2	Teacher VAM Estimates in Recent Studies for the 85th Percentile of Effectiveness Relative to the 50th Percentile, Reported in Standard Deviations of Student Test Scores	19
III.3	Number of Teachers with Effectiveness Estimates Reported and Share of Reported Estimates that Are Statistically Different from the Average, by Number of Cohorts Used in Estimation	23
III.4	Implied Percentage of Variation in Teacher Value-Added Within Districts and Schools	25
III.5	Counts and Percentages of Grade 5 Math and Reading Teachers in Effectiveness Quartiles Based on 3-Cohort Teacher VAMs that Include and Exclude Same-Subject Baseline Scores.....	27
III.6	Grade 4 Through 8 Math and Reading Teachers in Pilot Districts in Effectiveness Quartiles Based on Teacher VAMs with Fall or Spring Baselines Using PSSA Outcome Data	29
III.7	Key Characteristics of Teacher Effectiveness Estimates Based on Selected Non-PSSA Tests Administered in the Pilot Districts	30
III.8	Key Characteristics of Teacher Effectiveness Estimates Based on PSSAs in 3rd and 11th Grades in the Pilot Districts.....	32

IV.1	Danielson Framework Domains and Components, by Priority and Additional Components for the Pennsylvania Pilot.....	33
IV.2	Final Ratings in Pennsylvania and Chicago, by Number and Percentage of Teachers.....	35
IV.3	Sample Characteristics of Nonpilot and Pilot Teachers	36
IV.4	Regression Coefficients Indicating the Standard Deviation Increase in Teacher Value-Added that Is Predicted for a One-Unit Increase in Rubric Scores.....	39
V.1	Counts and Percentages of Principals in Effectiveness Quartiles Based on Principal Transitions Model and School VAM	46
V.2	Distribution of School Effectiveness Estimates for Selected PSSA Outcomes	47
V.3	Number of Schools with Effectiveness Estimates Reported and Share of Reported Estimates that Are Statistically Different from the Average, by Number of Cohorts Used in Estimation	50
V.4	Key Characteristics of School Effectiveness Estimates Based on Selected Tests Administered in the Pilot Districts	52
V.5	Key Characteristics of School Effectiveness Estimates Based on Nonassessment Outcomes.....	54
B.1	Data Sources	B-2
B.2	Descriptive Statistics on Student Characteristics, 2010-2011	B-3
B.3	Baseline and Analysis Student Sample Sizes for Teacher and School VAMs, by Outcome	B-4
B.4	Number of Teachers and Principals with VAM Estimates Reported from Multicohort and Single-Cohort VAMs	B-6
C.1	Sample Characteristics of Outcome Measures and Teacher VAMs Based on State Samples	C-1
C.2	Estimated Regression Coefficients from Selected Three-Cohort PSSA Teacher VAMs.....	C-2
C.3	Effect Sizes for Three-Cohort Teacher VAMs Expressed in Terms of One Year of Learning	C-3
C.4	Sample Characteristics of Outcome Measures and Teacher VAMs Based on Phase 1 Samples	C-4
C.5	Sample Characteristics of Outcome Measures and School VAMs Based on State Samples	C-5

C.6 Sample Characteristics of Outcome Measures and School VAMs
Based on Phase 1 Samples C-6

THIS PAGE LEFT BLANK FOR DOUBLE-SIDED PRINTING

FIGURES

1 Distribution of Teacher Effectiveness for 5th-Grade Math PSSA Scores xv

2 Distribution of Final Rating Scores for Phase 1 Teachers xvii

3 Distribution of VAM Scores for Phase 1 Teachers with VAM Estimates xviii

III.1 Distribution of Teacher Effectiveness and 95 Percent Confidence Intervals of Teacher Effectiveness Estimates for 5th-Grade Math PSSA Scores 21

III.2 Distribution of Teacher Effectiveness Estimates and 95 Percent Confidence Intervals of Teacher Effectiveness Estimates for 8th-Grade Reading PSSA Scores 22

IV.1 Distribution of Final Rating Scores for Phase 1 Teachers 34

IV.2 Distribution of Average Rating Scores for Phase 1 Teachers with VAM Estimates 37

IV.3 Distribution of VAM Scores for Phase 1 Teachers with VAM Estimates 38

V.1 Distribution of School Effectiveness Estimates and 95 Percent Confidence Intervals of School Effectiveness Estimates for Math PSSA Grade 5 Scores 49

V.2 Distribution of School Effectiveness Estimates for 9th-Grade Holding Power 55

THIS PAGE LEFT BLANK FOR DOUBLE-SIDED PRINTING

EXECUTIVE SUMMARY

The Commonwealth of Pennsylvania plans to develop a new statewide evaluation system for teachers and principals in its public schools by school year 2013–2014. To inform the development of this evaluation system, the Team Pennsylvania Foundation (Team PA) undertook the first phase of the Pennsylvania Teacher and Principal Evaluation Pilot—henceforth referred to as Phase 1—in 2010 and 2011 in collaboration with a broad stakeholder group that included leaders from the Pennsylvania Department of Education (PDE), the Pennsylvania State Education Association (PSEA), school districts, and the business community. The purpose of Phase 1 was to develop and implement a pilot set of performance measures to obtain lessons for improving the use of classroom observations and student data in evaluating teacher and principal performance. None of the results from Phase 1 had a bearing on actual evaluations or personnel decisions for any teacher or principal.

Phase 1 proceeded along two tracks. In the first track, observation-based rubrics for evaluating teacher and principal effectiveness were implemented on a trial basis in the Allentown, Cornell, and Mohawk Area school districts, and in Northwest Tri-County Intermediate Unit 5 (collectively referred to as Phase 1 pilot districts). Based on these rubrics, a set of preselected principals and teachers from the pilot districts were rated by their supervising superintendents and principals, respectively, in spring 2011. Lane and Horner (2011) discuss the results of this track.

This report presents findings for the second track of Phase 1. In this track, Mathematica Policy Research used student-level data to develop value-added models (VAMs) for estimating teacher and principal effectiveness. VAMs estimate the effects of educators on student achievement growth. VAMs belong to the class of models that are generally referred to as student growth models, but a VAM estimate is not a measure of student growth; rather, it is an estimate of an educator’s or a school’s *contribution* to student growth. VAMs can be appropriate for use in teacher or principal evaluations because they produce information about educator effectiveness. Other indicators like student proficiency rates and descriptive measures of student growth might be appropriate as targets for school accountability purposes, but they should not be viewed as indicating what a teacher or school has contributed to student learning.

After calculating these effectiveness estimates, Mathematica then examined whether Phase 1 teachers with higher classroom observation scores on specific professional practices covered by the pilot rubric had greater impacts on student achievement as measured by value-added.

Specifically, we address the following three primary research questions in this report:

1. How can VAMs be used to characterize the effectiveness of teachers at raising achievement according to multiple outcome measures?
2. Do specific teacher practices relate to larger contributions to student learning among Phase 1 teachers?
3. How can principals’ contributions to student learning be measured?

The U.S. Department of Education’s Race to the Top initiative is a prominent example of the interest among federal, state, and local policymakers in measuring educator effectiveness based on performance, and VAMs have been a focal point in these debates. In a VAM, the actual level of achievement demonstrated by an educator’s students is compared to the level that would be

predicted after accounting for students' own prior achievement histories and factors such as the characteristics of their family backgrounds and peers. The differential amount (above or below zero) is averaged across students taught by each educator and attributed to educators as their contribution to achievement. VAMs measure relative teacher performance based on the assessments that are used in the models. In principle, they can be applied to any quantifiable measure of student outcomes. As a measure of educator quality, a VAM's fairness depends on whether the method successfully removes influences outside an educator's control. VAMs do not indicate what level of value-added Pennsylvania should view as adequate in terms of an external standard for specifying whether students are learning "enough." VAMs also do not indicate whether the assessments on which they are based capture the skills that students ought to be learning in the classroom.

We find that VAMs based on multiple outcome measures can be informative tools for identifying highly effective and highly ineffective teachers and schools. However, larger samples of teachers than were available in Phase 1 are needed to ascertain the relationships between instructional practices and teachers' impacts on student outcomes. VAMs also face limitations in their ability to distinguish educators' true effects—especially the effects of principals—from factors beyond their control, and it is important to take these limitations into account when applying VAMs to a real, large-scale evaluation system. Subsequent phases of the pilot will require additional work to further explore and address these limitations.

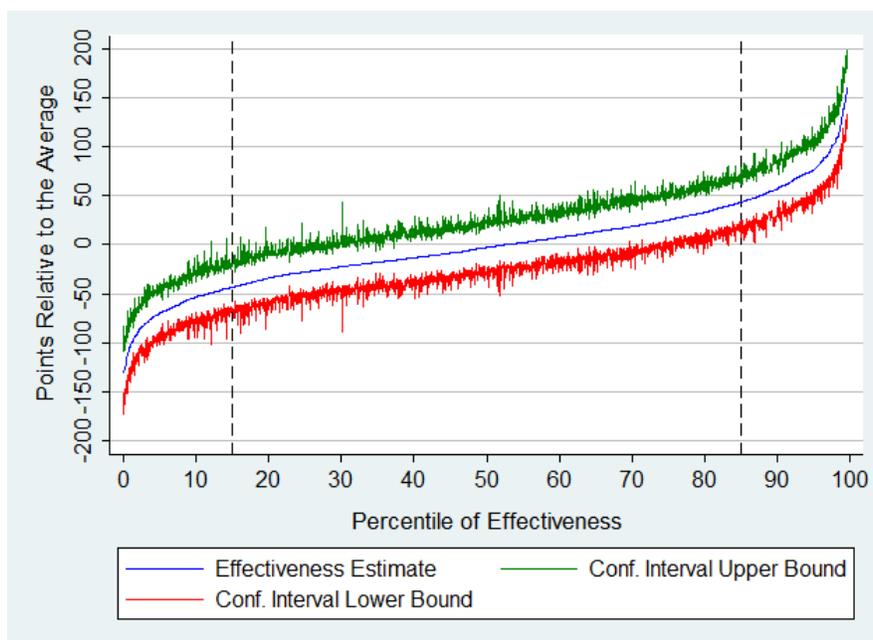
The following sections describe the main findings from the analyses and how these findings should inform the next phase of the pilot.

Using VAMs to Estimate Teacher Effectiveness

Teacher contributions to student achievement vary substantially across Pennsylvania.

The size of teachers' effects on students' Pennsylvania System of School Assessment (PSSA) scores varies substantially across the state in all PSSA subjects in grades 4 through 8. In Figure 1, we provide an example of a statewide distribution of teacher effectiveness by depicting it for 5th-grade math teachers. The blue curve indicates the value-added of individual teachers, who are rank ordered along the horizontal axis based on the estimated size of their contribution to 5th-grade math PSSA achievement. Value-added is expressed along the vertical axis in terms of additional PSSA scale points relative to the teacher in the middle of the distribution.¹ For instance, the teacher at the 85th percentile of effectiveness enables 5th-grade students to score an average of 44 PSSA points higher in math than they would have scored by having the 50th percentile teacher and 87 points higher than they would have scored by having the 15th percentile teacher.

¹ Value-added is calculated in terms of z-scores (see Appendix C). We convert z-score units to PSSA scaled scores for illustrative purposes in reporting results.

Figure 1. Distribution of Teacher Effectiveness for 5th-Grade Math PSSA Scores

Source: Mathematica calculations based on Pennsylvania data. The sample includes 2,836 teachers who taught 5th-grade students in each year between 2008-2009 and 2010-2011.

Note: See Figure III.1 for more information. Dashed lines demarcate the 15th and 85th percentiles.

Value-added data has an advantage over most other types of effectiveness information because it can indicate whether the effectiveness of two educators is statistically different. That is, a VAM can indicate with a high degree of confidence whether the actual effectiveness of teachers with low or high VAM estimates is likely to differ from the effectiveness of a teacher in the middle of the distribution. This is the purpose for the intervals around the blue curve in Figure 1, which are called confidence intervals. Statistically speaking, teachers with confidence intervals that are entirely above or below zero are said to be performing differently from (that is, either above or below) average. Such intervals are characteristic of nearly all Pennsylvania teachers below the 15th percentile and above the 85th percentile. In short, VAMs have the ability to delineate groups of teachers that differ in their performance estimates to an extent that could not have arisen by chance errors in estimation. Other types of evaluation data like classroom observation data can place teachers into performance categories but cannot indicate whether the performance of teachers across those categories is statistically different unless a confidence interval is reported.

Incorporating multiple student cohorts improves the reliability of effectiveness estimates.

A key design element for a VAM is the number of student cohorts—the full roster of students taught by a teacher in each single year—whose outcomes will factor into a teacher’s effectiveness estimate. Outcomes for multiple student cohorts carry potential information on a teacher’s contribution. Incorporating students from multiple cohorts in a VAM thus facilitates measuring a teacher’s effectiveness with greater statistical reliability. As shown in Table 1, a greater share of the effectiveness estimates can be statistically distinguished from average effectiveness in teacher VAMs that use three cohorts than in those that use one cohort. Greater reliability is a highly desirable feature for teacher evaluation measures, but the decision to incorporate data from multiple student

cohorts comes with tradeoffs. First, with more cohorts, a teacher’s effectiveness estimate will be less reflective of the teacher’s most recent performance. Second, fewer teachers will have estimates reported that are based on the full number of cohorts used in the VAM, although estimates can be calculated for all teachers based on the number of cohorts available to each.

Table 1. Number of Teachers with Effectiveness Estimates Reported Based on the Number of Cohorts in the VAM and Share of Reported Estimates that Are Statistically Different from the Average

Outcome	Number of Teachers with Estimates Reported		Percentage of Reported Estimates that Are Statistically Different from Average	
	1-Cohort Model	3-Cohort Model	1-Cohort Model	3-Cohort Model
Math PSSA, Grade 5	4,103	2,836	36.5	52.0
Reading PSSA, Grade 8	1,916	1,717	22.3	30.5
Science PSSA, Grade 4	4,187	2,854	27.7	49.8

Source: Mathematica calculations based on Pennsylvania data.

Note: See Table III.3 for more information.

There is more variation in teacher effectiveness within schools than across schools.

About 62 percent of the variation in estimated teacher effectiveness in Pennsylvania is observed within individual schools. This implies that across the state there are plenty of effective teachers in bad schools and ineffective teachers in good schools. This finding supports the conclusion that the most important factors to include in a VAM for isolating a teacher’s contribution are those that vary within schools.

The remaining 38 percent of the variation is explained by differences in schoolwide average value-added, and this part of the variation poses an analytic dilemma. Average value-added varies from school to school, but is this variation simply the result of the sorting of good and bad teachers, or are the schools affecting their teachers’ value-added? The data do not allow us to determine whether the 38 percent of teacher value-added is attributable to the teachers themselves (that is, because good teachers tend to land in the same schools with other good teachers) or to factors at the school that are outside the teachers’ control. If all of the 38 percent is related to schoolwide factors rather than to teachers, then the VAM should include a control for each individual school—thereby making teachers responsible only for the difference between their own value-added and the average value-added in their schools. This would involve the implicit assumption that average teacher quality is essentially equal in every school across the state, which seems implausible. It could also produce conflicting incentives for teachers. Good teachers in good schools could improve their value-added by moving to low-performing schools. However, absent any movement across schools, teachers could improve their value-added only by performing better than their colleagues down the hall.

Another approach would be to control explicitly for observable school characteristics in the VAM, but there are analytic challenges in determining how to ensure that these adjustments do not absorb true differences in teacher effectiveness across schools. Exploring potential ways of adjusting for school characteristics deserves further attention in Phase 2. For now, the teacher VAMs we use do not make any school-level adjustments, meaning that teachers are compared with all other teachers (of the same grade and subject) throughout the state.

VAMs based on non-PSSA outcomes have varying degrees of statistical reliability.

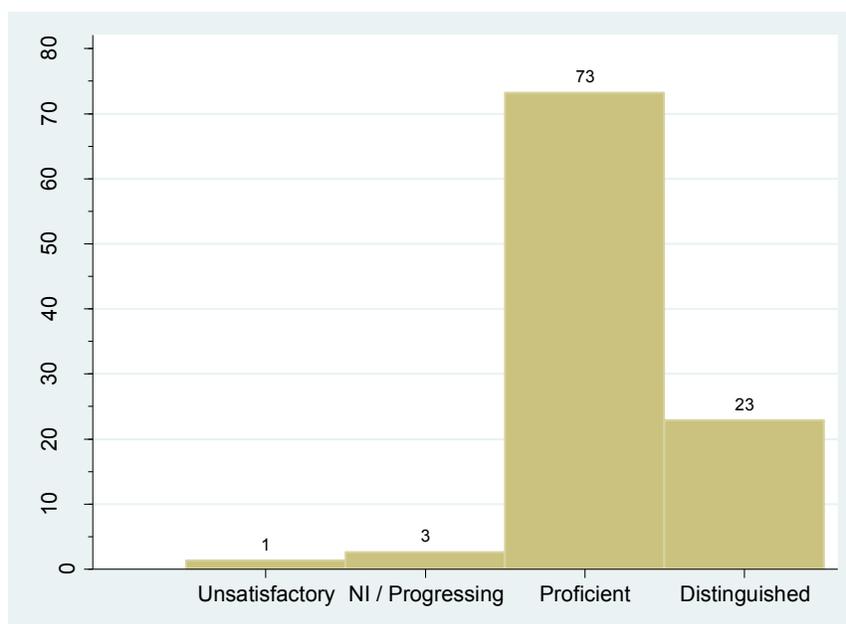
We estimated VAMs based on several non-PSSA outcomes and found notable differences in the ability of the VAMs to make reliable distinctions among teachers. For example, whereas 38 percent of teacher effectiveness estimates can be statistically distinguished from the average on the basis of a 1st-grade curriculum-based writing assessment in Allentown, only 18 percent can be distinguished from the average based on a 2nd-grade measure of early literacy skills. VAMs with greater reliability are likely to be better predictors of teacher abilities in the future as measured by value-added. Therefore, the differences in reliability could be factors in determining what weights PDE would like to place on different types of effectiveness estimates in the evaluation system.

Teacher Value-Added and the Pilot Observation Rubric

Principals rated nearly all Phase 1 teachers as proficient or distinguished.

In 2011, PDE found that, under the existing evaluation system, principals rated more than 99 percent of teachers across the state as satisfactory. Identifying the bottom 1 percent could be very useful for tenure or other personnel decisions, but the lack of variation in the other 99 percent was a cause for concern. During Phase 1, principals implemented a pilot rubric for teacher observations based on the Framework for Teaching by Charlotte Danielson that included three categories above an unsatisfactory rating. The pilot implementation produced nearly the same result in terms of the percentage of teachers at the low end of the evaluation scale. Specifically, 1 percent of all Phase 1 teachers were rated as unsatisfactory, 3 percent were rated as needing improvement—called progressing for new teachers—and 96 percent were rated as proficient or distinguished (Figure 2).

Figure 2. Distribution of Final Rating Scores for Phase 1 Teachers



Note: See Figure IV.1 for more information.

NI = Needs Improvement.

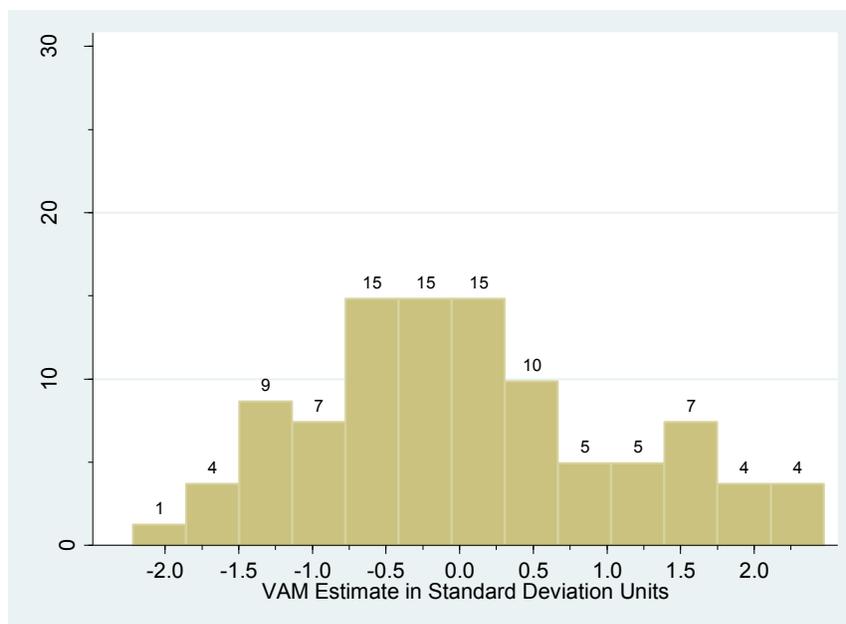
The distribution of observation scores includes a far greater proportion of teachers at the higher performance levels than would be expected based on a normal bell curve. With a small sample of

teachers in Phase 1, it is possible that the scores of these teachers are not representative of scores that would be obtained by teachers across Pennsylvania in a larger pilot. However, we do not see any compelling evidence to support this possibility based on comparisons of the characteristics of pilot teachers and other educators in Pennsylvania.

Value-added can provide information on a larger number of performance categories than most classroom observation measures typically provide.

The pilot observation rubric distinguished teacher performance in only a slightly more detailed way than the state’s existing form—which rates teachers as unsatisfactory or satisfactory—particularly at the low end of performance. In general, VAMs can provide for a large number of categories, which could be useful for policy purposes. Unlike most classroom observation rubrics, VAMs also provide information on the amount of statistical uncertainty attached to estimates in particular categories. In Figure 3, we illustrate the distribution of value-added scores among the 81 Phase 1 teachers for whom a value-added estimate could be assigned. The VAM distribution looks more like a traditional bell curve that is centered on the average possible value and in which teachers are more differentiated in terms of their performance.

Figure 3. Distribution of VAM Scores for Phase 1 Teachers with VAM Estimates



Source: Mathematica calculations based on Pennsylvania data.

Note: See Figure IV.3 for more information.

There are no statistically significant relationships between teachers’ observation scores and their value-added scores in the Phase 1 data.

Using statistical models, we tested the relationships between teachers’ estimated contributions to student learning and their observation scores for the 81 teachers with both types of effectiveness measures. The models compared the VAM score for individual teachers with their rubric ratings on each component and overall across components. The analyses sought to measure the predicted increase in teacher contributions to student learning from a one-level increase (for example, from proficient to distinguished) on any component of the observation rubric. Due to the small size of

the pilot and the compressed distribution of observation scores, none of the relationships we estimated are statistically significant. This could change in Phase 2 when a much larger number of teachers will be involved; the research literature includes several studies that indicate that teachers who have higher scores on observational rubrics make larger contributions to student achievement than teachers with lower scores. But if principals are unwilling or unable to differentiate among teachers in their observations, and if 96 percent of teachers again have ratings in the top two categories, we might again find no statistical relationship to value-added estimates. The value of a four-category rubric for professional practice depends on the willingness of the raters to use all of the categories.

Using VAMs to Estimate Principal Effectiveness

The best available method for distinguishing principals' effects on student outcomes from the effects of other school-specific factors can be applied only to a limited number of principals and therefore is not applicable to a real evaluation system.

A key analytic challenge of any principal VAM is to disentangle principals' true contributions to student outcomes from the influence of other school-level factors. A natural starting point for estimating principal effectiveness is to estimate the effectiveness of the principal's school. The complication is that a school's effectiveness can also reflect other school-specific characteristics and circumstances beyond the principal's control, most notably including the preexisting abilities of the school's teachers. Teachers have direct instructional contact with students, but principals can influence student achievement only indirectly.

The best available VAM for isolating pure principal effects, which we call the principal transitions model, calculates how the same school's value-added differs under the leadership of different principals. Thus, it measures how effective a principal is relative to the other principals who have served at the same school. This approach has the benefit of controlling for all school-specific factors beyond principals' control that remain constant over time.

From the statewide data, we identified two major reasons why this method cannot be applied to real-world evaluations of principals. First, it can generate effectiveness estimates for only a limited group of principals—those principals from schools that have undergone leadership transitions. In the statewide data, only a minority of schools underwent leadership transitions over a three-year period. Second, the principal transitions model also limits the ways in which principals can be compared on their performance. Comparisons can be made only within small networks of schools connected by a series of principal transfers. We found that most such networks encompassed only one or two schools, implying that this model measures a principal's effectiveness relative to a very limited comparison group.

VAMs for measuring school effectiveness provide informative but imperfect measures of principals' contributions to student learning.

An alternative model, which is applicable to real evaluations, gives each principal a value-added score based on the average effectiveness of the principal's school(s) during the analysis period. Although this model generates estimates for principals even if they have served in multiple schools, we call it a school VAM to emphasize the fact that it bundles together principals' true contributions with the effects of other school-level factors.

We assessed the degree to which effectiveness estimates from the school VAM deviate from pure principal effects. Estimates from the principal transitions model served as benchmarks with which estimates from the school VAM (for the same principals) were compared. We found a moderate degree of consistency between the effectiveness rankings produced by the two models. About half of principals are placed into identical quartiles of performance by the two models. However, a noticeable minority of principals receive a ranking from the school VAM that differs substantially from their ranking from the transitions model.

School VAM estimates actually capture the contributions of entire schools, including some factors beyond principals' control. Nevertheless, given the moderate consistency of these estimates with those from the transitions model, some of the variation in these estimates among principals is likely to capture true differences in principal quality.

VAMs can generally distinguish among schools with respect to impacts on student assessment scores.

There are sizable differences among schools in VAM estimates. By switching from the 15th to the 85th percentile school, a 5th-grade student who originally scored better than half of all students in the state on the math PSSA would end up scoring better than two-thirds of all students. Moreover, performance differences among schools are estimated with greater statistical reliability than those among teachers due to larger student samples per school. In three-cohort models, typically at least two-thirds of schools can be statistically distinguished from the average based on math PSSA outcomes, and at least half can be distinguished from the average based on reading PSSA outcomes. These are, of course, differences in the total value-added of each principal's school(s). The proportion of the variation that is attributable to the principals themselves (versus other school characteristics that might be outside principals' control) is unknown.

Schools differ in their effectiveness at keeping students enrolled in high school.

We examined VAMs based on a nontest outcome called holding power, or the extent to which high-school students stay enrolled in a Pennsylvania school the following year; this might be viewed as a proxy for a school's effectiveness in preventing dropout. Impacts on holding power differ greatly between the worst-performing schools and all other schools in the state. For instance, the bottom 6 percent of schools lower their 9th graders' probability of enrolling in the following year by more than 30 percentage points relative to the average school. It is worth noting that the validity of these estimates depends on the assumption that the statewide data system has complete records on student enrollment. These estimates also do not include 12th graders, so they do not capture actual graduation outcomes. The data to study 12th graders will not be available until Phase 2 at the soonest. Despite these caveats, school effectiveness estimates for holding power appear to be an informative tool for identifying high schools that perform poorly in keeping their students enrolled in Pennsylvania's public schools.

Looking Ahead to Subsequent Pilot Phases

We offer several recommendations that relate broadly to strategies for sampling educators from the pilot districts and steps for refining and improving the performance measures. With regard to sampling, we recommend oversampling educators for whom we can generate value-added estimates with the greatest validity and relevance to the future evaluation model. In particular, because a future statewide evaluation model will almost certainly include the PSSA, we recommend including a

substantial number of math and English language arts teachers from grades 4 through 8 and science teachers in grades 4 and 8. We also recommend oversampling middle school principals when a new principal evaluation instrument is developed. Given that all middle school grades are tested by the PSSA, value-added scores and rubric scores will cover exactly the same grades for this set of principals. Additionally, teachers and principals should be recruited for the pilot to provide for more variation in the observation measure. Focusing on a limited range of performance inhibits the pilot's ability to differentiate between the practices of more and less effective educators.

Several steps can also be taken to improve the performance measures from the VAMs and the observational rubric. First, the assessment properties of the student outcomes—especially district-administered assessments—and the observational rubrics should be evaluated. This includes assessing interobserver agreement, or the rate at which different observers independently agree on a teacher's observation rating, and observer drift, or the tendency of two raters to agree with each other more frequently over time. Second, the quality of data linkages in Pennsylvania's student data should continue to be evaluated. Third, additional nonassessment outcomes for principal evaluations should be examined, such as by developing value-added models based on 12th-grade graduation outcomes. Fourth, the pilot should continue its progress toward identifying how different types of effectiveness data will be integrated in the overall evaluation model. We look forward to continuing our work on these efforts in Phase 2.

THIS PAGE LEFT BLANK FOR DOUBLE-SIDED PRINTING

I. INTRODUCTION

A. The Pennsylvania Teacher and Principal Evaluation Pilot

The Team Pennsylvania Foundation (Team PA) recognizes that the evaluation of teachers and principals is a critical foundation for the education reforms envisioned by the state's leaders. To develop an evaluation system that is accurate and fair, between 2010 and 2011 Team PA undertook the first phase of the Pennsylvania Teacher and Principal Evaluation Pilot (referred to as Phase 1) from which lessons learned will inform the development of a full, statewide evaluation system by 2013–2014.² Phase 1 proceeded along two tracks in collaboration with a broad stakeholder group that included representatives from the Pennsylvania Department of Education (PDE), the Pennsylvania State Education Association (PSEA), school districts, and the business community. The tracks were designed to pilot the development and implementation of measures that would improve the use of both classroom observations and student data in evaluating teacher and principal performance. None of the results from Phase 1 had a bearing on actual evaluations or personnel decisions for any teacher or principal.

In the first track, steering committee subgroups initially developed new observation-based rubrics for evaluating teachers and principals during fall 2010. In January 2011, principals and superintendents from Allentown, Cornell, and Mohawk Area school districts, and from Northwest Tri-County Intermediate Unit 5 (collectively, the Phase 1 pilot districts), were trained in the new protocols.³ These school and district leaders then implemented the new rubrics on a trial basis during the spring semester in their own districts to 153 preselected teachers and 30 preselected principals, respectively. Lane and Horner (2011) documented the process, progress, and lessons learned from the trial implementation in preparation for Phase 2, which will scale up the pilot to include educators from approximately 100 school districts starting in 2012.

In this report, we present findings for the second track of Phase 1. The second track involved using student data to develop value-added models (VAMs) for measuring teacher and principal contributions to student learning, and then examining professional practices that are positively associated with VAM estimates. We used data from the entire state of Pennsylvania for most analyses; we used data for districts covered in the first track of Phase 1 for other analyses.

A VAM is a statistical model that predicts students' levels of achievement based on students' own achievement histories and other characteristics. The difference between students' actual and predicted achievement (above or below zero) is averaged and attributed to their teachers or schools as a measure of the educators' contributions to student learning. Mathematica developed the VAMs for Team PA in Phase 1 and conducted analyses to address the three primary research questions for this report:

² Phase 1 was supported through a grant from the Bill & Melinda Gates Foundation to Team PA.

³ The National Institute for School Leadership developed and administered the training.

1. How can VAMs be used to characterize the effectiveness of teachers at raising achievement according to multiple outcome measures?
2. Do specific teacher practices relate to larger contributions to student learning among Phase 1 teachers?
3. How can principals' contributions to student learning be measured?

Several of the analyses done in this Phase 1 report are based on small samples of teachers, principals, and schools. Findings from these analyses *should be viewed as providing suggestive evidence that merits further attention in Phase 2 of the pilot*. For instance, the study's second research question relies on data collected once on 153 preselected teachers who teach in four school districts that are not representative of the state in terms of the characteristics of their students.⁴ The much larger Phase 2 pilot will provide results that are more precise and more representative of Pennsylvania teachers. To prepare for Phase 2, we invite feedback on how to refine the VAMs in this report to best reflect policy goals for the statewide model evaluation system.⁵

B. Description of Value-Added Models

A well-constructed VAM uses the prior achievement histories of individual students to produce valid estimates of what educators contribute to achievement, regardless of the starting points of their students. VAM estimates overcome a main deficiency of most levels-based measures, such as the rate of student proficiency, which penalize teachers and schools that serve historically low-performing students. By accounting for other observable background characteristics—such as socioeconomic or disability status—of the students assigned to each teacher or principal, VAMs can also overcome a main deficiency of simple growth-in-achievement models that penalize teachers and principals who serve at-risk or hard-to-teach students. Despite these advantages, VAMs—like all measures of performance—are imperfect measures. We recommend basing teacher and principal policy decisions, when possible, on multiple types of information that are combined in an optimal way to ascertain an individual's effectiveness as accurately and completely as possible.

1. Conceptual Framework

The process of estimating a value-added model includes two conceptual steps. In the first step, the VAM makes a prediction about an outcome of interest, typically a student's assessment score in a subject. This prediction is based on factors that include students' own achievement histories and usually other background characteristics about students and their peers. The prediction is derived using data on the performance of other students, either across Pennsylvania or the pilot districts, and represents what we expect a student to achieve if served by the teacher or school in the middle of the effectiveness distribution. It is derived from outcomes achieved by the other students in the same year; the word *prediction* does not mean that a VAM can project a student's future achievement. In the second step, the VAM compares students' actual outcomes with their predicted outcomes.

⁴ Analyses of principal practices were not conducted in Phase 1 because the observation rubric for principals will undergo substantial changes during Phase 2 and because the principal pilot was so small that meaningful analyses would not have been feasible.

⁵ The overall structure of the teacher and principal evaluation system is under development by PDE. Mathematica is not aware of any plans to include the VAMs developed specifically for this report in the evaluation system.

The VAM score for a teacher or school is the difference between actual performance and the predictions averaged across all students taught by a given educator.

Thus, a value-added model addresses the following central question: *To what extent does the actual level of achievement demonstrated by an educator's students exceed (or fall short of) the level that would have been expected for students with similar achievement histories and similar background characteristics if they had been taught by the educator in the middle of the effectiveness distribution?* A VAM does not measure student achievement growth. It instead seeks to produce something approaching a causal inference about the individual contributions of educators to the learning of students under their charge. Given the available data, VAMs arguably represent the best method for estimating educators' contributions to student learning as measured by assessment scores, but there are likely to be at least some factors that limit the accuracy and validity of these estimates.

Rothstein (2010) concluded that teacher effectiveness measures according to most VAMs lack validity because some teachers are more likely than their colleagues to be assigned students with particularly high or low gains in the previous grade. Fortunately, the degree of bias from this kind of sorting of students might not be large. Kane and Staiger (2008) compared teacher VAM estimates in Los Angeles under a typical situation in which principals assigned students to teachers to VAM estimates in the following year when principals randomly assigned teaching assignments—thereby eliminating the possibility of bias due to the sorting of students. They found that a higher VAM score before random assignment was a positive and significant predictor of achievement differences when classrooms were assigned randomly. In addition, Koedel and Betts (2011) found that the sorting bias identified by Rothstein can be reduced to statistical insignificance by including students from multiple cohorts in teacher VAM estimates, rather than just one cohort as in the Rothstein study. Goldhaber and Chaplin (2011) found that even without using multiple cohorts of students, the bias identified by the specification tests Rothstein uses might be very small.

Another reason to be cautious about interpreting a VAM estimate is that VAMs likely cannot control for all of the relevant factors needed to distinguish completely the teacher's or the school's contribution from every other factor affecting the performance of students. A VAM can control only for those factors that are observable in the data. If there are other student, peer, and school characteristics that influence student performance and that are not captured in the VAMs, they can artificially inflate the VAM estimates for some teachers and deflate the estimates for others.

A final consideration for interpreting the performance measures produced by VAM methods is that VAMs do not measure student achievement growth in absolute terms. They place educators on a distribution of performance relative to other educators with students in the same grade and subject on the specific student assessment used as the outcome. The value of VAMs depends in significant part on the validity of the underlying student assessments in capturing what students ought to be learning. Because VAMs are not measures of student achievement growth, they cannot measure growth with respect to the Pennsylvania Academic Standards. VAMs measure the difference between actual and predicted scores for outcomes that are, at best, proximal measures of academic standards.

2. Advantages and Limitations

VAMs have been studied extensively and have been the subject of considerable policy discussion at the local, state, and national levels. The policy interest in value-added has risen recently in response to the U.S. Department of Education's Race to the Top (RTTT) initiative, which makes

competitive grants to states that agree to make student achievement part of annual evaluations of teacher and principal effectiveness. A recent issue brief found that eight states and the District of Columbia recently enacted new legislation to make student performance a major component of evaluations for general education teachers (Pennsylvania Clearinghouse for Education Research, 2011). Many states mandate that half of a teacher's evaluation must depend on student achievement (accounting for prior achievement).

To facilitate a broader understanding of value-added and its potential use as a component in teacher and principal evaluations, we list key strengths and limitations of the approach in Table I.1. In addition, in September 2010 Mathematica conducted a synthesis of information on the research and implementation of VAMs for Team PA (Lipscomb et al. 2010b).⁶ In that review, we selected 21 studies that represent key issues and results in the literature and examined varying degrees of value-added implementation in seven school districts or states.

Table I.1. Strengths and Limitations of Value-Added Models Relative to Other Evaluation Methods

Strengths	Limitations
Focuses on outcomes rather than practice so it might encourage educators to better tailor practice to student needs	Restricted to effectiveness as measured through outcomes that can be systematically measured
Provides an objective measure of performance at the level of the individual teacher or school	Applied only in tested grades and subjects
Produces estimates of educators' contributions to achievement growth that account for students' starting points and other observed characteristics	Connection between school value-added and principal effectiveness is unclear
Results known to differentiate among staff at least at the tails of the performance distribution	Communicating statistical methodology to nontechnical audiences can be difficult

Sources: Pennsylvania Clearinghouse for Education Research (2011) and Mathematica.

The research synthesis highlighted several general findings that were used, in turn, to inform the goals and subsequent analyses undertaken for this report. We found consistent support for the existence of a wide distribution of teacher effectiveness with respect to student test score growth. As one might expect, teacher quality is the most important school-based factor affecting students. In most studies, the top 15 percent of math and reading teachers were capable of raising the achievement of the median-performing student at least 5 to 8 percentile points with one year of teaching compared with the teacher with the median value-added score.

We also found that few research studies examined the application of value-added to principals, although numerous studies examined its application to teachers and schools.⁷ Due to the scarcity of research on principal value-added, we investigate in this report whether the average contribution to student achievement among educators at a principal's school approximates the principal's contribution, as the two should not be presumed to be synonymous. We ultimately conclude that,

⁶ The review is available online through Mathematica's web site at [http://www.mathematica-mpr.com/publications/PDFs/education/teacherprin_valueadded.pdf].

⁷ Dhuey and Smith (2011) is a recent addition to the principal value-added literature.

for many principals, it is impossible to distinguish the principal's contribution to student achievement from the contribution of other facets of the school (notably including the collective contribution of teachers). In consequence, throughout the report we label the principal-based measure as an estimate of the value-added of the *principal's school*, rather than an estimate of the value-added of the principal.

As indicated by Table I.1, value-added provides an objective measure of individual performance but one focused narrowly on test scores. The need to rely on assessment data has proven to be a practical challenge in extending value-added to an entire teaching staff. This limitation underscores the importance of determining through the Pennsylvania pilot study whether certain teacher or principal practices that can be measured through classroom observations in all grades and subjects are strongly tied to larger contributions to student achievement growth in tested grades and subjects.

The research literature also makes clear that an evaluation system can be considered fair only if it is based on valid and reliable measures. By validity, we mean whether the evaluation model measures what it intends to measure or whether it systematically over- (or under-) estimates performance for some teachers or principals. By reliability, we mean whether repeated measurements lead to a consistent result. Critics of value-added have voiced concerns that it is a noisy signal and that any of a litany of important factors can lead to the misclassification of some teachers as high or low performers (for example, nonrandom assignment of students into classrooms, small samples, incomplete statistical controls, or assessments that do not reflect the curriculum or standards). These concerns should not be swept under the rug. At the same time, we feel that they are not reasons to discard value-added analyses entirely. We share the view of a recent Brookings task force comprised of national experts on teacher quality in arguing that the best response is “to improve value-added measures continually and to use them wisely, not to discard or ignore the data” (Glazerman et al. 2010).

When the outcome is student test scores, value-added has been shown to be a better indicator of teacher effectiveness than teacher graduate degrees, certification, and experience after the initial five years of service (Goldhaber and Hansen 2010). Glazerman et al. (2010) also caution against setting unrealistic expectations for value-added as a performance measure, pointing out that the year-to-year correlation of value-added estimates for teachers—though modest—is as good as what has been found for measures used to make high-stakes decisions in other occupations. Value-added almost certainly provides better information for evaluating teacher and school effectiveness when compared against the alternative of maintaining the current system of evaluation in many school districts and states. In 2011, PDE found that 99 percent of teachers in the Commonwealth received a satisfactory rating for the 2009–2010 year (Team Pennsylvania Foundation 2011). In other words, the current system differentiates only a very small number of teachers with the absolute lowest ratings. Improving the evaluation framework will involve increasing the ability to differentiate high and low performance. It will also require ensuring that raters are trained to implement the new framework consistently for any new system to be deemed fair (Lane and Horner 2011).

In the following chapters, we present findings from analyses that address the study's three research questions. In Chapter II, we describe characteristics of the VAMs, such as the outcome measures, control variables, and applicability of estimates to Phase 1 teachers. We then present findings pertaining to teacher effectiveness measures using state-mandated and other assessments in Chapter III. In Chapter IV, we characterize relationships between teacher effectiveness and teacher practices to the extent possible in the Phase 1 pilot sample. We then present findings pertaining to principal and school effectiveness measures based on assessment and non-assessment data in

Chapter V. Finally, we provide a brief conclusion in Chapter VI with recommended next steps for this strand of the pilot study in subsequent phases. Interested readers are directed to Appendices A through C for technical information on the methodology, samples, and results, respectively.

II. CHARACTERISTICS OF VAMS ESTIMATED IN THIS REPORT AND THE APPLICABILITY OF EFFECTIVENESS MEASURES TO TEACHERS IN PHASE 1

The value-added models (VAMs) for this report include different outcome measures, control variables, and student samples. In this chapter, we provide a nontechnical description of the characteristics of the VAMs for teachers and schools that produce the results we discuss later in the report. We list the outcome measures, prior achievement controls, other background variables, and student samples that are included. We also show the extent to which Phase 1 teachers have at least one VAM estimate from across outcomes and therefore can be included in the analysis that examines relationships between value-added and observation-based measures.

A. Outcome Measures

We selected outcome measures for this report using the following two criteria that reflect goals for the pilot analysis:

1. The set of outcomes should include multiple measures of student outcomes, including non-PSSA test-based measures and nontest measures.
2. The value-added estimates based on the set of outcomes should include as many teachers from Phase 1 as possible with at least one estimate.

These selection criteria are consistent with the purpose of a pilot study in which findings are used to inform future development work and have no actual consequences for teachers, principals, or schools. In deciding whether to include specific outcomes, we did not assess the degree to which the measures correspond to the content that teachers are asked to teach or to which scores are indicators of skill acquisition by students. Our focus was in estimating VAMs to assess the extent to which attributions to teachers or principals are feasible. We withhold judgment on whether specific outcomes should or should not be included in Pennsylvania's model statewide evaluation system. Deciding which outcomes to include in the actual evaluation model will involve policy discussions that are outside the scope for Phase 1 (for example, discussions about a measure's degree of alignment with curriculum and standards, its validity and reliability, whether it is administered to all students or only to some students in a grade, the extent to which scores are malleable, and whether/how to allow for discretion at the district level in selecting measures).

In Table II.1, we list the student outcomes that are used in the primary VAM calculations for this report. The test-based outcomes come from the Pennsylvania System of School Assessment (PSSA), from Allentown's Progress Assessment (Progress), and from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). The PSSA is the statewide assessment that is given to all students in grades 3 to 8 and 11. It is also used for compliance with federal school accountability policies. Progress Assessments are curriculum-based measures that were developed by teacher committees in Allentown. They are administered to students multiple times during the year and are cumulative up to the date they are given. DIBELS includes several diagnostic measures that teachers can use to monitor students' early literacy and early reading skill development. The nontest outcomes include a student's rate of attendance and a measure that we constructed and refer to as

holding power.⁸ The holding power variable is a binary measure of whether high school students during a given academic year enroll in any Pennsylvania public school the following year, which we interpret as an indicator of students who are likely to complete high school.⁹ Although it will overestimate dropout rates across the state (because some students disappearing from the data are enrolled in private schools and others are enrolled outside the state), we expect it to permit a fair comparison among schools. We include attendance and holding power outcomes for school evaluation but not for teacher evaluation because these measures are likely to be affected by multiple staff at the building level.

Table II.1. Outcomes Considered in Value-Added Models for Teacher and School Evaluation in this Report

Outcome	Subject(s)	Grade	Teacher Evaluation	School Evaluation	Cohorts
PSSA (scaled score)	M, R	3	A, C, M	A, C, M	1
PSSA	M, R, S	4	PA	PA	3
PSSA	M, R, W	5	PA	PA	3
PSSA	M, R	6	PA	PA	3
PSSA	M, R	7	PA	PA	3
PSSA	M, R, W, S	8	PA	PA	3
PSSA	M, R, W, S	11	A, C, M	PA	1 (T); 2 (P)
Progress (raw score)	W	1	A	A	1
Progress	M, W	2	A	A	1
Progress	W	3	A	A	1
DIBELS (raw score)	R (NWF, PSF)	1	A	A	1
DIBELS	R (ORF)	2	A, C	A, C	1
Attendance (%)	--	4-12	--	A, M, N	1
Holding Power {0,1}	--	9	--	PA	3
Holding Power	--	10	--	PA	2
Holding Power	--	11	--	PA	1

Note: VAMs based on PSSA scores include students taking the modified version of the PSSA.

Subject abbreviations: M = Math; R = Reading; S = Science; W = Writing. DIBELS abbreviations: NWF = nonsense word fluency; ORF = oral reading fluency; PSF = phoneme segmentation fluency. Sample abbreviations: A = Allentown; C = Cornell; M = Mohawk; N = Northwest Tri-County; PA = Pennsylvania. T = Teacher; P = Principal.

-- indicates outcomes that are not specific to a particular academic subject or for teacher evaluation.

Table II.1 also indicates the subjects, grade levels, samples, and number of student cohorts for each VAM. All VAMs are estimated separately by subject and grade except for the attendance rate VAM. We included multiple grades of attendance information together to maximize sample sizes within the Phase 1 districts, because attendance data were not available to us statewide. For

⁸ Some researchers use the term *value-added* only when there is a baseline measure of the outcome. We use the term for models without baseline measures of the outcomes because the methodology is very similar; in particular, it still involves comparing actual and predicted values of an outcome.

⁹ In the VAMs for holding power, a student's enrollment decision in the following year is attributed only to the school that a student attends in the current year. This approach ignores any lingering effects of a student's previous schools. However, this approach is consistent with all other types of teacher and school VAMs that attribute a student's current-year test score growth only to the effects of the student's current-year teachers or schools.

Pennsylvania's evaluation system, our preference is for statewide samples whenever possible because the findings are the most inclusive; VAMs based on just a subset of districts are representative only of the districts they include. The lack of a statewide sample for the attendance VAM thus underscores an important point about VAMs: *The viability of any outcome measure in a VAM relies fundamentally on its availability across students who are relevant for the analysis.* When statewide samples were not available for other outcomes, we took the same approach of requesting the information directly from any Phase 1 district that collected it.

For three sets of PSSA models—grades 3 and 11 for teachers and grade 3 for schools—we are limited to pilot district samples even though the measures are collected statewide. Because grade 3 is the first year of state-mandated testing, there is no available baseline achievement measure that is collected across the Commonwealth. In order to include grade 3, we obtained student data on the fall administration of the grade 3 4Sight assessment from Phase 1 districts and used those scores to control for students' baseline achievement levels. A related problem affects the VAMs in grade 11 because students are not assessed statewide in grades 9 and 10. For the school VAMs, we are able to preserve the Pennsylvania sample by using students' grade 8 scores as their baselines, thus measuring contributions to achievement between grades 8 and 11. This is allowable for principal–school models because students are typically served by the same school during high school grades. Teachers, however, affect students in the year they educate them, making it critical to establish a baseline either at the end of grade 10 or at the beginning of grade 11. As we describe later in this chapter, we used fall 4Sight scores from Phase 1 districts, thus limiting the student sample to those districts.

In the final column of Table II.1, we show the number of student cohorts in each VAM. By cohort, we mean all the students a teacher educates or all the students attending a principal's school during an academic year. Incorporating multiple cohorts of students into a VAM can improve both the validity and the reliability of the estimates by averaging out random year-to-year fluctuations in student performance that affect teacher or school estimates from a single year of data (Schochet and Chiang 2010). Koedel and Betts (2011) showed that multiple cohorts improve validity as well because systematic biases offset one another over multiple years. Our primary models include all available student cohorts, up to three, moving backward in time from the most recent school year. For example, the three-cohort teacher VAM for grade 4 math includes all students a teacher taught in math between 2008–2009 and 2010–2011 who took the grade 4 math PSSA. For several outcomes, only one or two cohorts of students can be included using the data that we can access currently.¹⁰ Estimates based on VAMs that include fewer student cohorts will be measured with greater noise, but they also have the advantage of better reflecting immediate past performance.

In the future, the Pennsylvania Department of Education (PDE) might wish to pursue a different set of outcome measures, including measures not included in this report. We focused narrowly on the academic subjects covered by the Phase 1 pilot (that is, math, English-language arts, and science). We also considered—but ultimately did not pursue—models based on the Pennsylvania Alternate System of Assessment (PASA), the 4Sight, and core course passage rates.

¹⁰ School VAMs based on grade 11 PSSA data include two cohorts because we use students' scores from three years earlier as their baseline scores and have data back only to 2006–2007. One cohort of student data is available for outcomes based on pilot samples. The number of cohorts in holding power VAMs differs by grade. In future years, three cohorts will be available for all grades, including grade 12.

The PASA is given to students with severe cognitive disabilities instead of the PSSA if specified by their Individualized Education Program. On average, there are one or two PASA students per school and grade in Pennsylvania. Consistent with other studies, we report estimates for individual teachers and schools only when they are based on more than 10 students. Thus, the PASA data would not have been sufficient to estimate impacts for most teachers or schools in our sample. Moreover, at the school level, we found in exploratory work that including PASA only marginally increased the number of schools in Pennsylvania with at least one VAM estimate above the number obtained through the PSSA alone. Finally, there are technical issues related to involving the PASA that would be too resource-intensive to resolve for this report given the possible benefits of including it.¹¹ We thus exclude this measure and do not include students with severe cognitive disabilities in this report. However, we are able to include the vast majority of student with disabilities because most of them take either the PSSA or the modified version of the PSSA.

The 4Sight is a quarterly formative assessment that is intended for teachers as a low-stakes diagnostic indicator of student performance on content that mirrors the PSSA. We did not include it as an outcome measure (despite including it as a baseline measure for some VAMs) because it is given in the same subjects and grades as the PSSA, therefore meaning that it would not augment the coverage of value-added estimates to teachers. We prefer the PSSA as a measure because it is already used for school accountability, suggesting that teachers are motivated to have their students perform well on that test. Lastly, we examined the potential to use core course completion rates as a nontest outcome at the high school level. Though the data were available in the Phase 1 districts, we did not include those data because the small size of the Phase 1 pilot meant that we would not be able to present the findings without inadvertently identifying some schools.

B. Teachers with VAM Estimates from Phase 1 of the Pilot

Using the assessments listed in Table II.1, we were able to cover slightly more than half of the 153 teachers who participated in Phase 1 with at least one value-added estimate. Each teacher was observed by his or her principal in one grade and subject. Classroom observation data were not collected on pilot teachers in multiple subjects and grades even though the teachers might educate students in multiple subjects and grades. In Table II.2, we show how the pilot teacher sample was distributed across grades and subjects. The sample was selected by Dr. Suzanne Lane at the University of Pittsburgh, with input from Mathematica and superintendents in the pilot districts. It was limited to math, English-language arts, and science because assessment data are most often available in these subjects. Grades were selected to be representative of the K–12 spectrum. The sample sought to balance PSSA-tested grades and subjects and other grade/subject combinations in which the PSSA is not administered. More than half of the sample came from Allentown due to that district's size relative to the others.

¹¹ The VAM would have to account for how the PASA is reported in Pennsylvania's longitudinal student data on a categorical, rather than continuous, scale and is administered at three different levels of difficulty. Furthermore, there are substantial sample-selection concerns related to treating students who alternate between taking the PASA and a version of the PSSA in different years.

Table II.2. Summary of Teacher Evaluation Pilot, Phase 1

Subject	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Middle School	High School	Total
Math	6	4	8	0	11	14	16	59
English-Language Arts	10	7	8	0	7	16	12	60
Science	0	5	0	12	0	8	9	34
Total	16	16	16	12	18	38	37	153

Note: Participants included teachers from the following school districts: Allentown (84), Mohawk Area (39), Cornell (20), and Northwest Tri-County IU5 (10).

The final sample of the 153 teachers for Phase 1 included 79 fewer teachers than had originally been selected across these same grades and subjects. The sample reduction was due primarily to the loss of one school district and one charter school that were slated to participate. We doubt that the sample loss affected our success rate in mapping VAM estimates to participating teachers because these 79 teachers had been assigned fairly evenly across grades and subjects.

The more serious concern for the pilot is that the Phase 1 sample is under-powered. Only 81 Phase 1 teachers have a VAM estimate that we can use in Chapter IV for studying relationships between teacher practices and larger individual contributions to student achievement. Based on this sample size, we can detect at best a 0.30 correlation between value-added and classroom observation scores.¹²

In Table II.3, we report the number and percentage of Phase 1 teachers with at least one VAM estimate from the analyses undertaken for this report. Overall, 53 percent of Phase 1 teachers have at least one VAM estimate that can be included in the Chapter IV analyses. When a value-added estimate could not be assigned, it was for one of two primary reasons. First, assessments were not always available in subject/grades/districts covered by the pilot (for example, second-grade science, or second-grade math outside of Allentown). Second, teachers did not always educate more than 10 students with an assessment score in the subject for which they were observed—a minimum number of students that we specified based on prior studies reporting estimates that are not overly noisy due to small sample sizes. This latter constraint affected all Phase 1 teachers in Northwest and many in Allentown who teach primarily special education students and students for whom English is a second language.

¹² This power calculation assumes a power level of 0.80 and a 5 percent confidence interval.

Table II.3. Number of Teachers from Phase 1 with at Least One VAM Estimate that Can Be Used for Correlating Value-Added with Teacher Practices in Chapter IV

Subject	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 8	Grade 11	Total
Math	0	1	6	0	9	5	7	28
English-Language Arts	5	5	7	0	6	5	5	33
Science	0	0	0	11	0	4	5	20
Total	5	6	13	11	15	15	17	81

Finally, we were not able to use the data on any teacher in Cornell because the evaluators in that district assigned all Phase 1 teachers exactly the same score on all rubric items, meaning that there is no variation in the classroom observation data for this district. In principle, these teachers could still be included in the sample for studying relationships between teacher practices and teacher effectiveness as measured by value-added because their data are complete. But preserving them in the sample does not contribute any information to the analysis; rather, it adds only noise. The teacher counts in Table II.3 reflect the deletion of the Cornell teacher sample. Omitting these teachers, we are able to use data on 61 percent of the remaining Phase 1 teacher sample in the Chapter IV analysis.

C. Control Variables that Are Included in the VAMs

1. Baseline Student Achievement

All VAMs used in education make predictions about student performance based on students' own achievement histories. Most researchers include prior scores from multiple academic subjects regardless of the subject of the outcome measure. We selected baseline measures—listed in Table II.4—by following a two-part strategy that is applied to each VAM based on the particular sample used:

1. Include scores from all available subjects in either the fall of the current grade or the spring of the prior grade—treating grade 8 scores as prior-grade scores for grade 11 students and showing preference for including measures that would allow for a statewide analysis if possible.
2. Include a same-subject PSSA score from two prior grades if one is available—substituting math for science scores and reading for writing scores

Table II.4. Baseline Measures for Value-Added Models Estimated in this Report, by Outcome

Outcome	Subject(s)	Grade	Sample	Prior Grade Baselines	Other Baseline Controls
PSSA	M, R	3	A,C,M	--	4Sight, fall Gr. 3 (M, R)
PSSA	M, R, S	4	PA	PSSA, Gr. 3 (M, R)	--
PSSA	M, R, W	5	PA	PSSA, Gr. 4 (M, R, S)	PSSA, Gr. 3 (M or R)
PSSA	M, R	6	PA	PSSA, Gr. 5 (M, R, W)	PSSA, Gr. 4 (M or R)
PSSA	M, R	7	PA	PSSA, Gr. 6 (M, R)	PSSA, Gr. 5 (M or R)
PSSA	M, R, W, S	8	PA	PSSA, Gr. 7 (M, R)	PSSA, Gr. 6 (M or R)
PSSA (Teacher)	M, R, W, S	11	A,C,M	--	4Sight, fall Gr. 11 (M, R) PSSA, Gr. 8 (M, R, W)
PSSA (School)	M, R, W, S	11	PA	--	PSSA, Gr. 8 (M, R, W)
Progress	W	1	A	--	Progress, fall Gr. 1 (W); DIBELS, fall Gr. 1 (NWF, PSF)
Progress	M, W	2	A	--	Progress, fall Gr. 2 (M, W); DIBELS, fall Gr. 2 (ORF)
Progress	W	3	A	--	Progress, fall Gr. 3 (W); 4sight, fall Gr. 3 (M, R)
DIBELS	R (NWF, PSF)	1	A	--	Progress, fall Gr. 1 (W); DIBELS, fall Gr. 1 (NWF, PSF)
DIBELS	R (ORF)	2	A,C	--	DIBELS, fall Gr. 2 (ORF)
Attendance (%)	--	4-12	A,M,N	Attendance, Gr. 3-11;-- PSSA, Gr. 3-11 (M, R)	
Holding power {0,1}	--	9-11	PA	PSSA, Gr. 8 (M, R, W)	--

Notes: Baselines are given in the spring of the prior grade unless otherwise indicated.

Subject abbreviations: M = Math; R = Reading; S = Science; W = Writing. DIBELS abbreviations: ORF = oral reading fluency; NWF = nonsense word fluency; PSF = phoneme segmentation fluency. Sample abbreviations: A = Allentown; C = Cornell; M = Mohawk; N = Northwest Tri-County; PA = Pennsylvania.

Controls for prior achievement are the most important factors in any VAM because a student's own achievement history is the most important factor by far in predicting actual achievement at the end of the year—much more important statistically than the contributions of the teacher in any single year. Adding more extensive controls for prior student achievement can provide for better predictions about student achievement, enhancing the internal validity of effectiveness measures. But it typically comes at a cost of excluding students who lack scores on the additional assessments for which controls are being added. In other words, more extensive controls can yield better indicators of teacher effectiveness but the indicators are applicable to a smaller number of the teacher's students. This tradeoff between sample size and greater controls is clearest around the issue of whether to include a score from two prior grades ago (that is, grade 4 for 6th graders) because it has implications for whether mobile students can be included. We opted for the greater internal validity because we found that the direct sample loss was only 5 percent.¹³ However, we note that the students who are dropped might not be a random sample of students, as they could differ relative to other students on characteristics beyond their mobility.

¹³ The sample loss is lower than what would be found in the data systems of individual school districts because the statewide data retain the achievement histories of students who move between districts in Pennsylvania.

2. Additional Student- and Classroom-Level Variables

Along with controlling for baseline student achievement, most VAMs account for observable student background characteristics to help isolate further the contributions of educators to student achievement. The factors that are included in the VAMs are thought to be correlated with student performance while also being outside the control of teachers and schools. The standard list of controls would include measures related to students' socioeconomic status (for example, parent educational attainment, family income, or proxies such as eligibility for the free or reduced-price meals programs); family structure (for example, living in a single-parent household); or eligibility for programs such as special education. Unfortunately, there is usually a discrepancy between the variables that ideally would be included and the variables that are available in the data system. Researchers and policymakers are then left with a very difficult choice between estimating a model that could systematically over- or under-estimate teacher contributions due to less-than-complete controls and attempting to compensate at least partially for the omitted variables by including other measures that are available in the data. In practice, most data systems collect only limited information on student background characteristics, typically basic demographic variables such as gender, race/ethnicity, meals program eligibility, disability status, and English-language learner (ELL) status. Ultimately, most researchers and policymakers opt to include whatever information is available. At the same time, they acknowledge that a different set of variables would be preferable. The unavailability of student background controls is a difficulty in the short run, but data systems can be expanded over time to allow for a different set of variables to be used.

We adopted this same approach of including measures that are available in the data system both because we find that they are significant predictors of student performance and because there is a foundation for including them in prior research studies (Lipscomb et al. 2010b). The measures, listed in Table II.5, include variables for meals program eligibility, ELL status, categories of disability, mobility, grade repetition and age, flags for the modified version of the PSSA, gender, and race/ethnicity. We are not able to control for other measures of socioeconomic status, measures of family structure, or prior rates of student attendance in the data available.

Table II.5. Student and Classroom Control Variables Included in VAMs Estimated for this Report

Control Variable	Definition	Used in VAMs for Schools	Used in VAMs for Teachers
Free Meals	Free meals eligibility {0,1}	√	√
Reduced-Price Meals	Reduced-price meals eligibility {0,1}	√	√
English-Language Learner (ELL)	ELL in outcome year {0,1}	√	√
Specific Learning Disability (SLD)	Designation of SLD under IDEA {0,1}	√	√
Speech or Language Impairment (SLI)	Designation of SLI under IDEA {0,1}	√	√
Emotional Disturbance (ED)	Designation of ED under IDEA {0,1}	√	√
Intellectual Disability (ID)	Designation of ID under IDEA {0,1}	√	√
Autism (AUT)	Designation of AUT under IDEA {0,1}	√	√
Physical/Sensory Impairment	Designation of hearing impairment, visual impairment, deaf-blindness, or orthopedic impairment under IDEA {0,1}	√	√
Other Impairment	Designation of other health impairment, multiple disabilities, developmental delay, or traumatic brain injury under IDEA {0,1}	√	√
Mobility	Attended multiple schools during school year {0,1}	√	√
Grade Repeater	Repetition of the current grade {0,1}	√	√
Behind	More than 1.5 years older than expected for grade {0,1}	√	√
Age	Student age in years as of September 1	√	√
PSSA-Modified (outcome)	Outcome is a PSSA-M score (PSSA outcomes only) {0,1}	√	√
PSSA-Modified (baseline)	Baseline is a PSSA-M score (PSSA baselines only) {0,1}	√	√
Gender	Female {0,1}	√	√
Race/Ethnicity	Indicators for African American, Hispanic, Asian Pacific Islander, or other race/ethnicity {0,1}	√	√
Classroom-Level Characteristics	Separate classroom-level variables for free meals, reduced-price meals, ELL, special education, gender, and race/ethnicity (percentage of students in the classroom)		√
Classroom Size	Number of students in the classroom		√
Classroom Size Interactions with Student-Level Characteristics	Separate interaction terms between classroom size and the following student-level characteristics: ED, ID, AUT, physical/sensory impairment, free meals, and ELL		√

Note: The value of a classroom-level variable for a particular student is an average across the courses that a student takes in the subject assessed by the outcome measure during the year.

Abbreviation: IDEA=Individuals with Disabilities Education Act.

Among these variables, the inclusion of gender and race/ethnicity controls is the most controversial. The intent is not to set different standards for students. Rather, it is an empirical acknowledgement that in the absence of preferable measures, these variables explain a statistically significant portion of the variation in student performance even after accounting for prior student achievement and all the other variables in Table II.5. To the extent that gender and race/ethnicity

represent unobserved factors that differ across students and are outside the control of teachers and schools, the VAM estimates would systematically penalize or advantage certain teachers and schools if these controls were omitted. If fuller controls were available, we would expect that the amount of variation that gender and race/ethnicity control for would shrink and eventually become statistically insignificant.

In addition to the student-level variables that we include in all teacher and school VAMs, we also include several classroom-level variables in teacher VAMs that account for peer influences on achievement.¹⁴ These measures are intended to account for various inputs that are largely beyond the control of teachers but affect their overall workload. The controls include the average characteristics of students in the classroom, the classroom size, and interaction variables between classroom size and student characteristics that indicate more severe needs. When a student takes multiple courses during the year in a subject, the peer variables are averaged across classrooms. We do not include classroom-level variables in the school VAMs because the make-up of classrooms is within the control of school administrators. We also do not include any measures related to educators' own characteristics (for example, their years of experience) that might affect their effectiveness relative to other educators.

We adjust some teacher effectiveness measures by subtracting the average value-added score among teachers in a school or district. This type of adjustment has the potential to control better for school- or district-level influences that affect the performance of all teachers at a school or district. But it also has two implications that might be disadvantageous for a statewide evaluation system. First, this type of adjustment changes VAM inferences so that teachers are compared only with other teachers in their same district or school, rather than with other teachers in the state. Second, it might under-represent true differences in teacher contributions across districts and schools if highly effective (or ineffective) teachers tend to cluster. For these reasons, we use the adjusted estimates for assessing the extent to which the variation in teacher effectiveness is primarily across or within districts and schools but not as a part of our primary VAMs. We believe that identifying district- or school-level variables that could control for this variation without preventing a statewide comparison of effectiveness would be very informative in follow-up work.

¹⁴ The exception is for VAMs that include a single cohort of students in elementary grades. Because classrooms in elementary grades tend to be self-contained, it is not possible to separate a teacher's influence from the influence of students' peers with one year of teaching data.

III. VALUE-ADDED RESULTS FOR TEACHERS

In this chapter, we present findings from value-added models (VAMs) that are intended to produce measures of teacher effectiveness. Our focus is on characterizing the distributions of teacher effectiveness across outcome measures, subjects, and grades. We begin by discussing teacher quality distributions in selected grades and subjects based on Pennsylvania System of School Assessment (PSSA) scores, and we conclude that sizable differences in quality exist across Pennsylvania. We then describe issues related to statistical uncertainty and how VAMs quantify the extent of imprecision through confidence intervals. Next, we contrast the teacher effectiveness estimates with estimates obtained through several alternative specifications to examine the sensitivity of results. The alternative specifications adjust the VAMs for district and school factors, omit a same-subject prior score control, and control for prior achievement using a beginning-of-year score rather than an end-of-year score, respectively. In the final section of this chapter, we describe several key characteristics of the teacher effectiveness estimates generated by estimating VAMs on additional outcomes and student samples from the pilot districts.

A. Variation in Teacher Effectiveness Based on PSSA Outcomes

Consistent with findings on teacher quality in the research literature, we find sizable differences in teacher effectiveness across Pennsylvania, as measured by value-added in math, reading, and science. In Table III.1, we depict the variation in teacher effectiveness based on PSSA scores for three subject-grade combinations covered by the Phase 1 pilot (that is, 5th-grade math, 8th-grade reading, and 4th-grade science). Teacher impacts are reported in terms of PSSA scaled scores at different points in the teacher quality distributions. The table values represent the expected difference in scaled scores between students educated by a given teacher and students educated by the average-performing Pennsylvania teacher, controlling for the factors described in Chapter II.

Table III.1. Distribution of Teacher VAM Estimates for Selected PSSA Outcomes

Outcome	Effectiveness of the Teacher at the Indicated Percentile Relative to the Effectiveness of the Average Teacher (in PSSA scale points)					
	5th	15th	25th	75th	85th	95th
Math PSSA, Grade 5	-70	-43	-28	+25	+44	+77
Reading PSSA, Grade 8	-35	-22	-15	+14	+22	+38
Science PSSA, Grade 4	-60	-39	-26	+24	+38	+68

Source: Mathematica calculations reported in Appendix Table C.1 based on Pennsylvania student data.

Note: Findings are based on a three-cohort model with statewide samples of teachers and students. The sample of teachers consists of those who served as teachers in every year from 2008-2009 to 2010-2011 in the outcome subject and grade, and their VAM estimates are based on students in their classrooms during that period.

PSSA = Pennsylvania System of School Assessment; VAM = value-added model.

In a single year of instruction by a teacher at the 15th percentile, a 5th-grade student originally at the median of the statewide distribution of math scores would score 43 scale points lower on the math PSSA than he or she would score with a single year of instruction by an average-performing teacher. On the other hand, this student would score 44 scale points higher by having the 85th percentile teacher than by having the average teacher. Thus, the 85th and 15th percentile teachers differ in their effectiveness by 87 PSSA points. This scale point difference can also be interpreted with reference to the statewide distribution of student test scores. By switching from the 15th

percentile teacher to the 85th percentile teacher, a 5th-grade student originally at the median of the statewide distribution of math scores would be predicted to rise to the 65th percentile.¹⁵ Sizable variation in the teacher effectiveness estimates is also observed for other grade–subject combinations. The 85th and 15th percentile teachers differ in their estimated effects on PSSA scores by 44 points in 8th-grade reading and 77 points in 4th-grade science. The impact of one year of teaching by a teacher at the 85th percentile relative to a teacher at the 15th percentile is about 54 percent as large as the 2011 test score gap between African American and white students in 5th-grade math, 24 percent as large in 8th-grade reading, and 45 percent as large in 4th-grade science.¹⁶

Pennsylvania’s most-effective teachers are certainly capable of moving the academic needle in the Commonwealth. However, teacher effectiveness is just one of many school-based and nonschool factors that affect students during the year. Students’ own prior achievement scores are by far the most important predictors of their actual achievement scores. In Appendix Table C.2, we show estimated coefficients for the control variables that are included in the models reported in Table III.1. Baseline scores clearly have the most explanatory power, a finding that is common to all VAMs, not just to these three selected ones. The relationships between achievement and the other student- and classroom-level variables are typically statistically significant as well, due partly to a very large number of student observations, though they have much smaller magnitudes.

The teacher quality estimates that we find in Pennsylvania for math and reading are similar in size to those found by other researchers in different states and school districts (see Table III.2). We can compare the size of different teacher quality distributions by expressing the effectiveness estimates in terms of standard deviations of student scores relative to the average score. A standard deviation is approximately the amount by which the 85th percentile student score exceeds the 50th percentile score (or equivalently, it is approximately the amount by which the 50th percentile score exceeds the 15th percentile score). In Table III.2, we report teacher estimates from prior research in these terms for the teacher at the 85th percentile of effectiveness relative to the teacher at the 50th percentile. For example, the value of 0.20 would indicate that by switching from the 50th percentile teacher to the 85th percentile teacher, the score of a student originally at the median of the statewide distribution of scores would be predicted to rise by 0.20 standard deviations. This gain translates into an increase from the 50th percentile of student scores to the 58th percentile of scores. Smaller values in Table III.2 indicate that teachers are grouped tightly together in terms of their performance. Larger values indicate that teachers are spread farther apart.

¹⁵ To make this calculation, we divided 87 scale points by 223 scale points, the standard deviation of 5th-grade PSSA math scores (see Appendix Table C.1). Thus, an 87 scale-point difference amounts to a difference of 0.39 standard deviations in the distribution of student scores. In the assumed normal distribution for student scores, moving from the 50th to the 65th percentile is equivalent to an increment of 0.39 standard deviations.

¹⁶ The PSSA achievement gap between African American and white students in 2011 was approximately 160 scaled score points in 5th-grade math, 180 scaled score points in 8th-grade reading, and 170 scaled score points in 4th-grade science, among students with prior-grade scores in math and reading.

Table III.2. Teacher VAM Estimates in Recent Studies for the 85th Percentile of Effectiveness Relative to the 50th Percentile, Reported in Standard Deviations of Student Test Scores

Research Study	Math	Reading	Grade Range	Location
This study	0.16-0.23	0.09-0.16	4-8	Pennsylvania
Aaronson et al. (2007)	0.15	--	9	Chicago
Goldhaber and Hansen (2010)	0.22	0.10	4-5	North Carolina
Hanushek and Rivkin (2008)	0.13-0.20	--	4-8	Texas (1 large urban district)
Jacob and Lefgren (2008)	0.26	0.12	2-6	Western United States (1 midsize district)
Kane et al. (2008)	0.17-0.21	0.17-0.20	4-8	New York City
Kane and Staiger (2008)	0.16-0.19	0.13-0.16	2-5	Los Angeles, New York City, Boston
Koedel and Betts (2011)	0.18-0.24	--	4	San Diego
Lipscomb et al. (2010a)	0.15-0.20	0.11-0.14	4-8	Pittsburgh
Rothstein (2010)	0.15	0.11	5	North Carolina

Sources: Appendix Table C.1 and the individual articles, most of which are summarized in Lipscomb et al. (2010b).

Note: Findings from Lipscomb et al. (2010a) are for three-cohort VAMs using PSSA score outcomes.

PSSA = Pennsylvania System of School Assessment; VAM = value-added model.

We find that Pennsylvania teachers are capable of affecting test scores more in math than in reading, as indicated by larger standard deviations in math—a finding supported by external research studies.¹⁷ We also find that distributions of effectiveness are larger in elementary grades than in middle school grades (Appendix Table C.1). Kane et al. (2008) found a similar pattern in estimating value-added in New York City, as did Lipscomb et al. (2010a) in Pittsburgh. The remaining studies in the table did not examine value-added distributions by grade level, but larger effectiveness distributions tended to be found in elementary grades across these studies.¹⁸

Teachers are capable of producing larger achievement gains in 4th-grade math than in 8th-grade reading (for example) partly because students tend to make relatively larger improvements in math and in elementary grades. A useful way to compare the size of teacher effects across subjects and grades is to adjust them for the average annual gains that we expect for students in each subject and grade. Then the estimates can be interpreted as an effect size relative to the amount of learning that we expect for typical students. We made this adjustment using the expected gains measures that are reported in Hill et al. (2008).¹⁹ When averaged across grades and subjects, the results suggest that a typical student with a teacher at the 85th percentile learns roughly 40 percent more than other

¹⁷ Similar distributions for science and writing in Pennsylvania are more like distributions for math than for reading. External estimates for science and writing are available only in Lipscomb et al. (2010a) for Pittsburgh. The results in that study are comparable to results in the present study.

¹⁸ High school grades have been studied to a lesser extent, in part because state assessments are not usually administered in consecutive grades. Lipscomb et al. (2010a) present exploratory findings from VAMs based on 11th-grade PSSA outcomes and teaching data from 2009–2010. Relative to effectiveness distributions in grades 6–8 based on PSSA outcomes from 2009–2010, the 11th-grade effectiveness distributions were estimated to be similarly sized in math and smaller in reading.

¹⁹ The adjustment measures are based on seven nationally normed tests. The denominators used for the adjustments are larger in math than in reading, and larger in earlier grades than in later grades.

students taught by the 50th percentile teacher in the same grade and subject during the year in terms of typical annual growth (see Appendix Table C.3). The impacts are similar across subjects in grades 4 and 5 but in grades 6–8 the impacts still tend to be larger in math than in reading.

B. Statistical Uncertainty in the Teacher Effectiveness Estimates

1. The Use of Confidence Intervals in VAMs

All performance measures are somewhat imprecise because they are based on limited information. To help quantify the precision of value-added measures, it is customary to report them together with a range of values called a confidence interval. In a hypothetical example, a teacher may place at the 45th percentile in terms of value-added in math with a confidence interval that ranges from the 40th to the 52nd percentile. Because the confidence interval includes the 50th percentile, the estimate of that teacher's effectiveness is no different statistically from average.²⁰ In reporting effectiveness, educators are treated as performing at the average unless their confidence intervals are entirely above or below the 50th percentile.²¹

The inclusion of a confidence interval is a reminder that value-added measures—like all other performance measures—are *estimates* of performance. There is debate about the size of the confidence interval that is acceptable for use in decision making, but the very reporting of confidence intervals is a distinct advantage of value-added measures over other measures for which a confidence interval is not reported. In classroom observation data, for instance, there is rarely an attempt to quantify the degree of imprecision around scores, although such an error band certainly exists. That is, if observations could be conducted many times for the same teacher in the same school year, the outcomes would probably differ based on factors such as the degree of reliability between different observers or even inclement weather that makes it difficult for students to concentrate on some days. Typically, however, only one evaluation score is obtained out of this distribution of possible scores. That score might over- or under-represent a teacher's typical performance. Value-added measures are confronted with related issues, but they can provide an indication about the degree to which a teacher's actual performance might be higher or lower than what it is estimated to be.

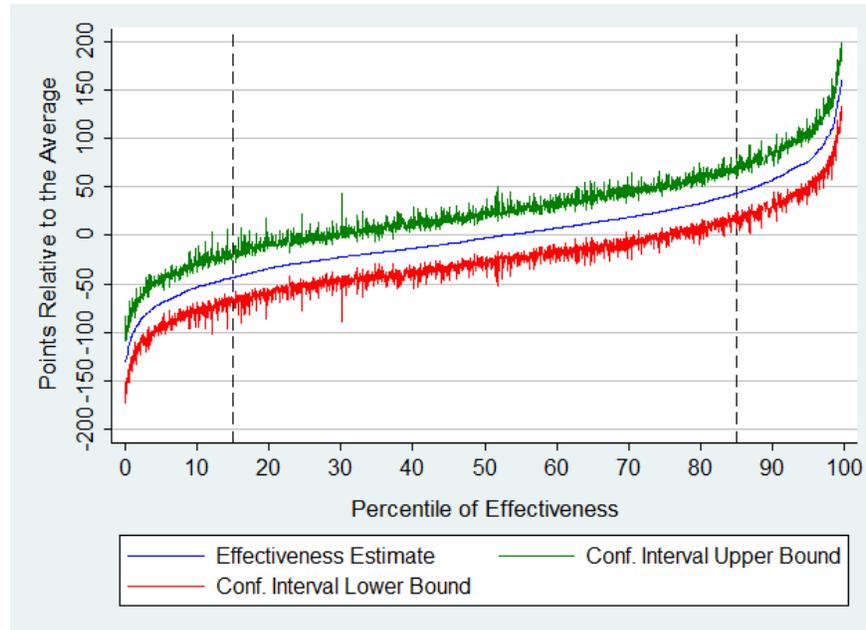
In Figures III.1 and III.2, we illustrate how confidence intervals are applied to value-added distributions in 5th-grade math and in 8th-grade reading, based on a single year of teaching for students taught between 2008–2009 and 2010–2011. We do not illustrate the distribution of effectiveness scores in 4th-grade science because it is similar. In each figure, the horizontal axis indicates percentiles of teacher effectiveness and the vertical axis indicates the additional contribution of a given teacher relative to the average teacher in PSSA scale points. The blue curve depicts the effectiveness distribution based on all Pennsylvania teachers teaching in the grade and subject. The green and red scatters above and below the blue curve represent the bounds of the confidence interval for each individual teacher estimate. Teachers with confidence intervals that

²⁰ The 50th percentile is the median value. We refer to this value as the average because we expect the median and the average teacher effectiveness estimates to be very close if not identical.

²¹ The statistical uncertainty of estimates relates directly to how much error there would be in classifying teachers or principals into performance categories on the basis of these estimates (Schochet and Chiang 2010).

include a score of zero (the score achieved by the 50th percentile teacher) cannot be distinguished statistically from average. Overall, the performance of 52 percent of teachers in 5th-grade math and 30 percent of teachers in 8th-grade reading are statistically different from average based on a single year of teaching (Appendix Table C.1). These tend to be the teachers who place at either end of the distribution. More teachers can be rated as above or below average in math because that subject has a larger distribution of estimates given the same-width confidence interval. Achieving an equal rate in reading would require even greater precision.²²

Figure III.1. Distribution of Teacher Effectiveness and 95 Percent Confidence Intervals of Teacher Effectiveness Estimates for 5th-Grade Math PSSA Scores



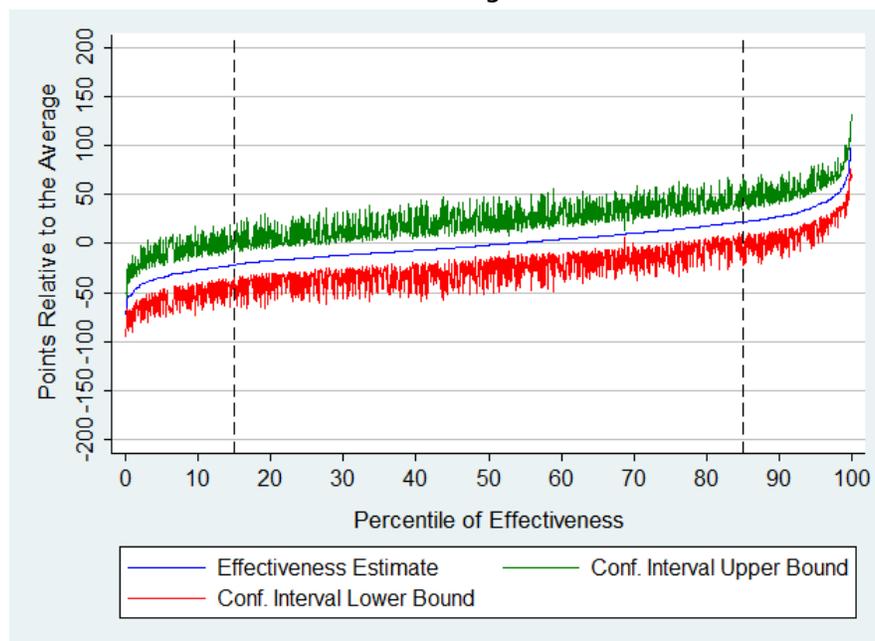
Source: Mathematica calculations based on data from PDE.

Note: Findings are based on a three-cohort model with statewide samples of teachers and students. The sample of teachers consists of those who served as teachers in every year from 2008-2009 to 2010-2011 in the outcome subject and grade.

PDE = Pennsylvania Department of Education; PSSA = Pennsylvania System of School Assessment.

²² As we discuss next, adding more years of data would likely reduce the confidence intervals. However, this change might also reduce the amount of variation in the teacher effectiveness measures so the impact on the fraction of teachers that are statistically different from zero would be ambiguous a priori.

Figure III.2. Distribution of Teacher Effectiveness Estimates and 95 Percent Confidence Intervals of Teacher Effectiveness Estimates for 8th-Grade Reading PSSA Scores



Source: Mathematica calculations based on data from PDE.

Note: Findings are based on a 3-cohort model with statewide samples of teachers and students. The sample of teachers consists of those who served as teachers in every year from 2008-09 to 2010-11 in the outcome subject and grade.

PDE = Pennsylvania Department of Education; PSSA = Pennsylvania System of School Assessment.

2. Incorporating Students from Multiple Cohorts

Value-added models vary in terms of the number of student cohorts that they include, but it is common to include multiple cohorts whenever a VAM is used for a high-stakes purpose such as a performance evaluation. By incorporating data from multiple cohorts, we can reduce the size of confidence intervals (that is, improve precision) through using data on more students. Teacher effectiveness measures based on multiple student cohorts are averages of a teacher's contributions to students taught during the years that are considered. In any single year, a teacher's students can perform unexpectedly well or poorly on an assessment for reasons other than the teacher's direct contribution. Such random fluctuations would affect the teacher's effectiveness rating for that year. To the extent that these random fluctuations tend to average out over time, the multicohort VAM provides a more reliable measure of performance. Averaging effectiveness measures across multiple cohorts also can be advantageous for reducing the effects of systematic fluctuations in scores. As mentioned earlier, Koedell and Betts (2011) found that a three-cohort VAM can reduce the potential bias in teacher effectiveness estimates that is due to nonrandom assignments of students in teachers' classrooms to statistical insignificance. Finally, multicohort VAMs can also better distinguish teacher effects from the effects of students' peers in the classroom, which cannot be separately identified in a single-cohort model unless teachers teach in multiple classrooms during the year. For all of these reasons, our primary VAMs—including the three presented in the prior section—incorporate the three most recent student cohorts or up to three if fewer are available.

The decision to incorporate data from multiple student cohorts comes with tradeoffs in terms of not reflecting immediate past performance and yielding fewer teachers with estimates based on the full number of cohorts. By definition, a three-cohort VAM evaluates performance over a longer period than a one-cohort VAM. A three-cohort VAM will thus apply to fewer teachers if policymakers decide not to report estimates for teachers when they have data from only one or two prior cohorts. In Table III.3, we illustrate this tradeoff for the selected outcomes discussed here. The main columns show the number of teachers with estimates based on the full number of cohorts in each specification (that is, one or three), and the percentage of those estimates that are statistically different from average.

Table III.3. Number of Teachers with Effectiveness Estimates Reported and Share of Reported Estimates that Are Statistically Different from the Average, by Number of Cohorts Used in Estimation

Outcome	Number of Teachers with Estimates Reported		Percentage of Reported Estimates that Are Statistically Significant from Average	
	1-Cohort Model	3-Cohort Model	1-Cohort Model	3-Cohort Model
Math PSSA, Grade 5	4,103	2,836	36.5	52.0
Reading PSSA, Grade 8	1,916	1,717	22.3	30.5
Science PSSA, Grade 4	4,187	2,854	27.7	49.8

Source: Mathematica calculations reported in Appendix Tables B.4 and C.1 based on data from PDE.

Note: Findings are based on statewide samples of teachers and students and a 95 percent confidence interval. The one-cohort model includes teachers with students in the outcome subject and grade in 2010-2011. The three-cohort model includes teachers with students in every year from 2008-2009 to 2010-2011 in the outcome subject and grade.

PDE = Pennsylvania Department of Education; PSSA = Pennsylvania System of School Assessment.

Moving from a one-cohort VAM to a three-cohort VAM decreases the number of teachers with effectiveness measures that are based on the full period that is considered but improves the precision of those estimates that are reported. For instance, the percentage of teachers with estimates reported in 5th-grade math declines by 30 percent but the share of them that are statistically significant increases by 40 percent. The gain in precision is not an artifact of using a different sample of teachers across specifications, because the percentages of statistically significant one-cohort VAM estimates among teachers who also have a three-cohort VAM estimate are nearly identical to the values reported in the table.

Adding additional cohorts of student data leads to a relatively larger reduction in the number of multicohort estimates based on the full panel of student cohorts for teachers in elementary grades than in middle school grades. This could at least partly be an implication of a requirement that we imposed whereby teachers have to be teaching students who take a particular subject and grade level assessment (for example, 5th-grade math PSSA) in *each* of the three prior school years. That is, elementary teachers who changed grade levels within the past three years would be excluded.²³ Fewer middle school teachers would be affected by changes to their teaching assignments to the extent that such changes affect the number of classes in which a teacher instructs students in the

²³ We would not actually have to limit the sample this way if the goal of the analysis was to calculate an overall value-added estimate for a teacher in a subject. In that case, we could require that a teacher have data from three prior years across grades and then calculate a composite estimate for that teacher.

subject and grade more than whether the teacher instructs any students at all in the subject and grade.

A compromise strategy (not considered for this report) could be to use three cohorts of student data in the VAM for a particular subject and grade level, and then report all estimates that apply to teachers with students in that subject and grade during the most recent year, regardless of whether they have been teaching that subject and grade for one, two, or three prior years. Under this approach, more teachers would have VAM estimates but the individual estimates would vary in terms of the number of student cohorts they include. This could be an attractive option if Pennsylvania wants to use value-added for teacher evaluations only when three years of data are available but also wants to provide value-added information to all teachers for diagnostic or professional development purposes. Decisions about the number of student cohorts to include in a VAM should be based foremost on the intended purpose of the VAM.

C. Sensitivity of Measured Effectiveness to Alternative VAM Specifications

The teacher effectiveness measures presented in this report depend on several model design elements in addition to choices about which outcomes, baselines, and student cohorts to include. We constructed the VAMs based on models that appear in the research literature but their features should be examined closely to ensure that they align with Pennsylvania’s policy preferences. In this section, we explore the sensitivity of teacher effectiveness estimates to three alternative specifications of the VAM to illustrate the types of decisions that policymakers must consider in constructing an effectiveness measure. First, we adjust estimates for factors that might vary at the district or school levels. Second, we assess the likely performance of VAMs when a same-subject baseline score is not available, as in science. Third, we compare the impacts on teacher effectiveness estimates of controlling for prior achievement using a beginning-year score versus an end-of-year score from the prior grade. We find that effectiveness estimates from the primary model might not be highly sensitive to alternative specifications for most teachers although alternative specifications do affect the effectiveness estimate for some teachers.

1. Adjusting Measured Effectiveness for District or School Factors

The distinguishing feature of a VAM is its emphasis on separately identifying the individual contributions of educators to the achievement growth of their students. Some analysts have included school-specific indicators (that is, dummy variables)—in addition to variables measured at the student and/or classroom levels—to control for factors such as working conditions at the school that might affect both student performance and a teacher’s ability to be effective in the classroom. When school-level indicators are excluded from VAMs (or similarly, when district-level indicators are excluded), teacher effectiveness measures incorporate any effect that schools have on student growth. This means that teachers at good schools (that is, those that improve student achievement more than others because of factors beyond the control of the teachers) will have an advantage in the sense that their estimated effects will be higher than similar teachers who teach at lower-quality schools. Adding school or district indicators factors out these across-school or across-district differences.

In a statewide evaluation system, however, including these indicator variables might not be desirable because the effectiveness estimates then implicitly compare teachers directly with other teachers in the same school or in the same district rather than with other teachers in Pennsylvania. An above-average teacher who does not perform quite as well as his or her colleagues at a very high-

performing school could actually be estimated to be below average in a model with school indicators. This would be especially undesirable if part of the reason a school performs well is the positive effect generated by having many good teachers. On the other hand, this could incentivize good teachers to move to bad schools and thereby promote equity. Factoring out average teacher impacts in a school can therefore lead to underestimating the teacher's influence if the average effect is simply due to the clustering of good (or bad) teachers rather than to a distinct school influence.²⁴ It also could undermine efforts to promote teamwork within a school because teacher effects would be measured only relative to other teachers in the same school or district. For these reasons, we do not include district or school indicators in our primary VAMs.

With controls for prior scores, student-level background characteristics, classroom-level characteristics, and teachers already added, it is possible that districts and schools do not have a substantial additional impact on student achievement. We can get an idea of the magnitude of the impact of omitting district and school effects in teacher VAMs by examining the change in standard deviation of estimated teacher effects after subtracting the within-district and within-school average effects from each estimate. We performed this analysis for math and reading by combining the teacher estimates from the three-cohort VAMs in grades 4 through 8, and then by subtracting the within-district or within-school average teacher effect from each individual teacher effect. Because the individual effectiveness distributions had different standard deviations, we first standardized them to a value of one before combining teachers across grades 4 through 8.²⁵ The adjusted distributions, summarized in Table III.4, consist of estimates that compare teachers with the average teacher in their district or school.

Table III.4. Implied Percentage of Variation in Teacher Value-Added Within Districts and Schools

Outcome	85th Minus 50th Percentile of VAM Estimates (in z-score units)			Implied Percentage of Total Variation in Teacher Value Added that Is Within Districts	Implied Percentage of Total Variation in Teacher Value Added that Is Within Schools
	Primary VAMs	Adjusted for Districts	Adjusted for Schools		
Math PSSA, Grades 4-8	1.00	0.91	0.77	83	59
Reading PSSA, Grades 4-8	1.00	0.93	0.81	87	65

Source: Mathematica calculations based on Pennsylvania data.

Note: Findings are based on three-cohort teacher VAM estimates for grades 4-8 that are reported in Appendix Table C.1. The implied percentage columns are calculated as the ratio of the square of each "adjusted" column value to the square of the corresponding primary VAM value. A z-score unit is a standard deviation of student scores.

PDE = Pennsylvania Department of Education; PSSA = Pennsylvania System of School Assessment.

²⁴ The contribution of principals is another school factor that might be confounded with effectiveness estimates for teachers. If teachers with high VAM scores simply serve at schools led by effective principals, the teacher's contribution might be less than what is measured by a model that excludes school indicators. However, because effective principals might recruit effective teachers, it is not clear that this variation should be removed from teacher effectiveness estimates.

²⁵ For teachers with students in multiple grades, their standardized grade-specific estimates were then averaged so that all teachers in grades 4 through 8 had one estimate for each subject. The district or school adjustment factor for an individual teacher is the average value-added across all teachers in the districts or schools where he or she teaches.

Most of the overall variation in Pennsylvania teacher effectiveness estimates in these grades and subjects is within individual schools rather than across them—a finding shared by many studies in the research literature. Most of the remaining variation is across schools within individual districts. The smallest portion of variation in Pennsylvania teacher effects is across districts. Specifically, we found that about 62 percent of the variation in the teacher effectiveness estimates is within schools, 23 percent is across schools within districts, and 15 percent is across districts. This is indicated in Table III.4 by averaging across rows the degree of remaining teacher quality variation in Pennsylvania after removing the average value-added of teachers in each district or school. For example, adjusting estimates from the primary models for districts leaves about 85 percent of the variation intact (that is, 83 percent in math and 87 percent in reading). Adjusting estimates from the primary models for schools reduces the amount of variation in the teacher effectiveness measures relatively more, but still 59 percent remains in math and 65 percent remains in reading.

These findings support a conclusion that the most important factors to include in a VAM are those that vary within schools. But the findings also indicate that about 38 percent of the variation in teacher effectiveness across Pennsylvania is across schools rather than within them. Thus, we cannot rule out the possibility that adding controls for certain school- or district-level factors could improve the validity of the estimates. An alternative method (not considered in this report) to adjusting for the average value-added in districts or schools that still accounts for the impact that districts and schools have on student achievement would be to include district- or school-level observable characteristics in the VAM. Examples of such characteristics could be the fraction of students eligible for free meals, the average years of experience among teachers, or the level of district funding per student. We did not pursue this approach for this report because we were concerned that the VAMs might not produce valid estimates for the relationships between school or district characteristics and outcomes. To estimate these relationships, the VAMs would have to rely only on year-to-year variation in the characteristics of the same school or district, which is much smaller and more transitory than the variation of interest across different schools and districts. Indeed, in preliminary analyses, we found that several coefficient estimates on student demographic variables measured at the school level had counterintuitive signs, which suggests that the VAMs might not produce valid estimates of these coefficients. We recommend that the Pennsylvania Department of Education (PDE) consider whether to control for factors that vary across districts or schools during Phase 2 of the pilot study and, if so, how best to do so.

2. Excluding a Prior Achievement Score from the Same Subject

Given the importance of students' own achievement histories in predicting current achievement, analysts seek to control for a prior assessment in which a student has a score in the same subject. Including a same-subject prior score is desirable but it is not a requirement for a VAM to operate because VAMs simply make a prediction about students' scores based on the factors that are controlled, whether they come from the same or different subjects. Because students are not always assessed in consecutive grades across subjects, a practical implication of extending the use of value-added broadly to teachers is that the model in some grades and subjects will not be able to incorporate a same-subject prior score.

An example comes from 4th-grade science. The VAM for 4th-grade science can only control for the incoming science abilities of students to the extent that they are related to prior achievement scores in math and reading, the two subjects tested by the PSSA in 3rd grade. Short of introducing a new 3rd-grade science assessment statewide, policymakers and analysts are left to decide whether to use the 4th-grade science VAM with available data or disregard it altogether. In the following

analysis, we infer the likely performance of the 4th-grade science VAM by simulating the impact on teacher effectiveness measures of intentionally omitting a same-subject prior score from VAMs for 5th-grade math and reading. We chose 5th grade for this diagnostic because 5th graders have three available scores from 4th grade: math, reading, and science. Specifically, we compared the math and reading teacher estimates obtained through VAMs that control for 4th-grade math and reading scores with estimates obtained by replacing the same-subject score with the 4th-grade science score. The rationale is that the former specification is what analysts and policymakers would like to estimate but the latter is equivalent to what can be estimated for 4th-grade science using available data.²⁶

We show results from the exercise in Table III.5, which aggregates estimates for math and reading for presentation purposes. The table rows indicate teacher effectiveness quartiles from the specification that controls for math and reading scores in 4th grade. The columns indicate effectiveness quartiles from the specification that replaces the same-subject prior score (that is, 4th-grade math or reading, depending on the outcome) with the 4th-grade science score. The values indicate the number and percentage of teacher estimates in each cell. Teachers are included in the analysis only if they have a VAM estimate under both specifications, but most included teachers are represented twice because 5th-grade teachers typically teach students in both subjects.

Table III.5. Counts and Percentages of Grade 5 Math and Reading Teachers in Effectiveness Quartiles Based on 3-Cohort Teacher VAMs that Include and Exclude Same-Subject Baseline Scores

	Quartile of Effectiveness Based on Teacher VAM that Controls for the Grade 4 Science Score Instead of the Same-Subject Baseline Score				Total
	1st (bottom)	2nd	3rd	4th (top)	
Quartile of Effectiveness Based on Teacher VAMs with Controls for Grade 4 Math and Reading Scores					
1st (bottom)	1,085 (74.8)	310 (21.4)	54 (3.7)	1 (0.1)	1,450 (100.0)
2nd	314 (21.7)	727 (50.2)	375 (25.9)	32 (2.2)	1,448 (100.0)
3rd	49 (3.4)	363 (25.1)	758 (52.3)	279 (19.3)	1,449 (100.0)
4th (top)	2 (0.1)	48 (3.3)	262 (18.1)	1,136 (78.5)	1,448 (100.0)
Total	1,450 (25.0)	1,448 (25.0)	1,449 (25.0)	1,448 (25.0)	5,795 (100.0)

Source: Mathematica calculations based on Pennsylvania data.

Note: In each table cell, the first value is the number of teachers in the given cell, and the second value (in parentheses) is the percentage of the row total that is represented by that cell.

The teacher estimates show a relatively high degree of correlation across specifications, with most estimates falling on the table's diagonal elements. For instance, the top-left cell indicates that

²⁶ For the diagnostic purpose of these analyses, we did not include controls for 3d-grade scores.

75 percent of the teachers with the highest 25 percent of effectiveness scores under the primary model that includes prior math and reading scores had effectiveness scores in the top 25 percent under the alternative model too. Of the teachers whose quartile position changes, nearly all of them move by just one quartile. Only three of nearly 5,800 teacher estimates, or 0.05 percent, move from the top quartile to the bottom quartile. The within-teacher correlations across specifications are 0.88 in math and 0.91 in reading.

Based on these results for 5th grade, we expect that a hypothetical 4th-grade science VAM that controlled for students' science achievement in 3rd grade would produce estimates that are relatively highly correlated to the estimates that can be obtained currently. That some estimates are off the diagonal elements indicates that the presence or absence of a same-subject control affects the placement of some individual teachers in the distribution of effectiveness. In addition, the models with the same-subject baselines explain a greater portion of the overall variation in student scores. Specifically, the adjusted r-squared value is 3 percentage points higher in reading (an increase from 0.66 to 0.69) and 11 percentage points higher in math (an increase from 0.64 to 0.75) in the VAM specification that includes a same-subject control. On balance, we believe the evidence does not support discarding VAMs altogether when a same-subject baseline is unavailable, but presumably the accuracy and precision of the measures would improve if a same-subject prior score were available.

3. Controlling for Students' Prior Achievement Histories with a Fall or Spring Score

Because assessments are typically administered in the spring, most VAMs control for students' prior achievement histories using scores obtained at the end of the prior grade. Teacher effectiveness estimates using this approach therefore incorporate the effects of students' summer experiences, which can confound estimates of teacher contributions during the academic year. Measuring a student's achievement growth by testing at the beginning and near the end of the school year might produce a better attribution of learning to the teacher. But it introduces several new concerns as well, because schools would have to increase the time and resources devoted to testing and some teachers might deemphasize the fall assessment to produce larger gains.

Although there are concerns with both approaches, we sought to examine whether they nevertheless produce similar measures of teacher effectiveness given the currently available data.²⁷ In Table III.6, we compare teacher effectiveness quartiles generated from VAMs that differ by whether they control for math and reading 4Sight assessment scores using fall or spring scores. Because we can only access 4Sight scores in the pilot districts, the analysis is therefore limited to students in Allentown, Cornell, and Mohawk during the 2010–2011 year in grades 4 through 8. Each VAM specification included controls for PSSA math and reading scores from the prior grade but not from two prior grades, in addition to the math and reading 4Sight assessment scores. To maximize the teacher sample, we standardized the individual grade and subject effectiveness distributions and then combined teacher estimates across grades 4 through 8 as in Table III.4. We included teachers only if they had an effectiveness estimate under both specifications, although teachers are represented more than once if they teach students in math and reading.

²⁷ If teachers are given incentives to perform well based on these measures then the results might change.

Table III.6. Grade 4 Through 8 Math and Reading Teachers in Pilot Districts in Effectiveness Quartiles Based on Teacher VAMs with Fall or Spring Baselines Using PSSA Outcome Data

	Quartile of Effectiveness Based on Teacher VAM with Beginning-Year Fall 4Sight Baselines				Total
	1st (bottom)	2nd	3rd	4th (top)	
Quartile of Effectiveness Based on Teacher VAMs with Prior-Grade Spring 4Sight Baselines					
1st (bottom)	66 (77.6)	19 (22.4)	0 (0.0)	0 (0.0)	85 (100.0)
2nd	15 (17.6)	53 (62.4)	17 (20.0)	0 (0.0)	85 (100.0)
3rd	4 (4.7)	13 (15.3)	54 (63.5)	14 (16.5)	85 (100.0)
4th (top)	0 (0.0)	0 (0.0)	14 (16.5)	71 (83.5)	85 (100.0)
Total	85 (25.0)	85 (25.0)	85 (25.0)	85 (25.0)	340 (100.0)

Source: Mathematica calculations based on data from Pennsylvania and pilot districts' records.

Note: In each table cell, the first value is the number of teachers in the given cell and the second value (in parentheses) is the percentage of the row total that is represented by that cell.

PSSA = Pennsylvania System of School Assessment; VAM = value-added model.

The findings from this analysis support a conclusion that the timing of baseline controls might not make a large difference in the effectiveness measure for most teachers in these three districts. As in Table III.5, most of estimates are placed on the diagonal elements and nearly all the remaining estimates are off the diagonal by one quartile. The within-teacher correlations across specifications are 0.95 in math and 0.93 in reading. A comparison of the adjusted r-squared values by subject and grade across the two specifications indicates that using the fall baselines leads to a slight increase of about one percentage point in the percentage of variation in student scores that is explained by the model. On the whole, we find that fall and spring baselines produce similar teacher effectiveness estimates in Allentown, Cornell, and Mohawk, but that broader analyses should be conducted to support a general conclusion. In Phase 2 of the pilot, the expanded sample of districts should facilitate such an opportunity to conduct an analysis that is more representative of students in Pennsylvania.

D. Key Characteristics of Teacher Effectiveness Estimates Based on Pilot District Samples and Other Outcomes

More teachers can be included in value-added analyses by analyzing additional outcomes beyond PSSA scores in 4th through 8th grades. For teachers with students in grades 4 to 8, adding outcomes has the potential to make available multiple sources of information on each teacher's impact on his or her students. In Phase 1, we applied VAMs for measuring teacher effects to several additional outcomes using student samples from the Phase 1 districts in 2010–2011. First, we estimated teacher VAMs for non-PSSA outcomes in the lower elementary grades, in which the PSSA is not administered. Second, we estimated teacher VAMs based on PSSA assessments in 3rd and in 11th grades—grades that cannot be included in statewide analyses because a prior score is not available on a statewide basis. Thus, in all of the following analyses, the samples include only

students and teachers from the Phase 1 districts. This section describes key characteristics of the effectiveness estimates from these VAMs, which appear to differentiate among teachers except in 11th grade, the grade in which a larger sample of students is needed.

1. VAMs Based on Non-PSSA Assessments Administered by Pilot Districts

As we discussed in Chapter II, the pilot districts administer a number of assessments in lower elementary grades that are not covered by the PSSA. Adding these lower elementary grades to the analysis samples would substantially expand the set of teachers with value-added scores. We generated effectiveness estimates for teachers in the pilot districts based on the Progress Assessment (in Allentown) and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (in Allentown and Cornell). We analyzed the same key characteristics of these effectiveness estimates as we did for the PSSA-based estimates—namely, the extent of variation across teachers and the level of precision.

The effectiveness estimates appear to differentiate among teachers. In Table III.7, we show the estimates for teachers at the 15th and 85th percentiles. We omit the other percentiles given the smaller teacher sample sizes, but the patterns are similar to the PSSA distributions described earlier in the chapter. The 85th and 15th percentile teachers in Allentown differ in effectiveness by 9 percentage points on the writing Progress Assessment in 1st grade and by 11 percentage points on the math Progress Assessment in 2nd grade. By switching from the 15th to the 85th percentile teacher, a student originally at the median of these score distributions would be predicted to rise to the 73rd to 75th percentiles. There is less variation in teachers' impacts on 2nd-grade DIBELS scores, for which the 9-point difference in effectiveness between the 85th and 15th percentile teachers is equivalent to moving a student from the median to the 59th percentile of scores. Overall, these results (as well as results for additional assessments shown in Appendix Table C.4) generally suggest sizable variation in teachers' contributions to student scores.

Table III.7. Key Characteristics of Teacher Effectiveness Estimates Based on Selected Non-PSSA Tests Administered in the Pilot Districts

Outcome	Effectiveness of the Teacher at the Indicated Percentile Relative to the Effectiveness of the Average Teacher (in test scale points)		Percentage of Single-Cohort Teacher Effectiveness Estimates that Are Statistically Distinguishable from the Average
	15th	85th	
Progress Assessment, Writing, Grade 1 (Percentage Points) ^a	-5	4	38.0
Progress Assessment, Math, Grade 2 (Percentage Points) ^a	-5	6	34.8
DIBELS, ORF, Grade 2 ^b	-5	4	18.2

Source: Mathematica calculations based on data from Pennsylvania and pilot districts' records.

Note: Findings are based on a 95 percent confidence interval and a one-cohort model with samples of teachers and students from the pilot districts in 2010-2011.

^aAllentown only.

^bAllentown and Cornell only.

ORF = oral reading fluency; PSSA = Pennsylvania System of School Assessment.

We caution, however, that analyzing the variation in teacher effects by the existing dispersion in student scores, as we have done in the preceding discussion, does not fully gauge whether this

variation is educationally meaningful. A closer examination of the content validity of these assessments—which is beyond the scope of our analysis—is necessary for determining whether this variation translates into substantive differences in students’ skills and knowledge and, thus, is suitable for informing questions about educators’ impacts on student performance.

The ability of the teacher VAM to make statistically reliable distinctions among teachers differs by outcome measure. For estimates based on the Progress Assessments, one-third of teachers can be statistically distinguished from average effectiveness. However, only 18 percent of teachers can be distinguished from the average based on their impacts on 2nd-grade DIBELS scores. Although the precision of the DIBELS-based estimates is not worse than that of the Progress-based estimates, there is less overall variation in teachers’ impacts on 2nd-grade DIBELS scores. Therefore, a greater *proportion* of the variation in the DIBELS-based effectiveness estimates is due to random fluctuations and other sources of imprecision, making it more difficult to identify high- and low-performing teachers with high degrees of confidence. These differences in reliability across different types of effectiveness estimates could be factors in determining what weights PDE would like to place on these estimates in evaluating teachers.

2. VAMs Based on PSSA Assessments in 3rd and 11th Grades

We also expanded the grade-level coverage of teacher VAM estimates by using fall 4Sight scores as baseline achievement measures for PSSA outcomes in 3rd and 11th grades. In Table III.8, we show the characteristics of the resulting estimates.

In grade 3, as in grades 4 to 8, teachers’ impacts on PSSA scores vary sizably. We find that the 15th and the 85th percentile teacher differ in effectiveness by 110 scale points in 3rd-grade math and by 55 scale points in 3rd-grade reading. Consistent with findings for higher grades on a statewide basis, 3rd-grade teacher impacts in Allentown, Cornell, and Mohawk are larger in math than in reading. Moreover, a larger percentage of teacher estimates can be distinguished from the average teacher in math than in reading, although the percentage in each subject is similar to those for one-cohort teacher VAMs statewide in grades 4 to 8. Overall, we interpret the results for the 3rd-grade VAMs as indicating that using fall 4Sight scores (or another measure, if available, that is aligned with PSSA content) as a baseline measure might be viable from an attribution standpoint. As indicated previously, there are potential resource and incentive-compatibility concerns involved with using a fall baseline in a teacher VAM that Pennsylvania policymakers should first consider carefully.

Table III.8. Key Characteristics of Teacher Effectiveness Estimates Based on PSSAs in 3rd and 11th Grades in the Pilot Districts

Outcome	Effectiveness of the Teacher at the Indicated Percentile Relative to the Effectiveness of the Average Teacher (in test scale points)		Percentage of Teacher Effectiveness Estimates that Are Statistically Distinguishable from the Average
	15th	85th	
PSSA, Math, Grade 3	-55	55	43.5
PSSA, Reading, Grade 3	-29	26	18.8
PSSA, Math, Grade 11	-21	17	3.4
PSSA, Reading, Grade 11	-17	12	0.0
PSSA, Writing, Grade 11	-19	24	0.0
PSSA, Science, Grade 11	-3	3	0.0

Source: Mathematica calculations based on data from Pennsylvania and pilot districts' records.

Note: Findings are based on a 95 percent confidence interval and a one-cohort model with samples of teachers and students from the pilot districts. The sample of teachers consists of those who served as teachers in 2010-2011 in Allentown, Cornell, or Mohawk school districts.

In contrast with the results from the 3rd-grade VAMs, the 11th-grade VAMs are not able to make reliable distinctions among teachers. The distributions of effectiveness for 11th-grade VAMS are more compressed than in lower grades, as indicated by a smaller difference between the effectiveness of the 15th and the 85th percentile teacher. The estimated teacher effects also have more imprecision than in the other models and, consequently, cannot distinguish even very high or low teacher contributions from the average.²⁸ The estimates are “shrunk,” or pulled, more heavily toward the average to account for their greater imprecision (see Appendix A for a description of the shrinkage process). Part of the explanation for low precision is probably a lesser degree of alignment between the assessments and 11th-grade courses, which might be less likely to focus extensively on the skills measured in the PSSA. Keystone exams presumably will be more appropriate for teacher value-added use because they will be better aligned. As is, the results from Phase 1 for the 11th-grade VAMs would not be viable for use in an actual evaluation model. Precision might improve in Phase 2 when a much larger student sample can be included, but we would not expect substantial precision gains if the low precision is due to poor alignment between the content of the assessments and 11th-grade courses.

²⁸ We include students' 8th-grade PSSA scores as additional control variables in the 11th-grade teacher VAM along with fall 4Sight scores from grade 11. The inclusion of the 8th-grade scores, meant to enhance the controls for students' prior achievement histories, could nevertheless decrease precision if a large number of 11th-grade students are missing 8th-grade scores and thus have to be dropped from the analysis sample. We estimated an alternative specification of the 11th-grade VAM that omitted the 8th-grade scores. The alternative models included approximately 19 percent more students (that is, from about 715 to 850) but did not improve the ability of the VAM to distinguish between teachers and led to a reduction in the model r-squared value.

IV. RELATIONSHIPS BETWEEN TEACHER PRACTICES AND VALUE ADDED

In this chapter, we describe the analyses we conducted to examine relationships between value-added and teacher practices, as measured by the Phase 1 teacher observation rubric. We begin by summarizing the observation data obtained and the characteristics of the teachers who participated during Phase 1. Collectively, Phase 1 teachers are not dramatically different from teachers across Pennsylvania in terms of their demographics, master’s degree attainment, and level of teaching experience. We then examine the variation in observation scores, which indicate that nearly all Phase 1 teachers were rated as either proficient or distinguished by their principals on the pilot rubric. Finally, we estimate the change in teacher value-added that is associated with a one-level increase in a teacher’s score on different rubric components. Due to the small size of the pilot and a compressed distribution of observation scores, none of the resulting correlations are statistically significant, although most are numerically positive. We expect that the Phase 2 data will yield considerably more precision to these analyses.

A. The Phase 1 Teacher Observation Rubric and Score Distribution

The Pennsylvania teacher evaluation rubric administered to teachers in Phase 1 was based on the Framework for Teaching developed by Charlotte Danielson. The Danielson Framework includes 22 components grouped into four domains—planning and preparation, the classroom environment, instruction, and professional responsibilities. For Phase 1 of the Pennsylvania pilot, the stakeholder group focused on the 11 priority components in Table IV.1, consistent with similar work to improve teacher effectiveness in Pittsburgh Public Schools.²⁹

Table IV.1. Danielson Framework Domains and Components, by Priority and Additional Components for the Pennsylvania Pilot

Priority Components	Additional Components
Domain 1: Planning and Preparation	
1c: Setting instructional outcomes	1a: Demonstrating knowledge of content and pedagogy
1e: Designing coherent instruction	1b: Demonstrating knowledge of students
1f: Designing assessment outcomes	1d: Demonstrating knowledge of resources
Domain 2: The Classroom Environment	
2b: Establishing a culture for learning	2a: Creating an environment of respect and rapport
2d: Managing student behavior	2c: Managing classroom procedures
Domain 3: Instruction	
3b: Using questioning and discussion techniques	3a: Communicating with students
3c: Engaging students in learning	3e: Demonstrating flexibility and responsiveness
3d: Using assessment in instruction	
Domain 4: Professional Responsibilities	
4a: Reflecting on teaching and student learning	4d: Participating in a professional community
4b: System for managing students’ data	4e: Growing and developing professionally
4c: Communicating with families	4f: Showing professionalism

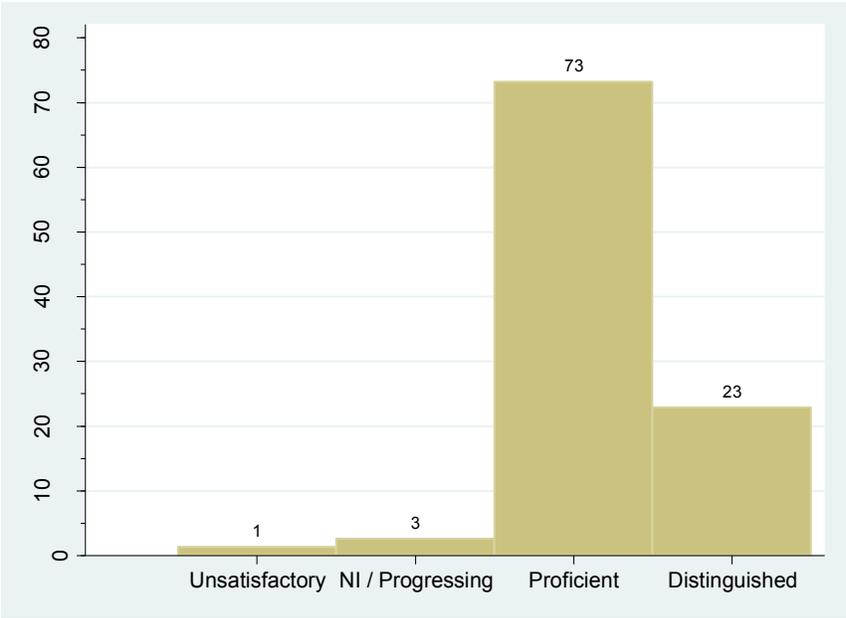
Source: Pennsylvania Teacher Evaluation Rubric from Phase 1.

²⁹ The Empowering Effective Teachers (EET) program is a joint project between Pittsburgh Public Schools and the Pittsburgh Federation of Teachers. Like Phase 1 of the Pennsylvania pilot, the EET program receives funding through the Bill & Melinda Gates Foundation.

Teachers were rated on a scale from one to four on each priority component (that is, 1 = Unsatisfactory; 2 = Needs Improvement for tenured teachers or Progressing for nontenured teachers; 3 = Proficient; and 4 = Distinguished). On the rubric, principals could also provide a short text description of the evidence on priority and additional components, although the narrative data are not used in this analysis. To help principals apply a consistent standard to their evaluations, the rubric included a component-by-component description of each level of performance. The evaluation matrix calculated an average rating across priority components and domains and a final rating that rounded the average rating to the nearest whole number.

In Figure IV.1, we show the distribution of final rating scores for the 153 Phase 1 teachers. *Among Phase 1 teachers, 96 percent were rated as proficient or distinguished. One percent of these teachers were rated as unsatisfactory—the same percentage that the Pennsylvania Department of Education (PDE) found across the state for 2009–2010 under the existing observation protocol that differentiates only between satisfactory and unsatisfactory performance.* Some principals rated every single teacher as proficient on all components. To the extent that more performance categories would be desirable for policy purposes, at the low end of performance the pilot rubric differentiated among teachers only slightly more than the current evaluation form does. The distribution of observation scores is more heavily concentrated in the proficient and distinguished categories than would be expected from applying typical results from the one-cohort value-added models (VAMs) at the precision levels reported in Table III.3. Specifically, approximately 30 percent of teachers in those models could be distinguished as above or below average. Depending on rubric rating definitions, we might therefore have expected a distribution in which about 15 percent of teachers are distinguished, 70 percent are proficient, and 15 percent are unsatisfactory or in need of improvement/progressing.

Figure IV.1. Distribution of Final Rating Scores for Phase 1 Teachers



Source: Observation data collected on Phase 1 teachers.

Notes: The data includes final rating scores on all 153 Phase 1 teachers from the four pilot districts.

NI = Needs Improvement.

The Consortium on Chicago School Research found a wider distribution of observation scores in a recent report documenting findings from a 2008 pilot study (Sartain et al. 2011). Like Pennsylvania’s pilot, the Teacher Evaluation Pilot in Chicago implemented an observation protocol based on the Danielson Framework, although the two pilots might have adapted the Framework to their own needs. The distribution of scores, contrasted with Phase 1 scores in Table IV.2, indicates that principals in Chicago rated a larger percentage of teachers at the basic level than did principals in Phase 1, and a smaller percentage of teachers as proficient or distinguished.³⁰ Unlike the Pennsylvania pilot, classroom observations in the Chicago pilot included an external observer; observers were practitioners with extensive and ongoing training in the Danielson Framework for Teaching. Chicago’s observing teachers were more conservative than principals in their ratings, with only 3 percent of teachers reaching the distinguished level.

Table IV.2. Final Ratings in Pennsylvania and Chicago, by Number and Percentage of Teachers

Characteristic	Pilot Teachers in Phase 1 Rated by Principal	Chicago Teachers	
		Rated by Principal	Rated by Observer
Distinguished	23	17	3
Proficient	73	53	67
Needs Improvement/Progressing (Basic in Chicago)	3	27	28
Unsatisfactory	1	3	2

Source: Ratings for Phase 1 teachers come from the pilot observation data. Ratings for Chicago teachers are reported in Sartain et al. (2011) Table 3.

Notes: The columns are based on data on 153, 4,747, and 4,852 ratings, respectively. Second observers were one of three individuals who were highly trained in the Danielson Framework for Teaching.

With data on only 153 Pennsylvania teachers, it is highly uncertain whether the distribution of scores obtained in Phase 1 is representative of scores that would be obtained by teachers across Pennsylvania in a larger pilot. However, we can glean from the data that Phase 1 teachers are not dramatically different from other Pennsylvania educators in terms of several broad demographic and professional characteristics. In Table IV.3, we compare teachers who participated in Phase 1 with all other educators in Pennsylvania based on their gender, race/ethnicity, educational attainment, and total years of experience.³¹ The samples are not statistically different by gender and master’s degree attainment. Relative to educators across Pennsylvania, a higher percentage of Phase 1 teachers were white and the distribution of total years of experience is more concentrated in the 6-to-10-years’ category and less concentrated in the 21-or-more-years’ category. Despite the statistical significance of the latter mean differences, the values are not dramatically different. Overall, we do not see any

³⁰ The general description of basic in Chicago—understanding the components of teaching but implementing them sporadically—corresponds to the descriptions of needs improvement/progressing in the Phase 1 rubric.

³¹ The comparison group includes other Pennsylvania educators rather than only Pennsylvania teachers because in the data we cannot differentiate teachers from other staff.

clear evidence to suggest that the distribution of observation scores would be substantially different if a larger sample of teachers were observed, based on the teacher characteristics in Table IV.3.³²

Table IV.3. Sample Characteristics of Nonpilot and Pilot Teachers

	Pilot Teachers in Phase 1 Districts	Nonpilot Educators in Pennsylvania	Statistically Significant Difference
Female	72	72	No
White	98	93	Yes
Master's Degree	46	51	No
Total Experience: 0-5 Years	20	25	No
Total Experience: 6-10 Years	34	23	Yes
Total Experience: 11-15 Years	20	19	No
Total Experience: 16-20 Years	12	12	No
Total Experience: 21 or more Years	14	21	Yes

Source: Mathematica calculations based on Pennsylvania data.

Note: The last column indicates statistically significant mean differences at the 5 percent level.

B. Observation Scores and Value-Added Scores for Phase 1 Teachers with VAM Estimates

As described in Chapter II, only 81 of the 153 teachers in the Phase 1 sample could be assigned value-added estimates. When only one assessment was available in a particular grade and subject, we applied estimates from that VAM to any respective Phase 1 teacher. When multiple assessments were available (as in 2nd-grade reading), we selected the VAM with the highest r-squared value.³³ If a teacher did not have a VAM estimate from the VAM with the highest r-squared among multiple assessments in a subject and grade, we used any VAM estimate that might be available for the teacher from the other assessments in the same subject and grade. This last step added only one additional teacher. All VAM estimates came from one-cohort models for 2010–2011 because the rubric covered teacher practices only in 2010–2011 and some outcomes were available in that year only.

There must be variation in observation scores to identify how changes in value added are associated with unit increases in observation ratings. However, the distribution of final ratings among these 81 teachers was even more skewed than in the overall sample. Thirty percent of these teachers received a distinguished rating and the remaining 70 percent received a proficient rating. None of these teachers had a final rating in the lower two categories. Given only two values for the final rating, we instead analyzed teachers' ratings on individual components and on their average score across priority components (that is, the final rating before rounding to the nearest whole number) because they are more continuous measures. We emphasize, however, that it is not clear

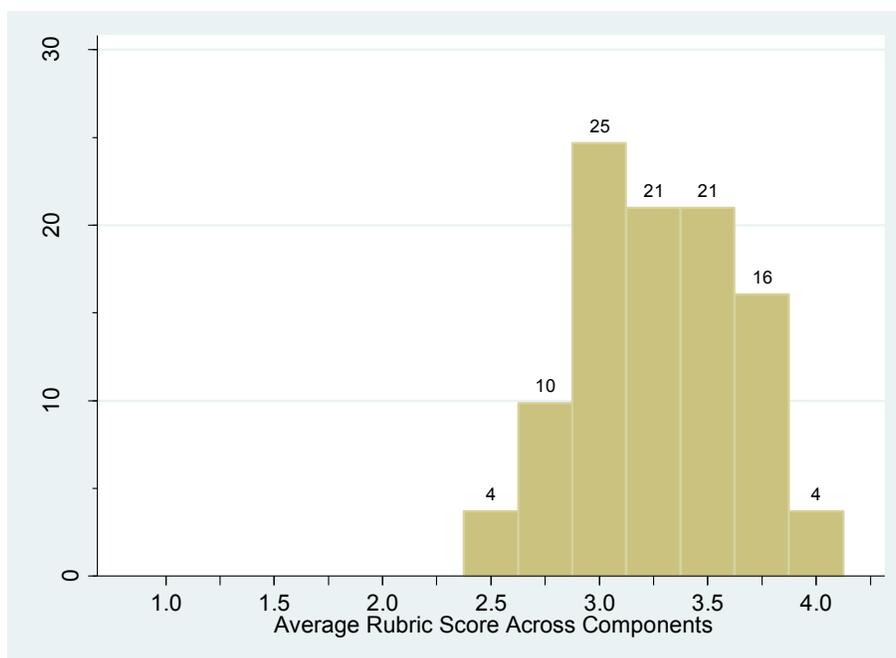
³² For instance, the percentage of new teachers (defined as having 0 to 5 years of total experience) is not statistically different across groups. If the Phase 1 sample included a relatively low proportion of new teachers, we might expect a less skewed distribution for Pennsylvania overall because new teachers learn on the job during the first five years.

³³ This assessment was Dynamic Indicators of Basic Early Literacy Skills (DIBELS) – Phoneme Segmentation Fluency in 1st-grade reading and DIBELS – Oral Reading Fluency in 2nd-grade reading.

that a score of 2.75 or 3.25, for example, is meaningfully different than a score of 3.0, because the measurement properties of the observation rubric have not been examined.

In Figure IV.2, we depict the distribution of average ratings across priority components among Phase 1 teachers with VAM estimates. The average ratings are represented on the horizontal axis from 1.0 to 4.0, grouped into categories of 0.25 for illustrative purposes. The vertical axis indicates the percentage of teachers with a given average rating across priority components.

Figure IV.2. Distribution of Average Rating Scores for Phase 1 Teachers with VAM Estimates



Source: Observation data collected on 81 Phase 1 teachers with VAM estimates.

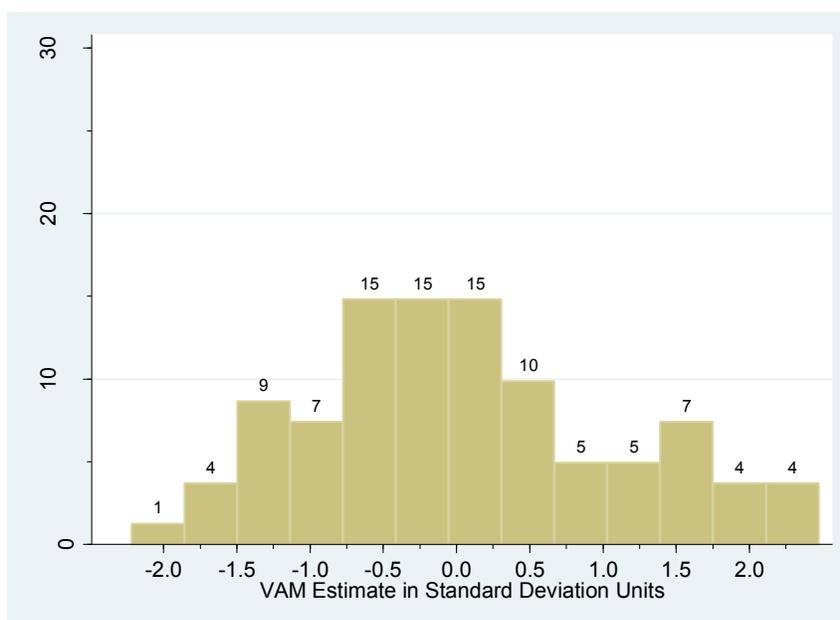
The distribution of final rating scores has fewer categories than the distribution of VAM scores for these same teachers, shown in Figure IV.3.³⁴ Because the teacher VAM scores are not measured in rubric levels, we illustrate the range of VAM estimates by separating estimates into 13 equal-sized categories to correspond with the 13 rubric rating categories that are possible with a gradation of increments of 0.25 each.³⁵ The VAM distribution spreads out more than the distribution of rubric scores, with teachers placing in each of the 13 categories. The VAM distribution also looks more like a traditional bell curve centered on the average possible value, unlike rubric scores that are concentrated in the upper range of possible values. Given the compressed distribution of observation scores and the small sample, we do not expect to see any statistically significant relationships with value-added estimates unless very small differences in observation scores

³⁴ To obtain the VAM score distribution in Figure IV.3, we standardized individual estimates in Allentown and Mohawk (that is, the two districts in which Phase 1 teachers have VAM estimates) by grade and subject for each assessment before assigning them to pilot teachers. Thus, the VAM scores across teachers come from equivalent distributions.

³⁵ VAM estimates in Figure IV.3 are measured in standard deviation units. See Chapter III for a description of standard deviation units.

meaningfully differentiate between teachers in terms of their contributions to student achievement growth.

Figure IV.3. Distribution of VAM Scores for Phase 1 Teachers with VAM Estimates



Source: Mathematica calculations based on Pennsylvania data.

Note: Before collecting scores from across assessment VAMs, we standardized each estimate distribution to have a zero mean and a standard deviation of one among Allentown and Mohawk teachers.

C. Relationships Between Value-Added and Observation Scores

Using statistical models, we tested the relationships between teachers' estimated contributions to student learning and their observation scores for the 81 teachers with both types of effectiveness measures. The models compared the VAM score for individual teachers—as depicted in the prior figure—with their rubric ratings, holding constant average differences in teachers' VAM scores across districts, subjects, grade levels, and assessments. We conducted separate analyses for each priority component and domain-level average, and for the overall average rating across components.

The findings are interpreted as the predicted increase in teacher contributions to student learning from a one-level increase on the observation rubric. Expressing relationships between value-added and observation scores in this way is likely to be more informative for policymakers than as a correlation coefficient because the magnitude of the relationship is expressed in terms of student learning. Larger magnitudes indicate larger gains in student achievement for a one-level increase on a rubric component. As in a VAM, the statistical model can also indicate whether a particular relationship is statistically different from zero. By holding constant the variation across teachers in their districts, subjects, grade levels, and assessments, this model is likely to yield estimated relationships that are more accurate than a simple correlation coefficient. Because the VAM estimates (the outcome variable) have been standardized by grade and subject, comparisons of teachers can be legitimately made only within grade and subject. Therefore, it is necessary to force comparisons of the rubric score (the independent variable) to be made only within grade and subject.

In Table IV.4, we report the estimated relationships for all priority components and domain-level averages and for the overall average across priority components. None of the relationships are statistically significant, most likely due to the small sample size in Phase 1. The strongest relationships (even though not significant) are in the instruction domain, which is encouraging for future work. If a hypothetical estimate of 0.4 were statistically significant, it would mean that a one-level increase in a component score is associated with a 0.4 standard deviation increase in teacher effectiveness as measured by value-added.³⁶ An increase of this magnitude is equivalent to the additional contribution of a teacher at the 65th percentile of VAM scores above the contribution of a teacher at the 50th percentile. A student with median test scores would be expected to perform at the 53rd percentile if taught by the former rather than the latter teacher.

Table IV.4. Regression Coefficients Indicating the Standard Deviation Increase in Teacher Value-Added that Is Predicted for a One-Unit Increase in Rubric Scores

Domain or Component	Name	Estimate	Standard Error	Statistically Significant
Domain 1 Avg.	Planning and Preparation	0.04	0.29	No
1c	Setting instructional outcomes	0.13	0.27	No
1e	Designing coherent instruction	0.09	0.26	No
1f	Designing assessment outcomes	0.03	0.24	No
Domain 2 Avg.	The Classroom Environment	-0.06	0.28	No
2b	Establishing a culture for learning	-0.11	0.25	No
2d	Managing student behavior	0.03	0.30	No
Domain 3 Avg.	Instruction	0.44	0.32	No
3b	Using questioning and discussion techniques	0.39	0.25	No
3c	Engaging students in learning	0.35	0.30	No
3d	Using assessment in instruction	0.15	0.27	No
Domain 4 Avg.	Professional Responsibilities	0.01	0.31	No
4a	Reflecting on teaching and student learning	-0.07	0.34	No
4b	System for managing students' data	-0.29	0.26	No
4c	Communicating with families	0.24	0.25	No
Average Across Domains		0.13	0.36	No

Source: Mathematica calculations based on Pennsylvania student data and teacher observation data.

Note: Estimates are statistically significant at the 5 percent level if the absolute value of the estimate divided by its standard error is at least 1.96.

To explore these relationships further, we conducted similar analyses on partitioned samples by district, then by subject, and finally by grade range. The grade range partitions initially restricted the sample to teachers in grades 1 through 8 because the 11th-grade VAM estimates were highly imprecise. We then restricted the sample further to include only teachers with VAM estimates based on statewide samples (that is, PSSA outcomes in grades 4 through 8). With the exception of a few

³⁶ Specifically, this relationship is between rubric component scores and *estimated* value-added. As Jacob and Lefgren (2008) indicate, this relationship is likely to underestimate the relationship between rubric scores and a teacher's actual contributions to student learning because of measurement error in the value-added estimates. By applying the adjustment factor that they propose, we estimate that the actual relationships could be up to 23 percent larger than in Table IV.4. None of the qualitative inferences would change, however, because the fundamental problem is a lack of precision.

relationships for individual components in subject-specific analyses, none of these latter relationships were statistically significant. Given the large number of comparisons we conducted, we do not report these few relationships because we cannot rule out that they are statistically significant by chance. We also tried expressing the observation ratings as a series of indicator variables for final rating categories rather than as a continuous measure, but the results were similarly imprecise.

Although the findings from Phase 1 cannot differentiate among teacher practices in terms of their relationships with contributions to student learning, we do not view such analyses as hopeless for the future. Measuring these relationships requires a reliable measure in which evaluators are trained extensively to differentiate between levels of performance according to the rubric, and larger sample sizes. In short, ratings of professional practice have been found to be related to teachers' value-added. For instance, Tyler et al. (2010) studied the relationship between Cincinnati's Teacher Evaluation System (TES) and student achievement growth in math and reading. Like Pennsylvania's evaluation rubric, the TES is modeled on the Danielson Framework. The researchers found that a one-unit increase in the overall TES score was associated with student achievement gains that would move a student with the 50th percentile score to the 57th or 58th percentile, depending on the subject. Classroom management skills and, in reading, the use of inquiry-based teaching also were associated with greater gains in student achievement. In Chicago, nearly all of the Danielson Framework components had statistically significant relationships with value-added scores (Sartain et al. 2011). The teachers with the lowest rubric ratings tended also to have the lowest VAM estimates and vice versa. Milanowski et al. (2004) examined data from three school organizations that used rubrics based on the Danielson Framework: Cincinnati Public Schools, the Vaughn Next Century Learning Center charter school in Los Angeles, and the Washoe County school district in Nevada. A one-unit change in teacher evaluation scores on student achievement in these districts was similar in Cincinnati and Washoe to findings in Tyler et al. (2010) for math and reading and larger in the Vaughn charter school. Jacob and Lefgren (2008) concluded that principals in Chicago are typically able to distinguish between the teachers whose contributions to student achievement are the largest and the smallest, but are less able to distinguish teachers in the middle of the distribution. Finally, Rockoff and Speroni (2010) found that 3rd through 8th grade teachers in New York City who received higher ratings during their first year of teaching made greater contributions to student achievement in future years. The relationships they estimated remain intact even after additionally controlling for a teacher's value-added estimate from the first year of teaching. The authors also found evidence that observers vary in how they apply rating standards. Overall, the findings suggest that evaluation ratings based on both subjective and objective performance measures include more information than is conveyed by each measure independently, and that observation ratings standards should emphasize a high rate of interobserver agreement.

Finally, the Measures of Effective Teaching project sponsored by the Bill & Melinda Gates Foundation recently released a report that found positive relationships between student achievement gains and teacher practices using five classroom observation instruments including the Framework for Teaching (Kane and Staiger, 2012). The report was based on findings from 1,333 teachers in Charlotte-Mecklenburg, Denver, Hillsborough County (FL), New York City, and Memphis, who taught students in fourth through eighth grades in math and in English-language arts. Given these promising findings in the literature, we recommend pursuing these analyses further in Phase 2 of Pennsylvania's teacher evaluation pilot.

V. VALUE-ADDED RESULTS FOR PRINCIPALS

In this chapter, we analyze school-level value-added models (VAMs), considering whether they might produce valid, useful measures of principal effectiveness. A particular focus of our analyses is to assess the extent to which these VAMs disentangle the impacts of principals from the influences of other school-level factors beyond the principals' control.

In what follows, we begin by describing a method that, in theory, represents the best available approach to isolating principals' true impacts on student outcomes. We discuss how the limitations of this method prevent it from being applicable to real evaluations, and we present an alternative method, the school VAM, that is practicable but less able to separate principals' effects from the effects of other factors at their schools. The school VAM is the focus of all subsequent analyses. We assess the extent to which the school VAM produces estimates that approximate pure principal effects, and we conclude that this VAM should be regarded as estimating the effects of entire schools, which include both principal effects and the influence of other school-level factors. The final sections of this chapter describe several key characteristics of the effectiveness estimates generated by the school VAM.

A. An Approach to Estimating Pure Principal Effects

1. Challenges in Isolating Principals' Contributions to Student Achievement

A valid estimate of principals' effectiveness would isolate their effects on student achievement from the effects of other factors beyond the principals' control. Following the basic approach used for teacher VAMs, a natural starting point for assessing principals' effectiveness is to examine the average difference between actual and predicted outcomes among the students enrolled in a principal's school. This difference captures the contribution made by a principal's school to student achievement under the principal's tenure. In other words, *the starting point for estimating principal effectiveness is to estimate the effectiveness of the principal's school.*

The complication is that a school's effectiveness reflects more than just the effectiveness of the school's principal. It also reflects other school-specific characteristics and circumstances beyond the principal's control. First, preexisting teacher abilities—the abilities that teachers bring to the classroom regardless of the principal under whom they serve—contribute to school effectiveness. Principals affect student outcomes primarily by enabling their teachers to be more or less effective than expected, given their preexisting abilities. However, the mix of preexisting teacher abilities in a principal's school is often beyond his or her control. For example, a school located near a prestigious university might attract more highly motivated or capable teachers than a school in a less amenable location. The mix of abilities in a school's teaching staff can also reflect hiring decisions made by a principal's predecessor, and the current principal might have little flexibility to alter these decisions in the short run. Second, any differences among schools with respect to characteristics and

resources that are not accounted for in the VAMs—such as differences in funding, facilities, and neighborhood quality—can also lead to differences in the estimated effectiveness of the schools.³⁷

Thus, a key analytic challenge of any statistical method that aims to identify the achievement effects of principals is to disentangle principals' true contributions to student outcomes—that is, pure principal effects—from the influence of other school-level factors. An estimation method used by some previous studies aims to address this challenge, which we discuss next.

2. Principal Transitions Model: Basic Structure

A type of VAM that we refer to as the *principal transitions model* provides an approach to distinguishing principal effects from the effects of other school characteristics and circumstances. Starting from estimates for schools' contributions to student outcomes, the model takes a further analytic step. It calculates how the same school's contribution differs under the leadership of different principals, and these differences serve to measure how effective a principal is relative to the other principals who have served at the same school. For example, if student outcomes relative to predicted outcomes rise when principal B succeeds principal A at a given school, then B is deemed to be more effective than A. Thus, the name of this model refers to the fact that only schools with leadership transitions during the considered period can be included in the analysis.

Because the principal transitions model is fundamentally based on comparing principals who have served at the same school, it controls for certain types of school-specific factors that are beyond principals' control. Specifically, the model controls for any school characteristics and circumstances that remain constant during the analysis period. These school-specific factors are common to all principals who have served at the same school, so they cannot contaminate comparisons of effectiveness among these principals. For example, if a school's proximity to a prestigious university does not change over time, then the resulting advantage in teacher recruitment will benefit equally all principals who have led this school and, thus, will not generate *differences* among the effectiveness estimates of these principals.

In the sparse research literature on the variation in and correlates of principal effectiveness, the principal transitions model is the most common type of principal VAM used by researchers. In fact, to our knowledge, all existing studies that have generated value-added estimates of principal effectiveness have used variants of the principal transitions model, either exclusively or in conjunction with alternative models (Branch et al. 2011; Dhuey and Smith 2011; Coelli and Green, forthcoming).³⁸ The popularity of this model stems from its ability to control for constant, school-specific influences on student achievement.

³⁷ The same types of school-level differences can be reflected in teacher value-added estimates as well, but the problem is less severe for teachers because the bulk of the variation in teacher effectiveness estimates is observed within—rather than between—schools (see Chapter III).

³⁸ Coelli and Green (forthcoming) augment the principal transitions model to allow a principal's impact on a school's effectiveness to evolve gradually over time with the principal's tenure at the school.

3. Limitations of the Principal Transitions Model

Although the principal transitions model might suit the research purposes of the previously described studies, it cannot be applied to real-world evaluations of principals. We identified and explored several limitations of this model and found them to be too severe to enable the model to be used in practice.

One major limitation of the principal transitions model is that it can generate effectiveness estimates for only a limited group of principals. Specifically, it can include only principals who have led schools in which a leadership transition has occurred during the analysis period. In a model with only one student cohort—that is, an analysis period of one year—the model is completely infeasible due to the lack of leadership transitions. Even over a three-year period, only a minority of schools undergo leadership transitions. For example, among all schools that contain students with 5th-grade Pennsylvania System of School Assessment (PSSA) math outcomes, 35 percent of schools (encompassing about 49 percent of principals) experienced at least one leadership transition during the 2008–2009 through 2010–2011 school years. In other words, excluding schools without leadership transitions would reduce by half the number of principals that can have effectiveness estimates. Thus, too many principals would have no effectiveness estimates if the principal transitions model were used for real evaluations.

For the principals who can have effectiveness estimates, the principal transitions model also limits the ways in which these principals can be compared on their performance. Comparisons can be made only within small connected networks of schools. Each connected network is a set of schools such that every member school has had at least one of its principals transfer to at least one other member school during the analysis period. By virtue of these transfers, principals from different schools within the same network can be compared; for example, if two principals from different schools are compared with a third principal who has served at both schools, they can, by implication, be compared with each other. However, the principal transitions model cannot determine how well a principal performed relative to another principal in a different network.

Connected networks are typically very small in the three-year period (2008–2009 through 2010–2011) covered by our analysis. Again, consider the set of principals from schools with 5th-grade PSSA math outcomes. Even among principals whose schools have undergone leadership transitions, 61 percent of these principals belong to networks with only a single school—their own school. In these networks, neither the predecessor(s) nor the successor(s) in the leadership of the school were observed in any other school with 5th-grade PSSA outcomes. Another 22 percent belong to networks with exactly two schools. Thus, only 17 percent of principals from schools with leadership transitions belong to connected networks with three or more schools, representing less than 9 percent of principals from all schools with or without leadership transitions. The key consequence is that the principal transitions model can determine how effective a principal is only relative to a very limited group of other principals. A meaningful evaluation system would need an assessment of a principal's effectiveness relative to a much broader comparison group.

Because the principal transitions model cannot be used in a real evaluation system, it is necessary to turn to an alternative value-added approach. We discuss this alternative approach next.

B. The School VAM as the Basis for Evaluating Principals

1. Analytic Approach

Given that the principal transitions model is infeasible for actual evaluations, the most straightforward alternative is simply to evaluate principals on the basis of their schools' contributions to student achievement. For each principal, this method—the school VAM—calculates the average effectiveness of the school(s) led by the principal during the analysis period. In other words, each principal receives a value-added score based on the difference between actual and predicted student outcomes averaged over all of the schools under his or her leadership during the analysis period.

The major limitation of using a school VAM to evaluate principals is that it bundles principals' true contributions with the effects of other school-level factors. Any types of differences across schools that are not accounted for in the model—such as differences in preexisting teacher abilities or school resources—could lead to differences in the value-added scores that principals receive. Compared with estimates from the principal transitions model, estimates from the school VAM have less validity as measures of pure principal effects.

However, the school VAM has several advantages. It does not suffer from the limitations of the principal transitions model; school value-added scores can be calculated for, and compared among, all eligible principals. This method also has the advantage of being conceptually straightforward: principals are held responsible for the extent to which their schools—including the teachers under their authority—affect student outcomes. Various districts and states, including Dallas and Tennessee, have implemented the approach of using school effectiveness measures to evaluate principals (see Lipscomb et al. 2010b). A school VAM has also been used in prior research to gauge the variation in principal effectiveness (Branch et al. 2011).

By virtue of being potentially applicable to a real evaluation system for principals, the school VAM is the central focus of the remainder of this chapter. Nonetheless, the following point is worth reiterating: *effectiveness estimates from school VAMs actually capture the effects of entire schools, including the effects of all instructional staff and educational inputs located at these schools.*

2. Comparison of School VAM and Principal Transitions Model

To interpret properly the effectiveness estimates from the school VAM, it is important to quantify the degree to which they deviate from pure principal effects. If these deviations were small, we could infer that school-level factors beyond the principals' control had only a small influence on school effectiveness, and we could regard estimates from the school VAM as being primarily indicative of principals' contributions. Large deviations, on the other hand, would suggest these estimates were poor measures of principal effects.

To carry out this analysis, we use estimates from the principal transitions model as benchmarks with which estimates from the school VAM are compared. Despite being inapplicable to real evaluations, estimates from the transitions model nevertheless represent our best estimates of pure principal effects and, as such, can serve as a useful point of comparison. Thus, to the extent that the two models yield more similar results, there will be greater justification for interpreting estimates from the school VAM as primarily reflecting principals' true contributions to student achievement.

Using Pennsylvania data, we obtained effectiveness estimates from the school VAM and the principal transitions model for the set of principals that can be included in both models. Because the initial estimates from the school VAM compare principals from all schools—not just the principals in the same connected network, as in the principal transitions model—we first converted these estimates so that they would have the same meaning as those from the transitions model. From each principal’s effectiveness estimate, we subtracted the average effectiveness estimate in the principal’s network. As a result, the final estimates from both models capture the deviation of a principal’s effectiveness from the average in the same network. For each of the two models, we placed principals into quartiles based on how much they outperform or underperform the average principal in their network.

Table V.1 compares the quartiles into which principals are placed based on the two models. This table addresses the question: *To what extent do the school VAM and principal transitions model rank principals similarly on their measured effectiveness?* Each row of the table represents a particular quartile of principals from the transitions model, and row entries show the number and percentage of those principals who are placed into each of the four quartiles based on the school VAM. The diagonal entries of the table represent the cases in which the two models coincide in placing principals into the same effectiveness quartile. For ease of presentation, we show results based on two outcomes—5th-grade PSSA math scores and 8th-grade PSSA reading scores—and pool the analyses related to both outcomes together into a single table.

We find a moderate degree of consistency between the effectiveness rankings produced by the two models. As shown by the diagonal (upper left to lower right) entries of Table V.1, about half of the principals in the analysis are placed into identical quartiles by the two models. Moreover, for most principals, their effectiveness estimates based on the school VAM differ by no more than one quartile from their effectiveness estimates based on the principal transitions model. The simple correlations between effectiveness estimates from the two models—0.39 in 5th-grade math and 0.58 in 8th-grade reading—also yield the same conclusion that the two models are moderately consistent with each other.

Although these results are encouraging, a noticeable minority of principals still receive a ranking from the school VAM that is substantially different from their transitions model ranking. For example, of the principals in the bottom quartile identified by the transitions model, 27 percent are in the top two quartiles identified by the school VAM. Similarly, 24 percent of principals in the top quartile from the principal transitions model are in the bottom two quartiles from the school VAM. Notably, there is less consistency between the school VAM and the principal transitions model than there is between the various teacher VAMs from Chapter III that used different baseline achievement controls (see Tables III.5 and III.6). The quartiles of effectiveness into which teachers were placed by these different VAMs rarely differed by more than one quartile.

Table V.1. Counts and Percentages of Principals in Effectiveness Quartiles Based on Principal Transitions Model and School VAM

	Quartile of Effectiveness Based on School VAM				Total
	1st (bottom)	2nd	3rd	4th (top)	
Quartile of Effectiveness Based on Principal Transitions Model					
1st (bottom)	69 (53.5)	25 (19.4)	15 (11.6)	20 (15.5)	129 (100.0)
2nd	31 (23.7)	63 (48.1)	26 (19.8)	11 (8.4)	131 (100.0)
3rd	13 (10.4)	25 (20.0)	58 (46.4)	29 (23.2)	125 (100.0)
4th (top)	16 (12.6)	15 (11.8)	29 (22.8)	67 (52.8)	127 (100.0)
Total	129 (25.2)	128 (25.0)	128 (25.0)	127 (24.8)	512 (100.0)

Source: Mathematica calculations based on Pennsylvania data.

Note: In each table cell, the first value is the number of principals in the given cell and the second value (in parentheses) is the percentage of the row total that is represented by that cell. Findings are based on statewide samples of principals with effectiveness estimates in either grade 5 math or grade 8 reading; principals with both types of effectiveness estimates are counted twice. The sample of principals consists of those who served as principals at any time from 2008-2009 to 2010-2011. Only principals with effectiveness estimates from both the school VAM and the principal transitions model are included in the analysis. To construct this table, principals were placed into cells separately for each of the two examined outcomes, and the resulting counts in each cell were aggregated across the two outcomes.

Overall, the results from Table V.1 suggest that *estimates from the school VAM are an informative but imperfect measure of principals' contributions to student learning*. Given the moderate consistency of these estimates with those from the transitions model, some of the variation in these estimates among principals is likely to capture true differences in principal quality. However, the discrepancies with the transitions model also suggest that some of the variation in these estimates is picking up school-level differences outside of the principals' control. Based on this evidence, we continue to believe that the school VAM estimates measure the contributions of entire schools to student achievement, and should not be described as principal value-added measures.

C. Key Characteristics of School Effectiveness Estimates Based on PSSA Outcomes

We proceed to describe several empirical features of the effectiveness estimates from the school VAM, including the amount of variation and the extent of statistical uncertainty in these estimates. All of these features influence the extent to which the VAM can distinguish effective and ineffective schools.

For ease of exposition, we will refer to schools as the entities being compared when describing school VAM estimates. As described previously, the method actually generates one estimate per principal based on the effectiveness of the principal's school(s) under his or her tenure; a principal who has led multiple schools is assigned a single estimate based on the average effectiveness of those schools. Nevertheless, referring to schools as the units of comparison emphasizes, once again,

that the school VAM fundamentally measures the contributions of entire schools—not only principals—to student outcomes.

This section focuses on school VAM estimates for which PSSA scores are the outcomes of interest. As discussed in Chapter II, VAMs based on PSSA outcomes yield effectiveness estimates for schools across the entire state, enabling the analysis of VAM characteristics to be supported by large samples. Unless otherwise noted, most of the following analyses pertain to VAM estimates that use data from three student cohorts.

1. Variation in Measured Effectiveness Across Schools

In order for VAMs to be informative in distinguishing effective and ineffective schools, it is necessary that there exist meaningful variation in the effectiveness estimates across schools. To document this variation, we calculated the extent to which schools at selected percentiles differ from the average school. Table V.2 presents these measured differences for selected grade-subject combinations. In addition, Appendix Table C.5 expresses differences in effectiveness estimates in terms of standard deviations of student scores for all grade–subject combinations.

Table V.2. Distribution of School Effectiveness Estimates for Selected PSSA Outcomes

Outcome	Effectiveness of the School at the Indicated Percentile Relative to the Effectiveness of the Average School (in PSSA scale points)					
	5th	15th	25th	75th	85th	95th
Math PSSA, Grade 5	-67	-40	-26	24	43	69
Math PSSA, Grade 11 ^a	-94	-50	-33	35	53	85
Reading PSSA, Grade 8	-45	-28	-18	19	29	46
Science PSSA, Grade 4	-68	-43	-26	27	43	69

Source: Mathematica calculations based on Pennsylvania data.

Note: Unless otherwise noted, findings are based on a three-cohort model with statewide samples of schools, principals, and students. The sample of principals consists of those who served as principals in every year from 2008-2009 to 2010-2011.

^a Findings are based on a two-cohort model because three years must elapse between the baseline and outcome scores.

The results indicate that effectiveness estimates vary considerably across schools. For example, by attending the 15th percentile school, a 5th-grade student would score 40 scale points lower on the math PSSA than he or she would score by attending the average school. On the other hand, this student would score 43 scale points higher in the 85th percentile school than in the average school. Thus, the 85th and 15th percentile schools differ in their effectiveness by 83 PSSA points. This scale point difference can also be interpreted with reference to the statewide distribution of student test scores. By switching from the 15th to the 85th percentile school, a 5th-grade student originally at the median of the statewide distribution of math scores would be predicted to rise to the 65th percentile.³⁹

³⁹ To make this calculation, we divided 83 scale points by 223 scale points, the standard deviation of 5th-grade PSSA math scores shown in Appendix Table C.5. Thus, an 83 scale point difference amounts to a difference of 0.37

Sizable variation in the school effectiveness estimates is also observed for other grade–subject combinations. The 85th and 15th percentile schools differ in their estimated effects on PSSA scores by 103 points in 11th-grade math, 57 points in 8th-grade reading, and 86 points in 4th-grade science. As we found for teachers in Chapter III, variation in schools’ effects appears to be smaller in reading than in math for every grade level (see Appendix Table C.5).

Our results for the variation in school effectiveness are in the range of the variation found by prior research. For example, for math PSSA outcomes, we find that the difference in effectiveness between the 85th and 50th percentile schools, expressed in standard deviations of student scores, ranges from 0.15 to 0.21 across grades (see Appendix Table C.5). Similarly, using math test scores from Texas in grades 3 through 8, Branch et al. (2011) found a corresponding difference of 0.21 standard deviations.

Notably, the variation in school effectiveness estimates is similar in magnitude to the variation in teacher effectiveness estimates described in Chapter III. The reason that effectiveness estimates do not vary considerably more for teachers than for schools is that VAMs account for imprecision in both types of estimates by “shrinking,” or pulling, these estimates toward their respective averages. Teachers’ estimates are pulled more heavily toward the average due to their greater imprecision. Appendix A describes this shrinkage approach in further detail.

In summary, there is meaningful variation in schools’ estimated impacts on PSSA scores. As a consequence, if an evaluation system for principals used the school VAM, it would be feasible for the system to delineate groups of principals that differed in their school performance estimates by a substantively important magnitude. This does not mean, however, that school VAM estimates are *valid* measures of principal performance, because they also include aspects of school performance that are outside of principals’ control.

2. Statistical Uncertainty in the School Effectiveness Estimates

As with the teacher effectiveness estimates examined in Chapter III, the effectiveness estimates for schools contain some degree of statistical uncertainty. To the extent that there is less uncertainty in these estimates, chance errors in estimation—due, for instance, to random fluctuations in the composition of a school’s students—exert less influence on these estimates. Thus, quantifying statistical uncertainty is important for determining which of the measured differences across schools are unlikely to have arisen purely by chance.

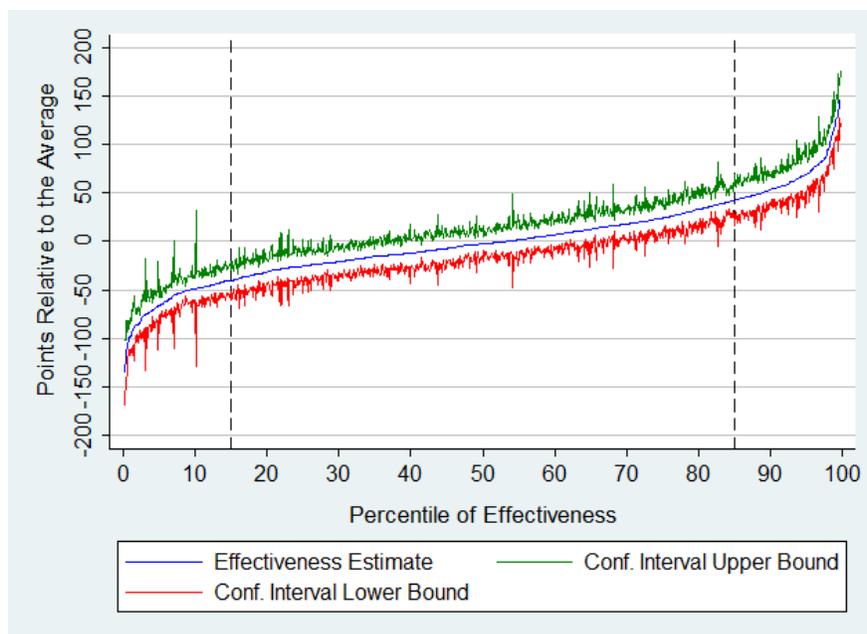
Figure V.1 provides a visual depiction of both the statistical uncertainty in the school effectiveness estimates and the variation in these estimates across schools based on 5th-grade PSSA math outcomes. As with similar figures for teachers in Chapter III, this figure plots (on the vertical axis) the school effectiveness estimates, as well as the lower and upper bounds of the 95 percent confidence intervals for these estimates, against the school’s percentile rank (on the horizontal axis). Estimates whose confidence intervals lie completely above or completely below zero—defined to be

(continued)

standard deviations in the distribution of student scores. In the assumed normal distribution for student scores, moving from the 50th to the 65th percentile is equivalent to an increment of 0.37 standard deviations.

the effectiveness of the average school—are statistically distinguishable from average school effectiveness.

Figure V.1. Distribution of School Effectiveness Estimates and 95 Percent Confidence Intervals of School Effectiveness Estimates for Math PSSA Grade 5 Scores



Source: Mathematica calculations based on Pennsylvania data.

Note: Findings are based on a three-cohort model with statewide samples of schools, principals, and students. The sample of principals consists of those who served as principals in every year from 2008-2009 to 2010-2011. In the figure, vertical dotted lines are drawn at the 15th and 85th percentiles.

As shown in Figure V.1, nearly all of the schools in the top and bottom quartiles of the performance distribution have effectiveness estimates that are statistically distinguishable from average effectiveness. For example, both the 15th and 85th percentile schools, represented by dotted lines in Figure V.1, differ from average effectiveness by a statistically significant extent. Naturally, the share of schools that is statistically different from the average declines as the estimates move closer to the middle of the performance distribution. Effectiveness estimates between the 40th and 60th percentiles are generally not statistically different from average effectiveness. The figure thus provides a visual indication that the VAM can identify highly effective and highly ineffective schools but is less able to make distinctions among schools near the middle of the performance distribution.

To summarize the extent of statistical uncertainty in the school effectiveness estimates, Table V.3 provides the number and share of schools whose effectiveness estimates are statistically different from average effectiveness. For the three-cohort models—the main focus of our analysis—the share of schools that can be distinguished from the average ranges from 58 to 69 percent, depending on the outcome measure. Comparing these results with those from Table III.3, we find that larger fractions of schools than teachers have effectiveness estimates that are statistically distinguishable from the average. The reason is that schools' effectiveness estimates are typically based on larger samples of students and, hence, have greater precision.

Table V.3. Number of Schools with Effectiveness Estimates Reported and Share of Reported Estimates that Are Statistically Different from the Average, by Number of Cohorts Used in Estimation

Outcome	Number of Principals		Percentage of School Effectiveness Estimates that Are Statistically Distinguishable from the Average	
	1-Cohort Model	3-Cohort Model	1-Cohort Model	3-Cohort Model
Math PSSA, Grade 5	1,336	1,079	53.0	66.4
Math PSSA, Grade 11	626	577 ^a	59.7	68.3 ^a
Reading PSSA, Grade 8	755	580	41.2	58.3
Science PSSA, Grade 4	1,427	1,166	54.1	69.0

Source: Mathematica calculations based on Pennsylvania data.

Note: Findings are based on a 95 percent confidence interval and statewide samples of schools, principals, and students. In the one-cohort model, the sample of principals consists of those who served as principals in 2010-2011. Unless otherwise noted, the sample of principals in the three-cohort model consists of those who served as principals in every year from 2008-2009 to 2010-2011.

^a Findings are based on a two-cohort model because three years must elapse between the baseline and outcome scores.

Given that statistical imprecision is a less severe problem for estimating school effectiveness than for estimating teacher effectiveness, an issue for consideration by the Pennsylvania Department of Education (PDE) is whether to use fewer than three cohorts for the school VAM. A key advantage of using a smaller number of (the most recent) cohorts is that the VAM provides a more up-to-date measure of school performance. This advantage must be weighed against the decrease in the precision of the estimates. For the outcomes shown in Table V.3, the fractions of schools that are statistically distinguishable from the average are lower in one-cohort models than in three-cohort models by 9 to 17 percentage points. For example, in a school VAM based on 5th-grade PSSA math scores, 53 percent of schools are statistically different from the average in a one-cohort model, whereas the corresponding percentage is 66 percent in a three-cohort model. In other words, it is more difficult for one-cohort models to distinguish true performance differences among schools from random fluctuations in the outcomes of their students. Of note, however, is that the shares of schools that are statistically distinguishable from the average in one-cohort models are at least as high as the corresponding shares for teachers in *three-cohort* models (see Tables III.3 and V.3). Thus, the decrease in precision from using fewer cohorts could be more tolerable in the case of schools than in the case of teachers.

D. Key Characteristics of School Effectiveness Estimates Based on Outcomes Other than PSSA Scores

Analyzing additional outcomes beyond PSSA scores has the potential to provide a more comprehensive picture of each school's impact. In Phase 1, we explored two broad categories of additional outcomes to which we applied school VAMs. First, as we did for teachers in Chapter III, we estimated school VAMs based on assessments administered by pilot districts to the lower elementary grades—grades excluded from the PSSA-based models. Second, we examined schools' impacts on key nonassessment outcomes—holding power and attendance—that are regarded as important precursors of academic success. This section describes key characteristics of the effectiveness estimates from these VAMs.

1. VAMs Based on Assessments Administered by Pilot Districts

Given that the PSSA does not cover all grade–subject combinations, the use of data from district assessments can expand the set of grades and subjects that can be included in school VAMs. As we discussed in previous chapters, the pilot districts administer a number of assessments at lower elementary grades not covered by the PSSA. However, expanding coverage of VAMs to the lower elementary grades yields a different type of benefit for principal evaluation systems than for teacher evaluation systems. Although adding lower elementary grades to the analysis samples would substantially expand the set of teachers with value-added scores, it would lead to only modest increases in the number of principals with school value-added scores. Most schools with lower elementary grades also contain upper elementary grades covered by the PSSA, enabling their principals already to have school VAM estimates based on these PSSA outcomes. For example, in a three-cohort model, whereas 1,249 principals would have a school VAM estimate from at least one 4th- or 5th-grade PSSA assessment, the inclusion of all remaining elementary grades (K–3) into the VAMs would bring, at most, 51 additional principals into the analyses—a 4 percent increase. Thus, for principal evaluations, the primary benefit of applying VAMs to the lower elementary grades is to be able to measure elementary schools’ effectiveness based on the widest possible set of grades.

To assess the potential for including district assessments in school VAMs, we generated effectiveness estimates for schools in the pilot districts based on locally administered assessments (see Chapter II for a discussion of the assessments and samples). We analyzed the same key characteristics of these effectiveness estimates as we did for the PSSA-based estimates—namely, the extent of variation across schools and the level of precision.

Schools in the pilot districts appear to differ in their impacts on district assessment scores by meaningful magnitudes. Table V.4 shows the effectiveness estimates (relative to average effectiveness) for the schools at the 15th and 85th percentiles; we omit other percentiles to maintain participants’ confidentiality, given the small sample sizes in this analysis. The 15th and 85th percentile schools in Allentown differ in effectiveness by 14 percentage points on the Writing Progress Assessment in 1st grade and by 9 percentage points on the Math Progress Assessment in 2nd grade. To interpret these differences, it is again instructive to convert them to increments within the distribution of student test scores. A student who would have had the median Progress score in Allentown if assigned to the 15th percentile school would, instead, be at the 71st to 84th percentiles of Progress scores if assigned to the 85th percentile school. There is less variation in schools’ impacts on 2nd-grade Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scores, even though the sample includes two districts (Allentown and Cornell). In the distribution of DIBELS scores, the eight-point difference in effectiveness between the 15th and 85th percentile schools is equivalent to moving a student from the median to the 59th percentile of scores. Overall, however, these results (as well as results for additional assessments shown in Appendix Table C.6) generally suggest sizable variation in schools’ contributions to student scores on the pilot districts’ assessments.

Table V.4. Key Characteristics of School Effectiveness Estimates Based on Selected Tests Administered in the Pilot Districts

Outcome	Effectiveness of the School at the Indicated Percentile Relative to the Effectiveness of the Average School (in test scale points)		Percentage of School Effectiveness Estimates that Are Statistically Distinguishable from the Average
	15th	85th	
Progress Assessment, Writing, Grade 1 (Percentage Points) ^a	-8	6	53.8
Progress Assessment, Math, Grade 2 (Percentage Points) ^a	-4	5	66.7
DIBELS, ORF, Grade 2 ^b	-4	4	13.3
PSSA, Math, Grade 3 ^c	-35	31	37.5
PSSA, Reading, Grade 3 ^c	-26	25	37.5

Source: Mathematica calculations based on Pennsylvania data.

Note: Findings are based on a 95 percent confidence interval and a one-cohort model with samples of schools, principals, and students from the pilot districts. The sample of principals consists of those who served as principals in 2010-2011.

^aAllentown only.

^bAllentown and Cornell only.

^cAllentown, Cornell, and Mohawk only.

DIBELS = Dynamic Indicators of Basic Early Literacy Skills; ORF = oral reading fluency.

As indicated in Chapter III, we caution that expressing the variation in school effects in terms of the existing dispersion in student scores does not fully gauge whether this variation is educationally meaningful. A closer examination of the content validity of these assessments is necessary for determining whether this variation translates into substantive differences in students' skills and knowledge.

The ability of the school VAM to make statistically reliable distinctions among schools differs by outcome measure. As shown in the last column of Table V.4, for estimates based on the Progress Assessments, one-half to two-thirds of schools can be statistically distinguished from average effectiveness. However, only 13 percent of schools can be distinguished from the average based on their impacts on second-grade DIBELS scores. As we described in Chapter III, a greater proportion of the variation in the DIBELS-based effectiveness estimates is due to random fluctuations and other sources of imprecision, making it more difficult to identify high- and low-performing schools with high degrees of confidence.

In addition to using outcomes from non-PSSA assessments as a means of expanding grade-level coverage, we also explored applying school VAMs to 3rd-grade PSSA outcomes by using 4Sight scores as baseline achievement measures. The final two rows of Table V.4 show the characteristics of the resulting estimates. Consistent with schools' impacts on PSSA scores at other grade levels, impacts on 3rd-grade PSSA scores vary substantially across schools; the 85th and 15th percentile schools differ in effectiveness by 66 PSSA points in math and 51 PSSA points in reading. A slightly lower proportion (38 percent) of schools can be statistically distinguished from the average school on the basis of the 3rd-grade PSSA VAMs than on the basis of the other one-cohort PSSA VAMs (shown previously in Table V.3). One reason is that the baseline scores and other student characteristics do not explain as much of the outcome variance in the 3rd-grade PSSA VAMs as in

most of the other PSSA VAMs that we estimated.⁴⁰ However, it is unclear whether the pilot districts are unique in this respect, or whether it is generally the case statewide that fall 4Sight scores have less power than prior PSSA scores to predict current-year PSSA scores. Later pilot phases, with a larger sample of districts, can address this question more definitively.

The results for the 3rd-grade VAMs indicate that using 4Sight scores as baseline measures is one potentially feasible strategy for including 3rd grade in the school VAMs. The benefits of including 3rd grade through this strategy should be weighed against several additional considerations. First, there are likely to be costs of introducing the 4Sight assessment into districts that currently do not use it. Second, like any other fall assessment, the fall 4Sight would be used only as a baseline measure and not as an outcome measure in any VAM, which might give educators an incentive to deemphasize—or even intentionally depress performance on—this assessment. Third, the fall 4Sight is not the only baseline measure that could potentially be used in a 3rd-grade VAM, but it is likely to be the one most closely aligned with the content of the 3rd-grade PSSA.

2. VAMs Based on Nonassessment Outcomes

Outcomes for VAMs do not necessarily have to be limited to test scores. If using multiple measures of student outcomes in evaluations is deemed to be a priority, then schools' impacts on nonassessment outcomes merit consideration. In fact, for various types of nonassessment outcomes, estimating the impacts of entire schools is more feasible than estimating the impacts of individual teachers. For example, a student's persistence in remaining enrolled in high school—which we call holding power—and a student's attendance rate are likely to be affected by multiple teachers who instruct this student, as well as by a myriad of other factors at the student's school, such as school culture. Although the influence of all of these school-based factors can be bundled together into an estimate of the school's impact, it would be much more difficult to isolate the effects of individual teachers on these outcomes.

We estimated school VAMs for the two nonassessment outcomes mentioned previously—holding power and attendance rate—because they are regarded as important intermediate outcomes that feed into educational attainment and achievement outcomes. A school's effectiveness estimate for a nonassessment outcome has a similar meaning as it does for an assessment outcome: it is the difference between what the school's students achieve and what they would have achieved if they had been assigned to the average school.

Table V.5 shows the distribution of school effectiveness estimates for nonassessment outcomes. Because the holding power VAMs are based on large, statewide samples and the attendance rate VAMs are based on the pilot districts only, we show several more percentiles of the school effectiveness distribution for holding power than for attendance.

⁴⁰ As shown in Appendix C.6, the R-squared values from the 3rd-grade PSSA VAMs range from 0.64 to 0.69. However, the R-squared values (not shown in the tables) for the one-cohort models based on 5th-grade math PSSA and 8th-grade reading PSSA scores are 0.77 and 0.74, respectively.

Table V.5. Key Characteristics of School Effectiveness Estimates Based on Nonassessment Outcomes

Outcome	Effectiveness of the School at the Indicated Percentile Relative to the Effectiveness of the Average School (in Percentage Points)						Percentage of Effectiveness Estimates that Are Statistically Distinguishable from the Average
	5th	15th	25th	75th	85th	95th	
Holding Power, Grade 9	-30.8	-21.3	2.4	7.0	7.4	10.6	92.2
Holding Power, Grade 10	-32.5	-21.2	2.8	6.8	7.3	11.0	92.0
Holding Power, Grade 11	-81.5	1.4	7.0	10.7	11.3	20.6	90.0
Attendance Rate, Grades 4-12	-- ^a	-0.5	-- ^a	-- ^a	0.6	-- ^a	13.8

Source: Mathematica calculations based on Pennsylvania data.

Note: Findings on holding power in grades 9 and 10 are based on a two-cohort model with statewide samples of schools, principals, and students; for these outcomes, the sample of principals consists of those who served as principals in both 2009-2010 and 2010-2011. Findings on holding power in grade 11 are based on a one-cohort model with statewide samples of schools, principals, and students. Findings on attendance are based on a one-cohort model with samples of schools, principals, and students from Allentown, Mohawk, and Northwest Tri-County. For the one-cohort models, the sample of principals consists of those who served as principals in 2010-2011. All analyses are based on a 95 percent confidence interval.

^a Due to the small sample sizes in this analysis, effectiveness estimates at these percentiles are suppressed in order to protect the confidentiality of the sample members.

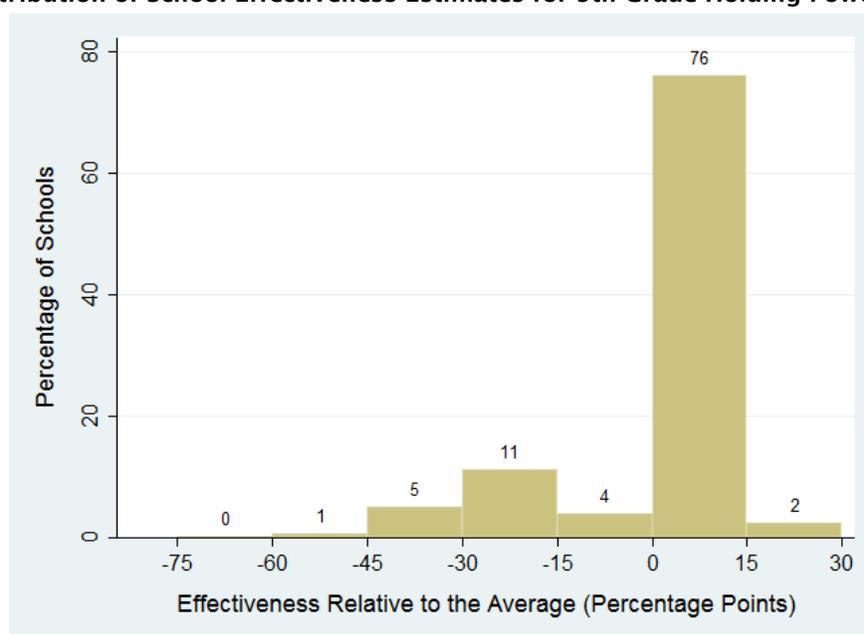
There are striking differences among schools in their effectiveness at keeping students enrolled in high school. A 9th-grader who attends the 85th percentile school is 29 percentage points more likely to enroll in a Pennsylvania public school again in the following year than if he or she had attended the 15th percentile school. Large differences are also observed at the other high school grades.

The distribution of schools' effects on holding power is highly skewed, as shown in both Table V.5 and the histogram in Figure V.2. The bottom 6 percent of schools have extremely negative impacts on holding power, lowering their students' probability of staying enrolled by more than 30 percentage points relative to the average school. These schools pull the average effectiveness in the sample downward by such a magnitude that a large majority (78.5 percent) of schools have effectiveness estimates that are above average, many of which are statistically significantly above average. Moreover, nearly all of these above-average schools have effectiveness estimates that are within 15 percentage points of one another. These patterns suggest that there are stark differences between the worst-performing high schools and the other schools in the state with respect to their impacts on keeping students enrolled.

On the basis of these results, school effectiveness estimates for holding power appear to be an informative tool for identifying high schools that perform poorly in keeping their students enrolled in Pennsylvania's public schools. However, a number of factors merit attention when interpreting these estimates. First, these estimates are only as valid as the underlying data on student enrollment. To the extent that any students' enrollment records are missing from the Pennsylvania Information Management System (PIMS), the schools that these students attended in the previous year will be erroneously regarded as having failed to "hold on" to these students. Second, holding power in

grades 9 through 11 is an intermediate outcome for a final outcome that is usually of greater interest—graduation. We could not estimate schools’ impacts on graduation (and/or holding power in grade 12) in Phase 1 because there were insufficient years of data to control for the 8th-grade PSSA scores of 12th-graders.⁴¹ With an additional year of student data, it will be feasible to estimate VAMs with graduation outcomes, enabling us to examine whether schools’ impacts on graduation outcomes are closely associated with their impacts on the intermediate holding power outcomes. With these caveats in mind, the Phase 1 evidence indicates that the holding power estimates provide additional information about high schools’ impacts on students beyond the information contained in effectiveness estimates for 11th-grade PSSA scores.⁴²

Figure V.2. Distribution of School Effectiveness Estimates for 9th-Grade Holding Power



Source: Mathematica calculations based on Pennsylvania data.

Note: Findings are based on a three-cohort model with statewide samples of schools, principals, and students. The sample of principals consists of those who served as principals in every year from 2008-2009 to 2010-2011.

On the other hand, estimates of schools’ contributions to attendance outcomes are largely uninformative for identifying effective and ineffective schools, at least in the Phase 1 sample. Students enrolled in the 85th percentile school are only one percentage point more likely to be in

⁴¹ In theory, it would have been possible to estimate a VAM for graduation outcomes at the end of spring 2011 with controls for 8th-grade PSSA scores from spring 2007. However, we would not have had data to check the accuracy of the graduation measure—in particular, to check whether students deemed to have graduated actually continued to be enrolled in the following year. Moreover, measuring graduation status at the end of spring 2011 would ignore cases of graduation that occurred as a result of summer school (in summer 2011).

⁴² In fact, there is very little correlation—and, in some cases, a slight negative correlation—between the school effectiveness estimates for holding power and those for 11th-grade PSSA scores. For example, correlations between schools’ impacts on 11th-grade math PSSA scores and their impacts on holding power range from -0.05 to -0.03, depending on the grade at which holding power is measured. For reading, correlations with the holding power estimates range from -0.05 to 0.

attendance than they would have been if enrolled in the 15th percentile school. Thus, there is little variation across schools in their effects on attendance rates and, as a result, relatively few (14 percent of) schools are statistically distinguishable from the average on this measure of effectiveness. These findings indicate that the school VAMs for attendance outcomes could not make meaningful, reliable distinctions among schools in the Phase 1 sample. In subsequent phases of the pilot, we will determine whether these findings continue to be observed in a larger pilot sample. In moving to larger samples, the viability of attendance as an outcome will also depend on whether it is measured consistently across districts in the state. Thus, the extent of variation in schools' impacts on attendance and the uniformity with which attendance is measured are the two critical factors for determining whether this outcome should be used in the model evaluation system.

VI. LOOKING AHEAD TO SUBSEQUENT PHASES OF THE PENNSYLVANIA TEACHER AND PRINCIPAL EVALUATION PILOT

As Pennsylvania continues its efforts to improve teacher and principal evaluations in coming years, lessons learned from Phase 1 will be important for the future development of a statewide evaluation system. In the track of the Phase 1 pilot that pertained to measuring effectiveness through the use of student data, Mathematica sought to address research questions related to how value-added models (VAMs) can be used to characterize the effectiveness of teachers and principals at raising student achievement according to multiple outcome measures and whether specific teacher practices relate to larger contributions to student achievement. Consistent with the initial phase of a pilot experiment, the findings from our analyses point to areas of progress and areas in which further attention should be directed during later pilot phases.

We have documented large variation in the estimated effectiveness of teachers and schools based on contributions to growth in assessment scores. The effectiveness estimates can distinguish between educators at the high and low ends of the distribution (provided that student samples sizes are not very small), and we have identified potential outcomes in which the variation in student performance is sufficient to make attributions of effectiveness to individual educators. However, the Phase 1 pilot's small number of teachers was a significant limiting factor for drawing inferences about teacher practices that relate to larger contributions to achievement growth and, to a lesser extent, to the extension of value-added methods beyond grades 4 through 8 that are covered by the Pennsylvania System of School Assessment (PSSA). We also concluded that it is not possible to isolate a principal's contribution to achievement growth using estimates of school-wide effectiveness (though this fact does not necessarily preclude the use of school-level VAM estimates for principal accountability purposes).

Despite these limitations in the Phase 1 analyses, the findings offer much new information and perspective on how to focus the pilot moving forward. Improving the measurement of teacher and principal effectiveness is a difficult challenge for which there is no quick fix, but a challenge that Pennsylvania's leaders have taken on because the ultimate issue at stake is improving student achievement. To conclude our final report from Phase 1 of the Pennsylvania Teacher and Principal Evaluation Pilot, we offer the following list of recommendations in preparation for Phase 2 and later phases that relate broadly to issues of sampling and measurement:

Sample of Teachers and Principals to Be Included in Later Phases

- **Oversample math and reading teachers in grades 4 through 8 and science teachers in grades 4 and 8.** Because a future statewide evaluation model will almost certainly include the PSSA in some capacity, we recommend including a substantial number of teachers in the pilot for whom value-added can be calculated relative to statewide performance. In conjunction with an overall larger Phase 2 sample, we expect that including more of these teachers will be beneficial for identifying relationships between value-added and observation scores.
- **Sample teachers in other grades based on whether outcome and baseline student scores are available in their districts.** For instance, recruit 3rd- or 11th-grade teachers in districts that administer a beginning-year 4Sight assessment in those grades. For purposes of the pilot, Pennsylvania might also seek to incorporate teachers in non-PSSA grades by recruiting among districts that use the same additional standardized

assessments in both non-PSSA and PSSA grades to assess which assessments are most predictive of student achievement as measured by the PSSA.

- **Recruit teachers for subsequent pilot phases to provide for more variation in the observation measure.** Recruit teachers based on the grades and subjects that they teach and on the availability of outcome and baseline assessment data for their students. This method can best facilitate analyses that support pilot goals about understanding relationships between observation scores and value-added. Recruiting only those teachers who are thought to be effective for the pilot might narrow the distribution of observation scores at the end of the year and, consequently, inhibit the pilot's ability to differentiate between the practices of more and less effective teachers.
- **Oversample middle school principals when a new principal evaluation instrument is developed.** Middle school grades are all tested by the PSSA, unlike grade ranges in elementary and high schools. A sample of middle school principals can thus provide for the cleanest analysis of which principal practices are related to larger growth in student achievement because value-added and rubric-based scores will cover the same grades.

Measurement Issues for Later Phases

- **Assess interobserver agreement and observer drift in the observation data.** Later pilot phases should evaluate the interobserver agreement of the teacher observation data. The pilot should also examine the potential for observer drift in evaluation ratings.
- **Evaluate the quality of data linkages in Pennsylvania's student data.** The validity of effectiveness measures based on student data relies fundamentally on the quality of the underlying data system. Most important would be to assess the quality of the student–teacher–principal links. One way to do this would be to provide Phase 2 teachers with the opportunity to verify the data to their classes. It would also be important to assess how educators such as special education teachers are included in the statewide data system and whether a consistent approach is used across districts.
- **Continue to develop nontest outcomes for school value-added measures.** An additional year of data will enable us to extend the concept of holding power to dropout prevention using data on 12th-grade outcomes. Value-added measures for student attendance and core-course completion could be more viable with an expanded sample of Phase 2 schools—or even better, a statewide sample—provided that data are available and that the information is collected consistently across districts.
- **Continue toward defining the evaluation system's structure.** By the conclusion of the pilot, state leaders will put forward a final evaluation model. This model should define the types of teacher and principal effectiveness measures that will be included and how effectiveness information will be integrated across measures. It should also establish standards for accepting elective effectiveness measures proposed by individual school districts. Continued progress toward defining the overall structure will help focus policy goals for the pilot during Phase 2.

REFERENCES

- Aaronson, D., L. Barrow, and W. Sander. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics*, vol. 25, no. 1, 2007, pp. 95–135.
- Bollinger, C. R., and J. Minier. "On the Robustness of Coefficient Estimates to the Inclusion of Proxy Variables." Working paper. Lexington, KY: University of Kentucky, 2009. <http://gatton.uky.edu/faculty/minier/bollingerminier.pdf>.
- Branch, G., E. Hanushek, and S. Rivkin. "Estimating Principal Effectiveness." Working paper. Dallas, TX: University of Texas at Dallas, 2011.
- Coelli, M., and D. Green. "Leadership Effects: School Principals and Student Outcomes." *Economics of Education Review*, forthcoming.
- Dhuey, E., and J. Smith. "How Important Are School Principals in the Production of Student Achievement?" Working paper. Toronto, ON: University of Toronto, 2011. Available at http://homes.chass.utoronto.ca/~edhuey/datastore/files/docs/dhuey_smith_princ_nov2011.pdf.
- Glazerman, S., S. Loeb, D. Goldhaber, D. Staiger, S. Raudenbush, and G. Whitehurst. "Evaluating Teachers: The Important Role of Value-Added." Washington, DC: Brown Center on Education Policy at Brookings, 2010.
- Goldhaber, D., and D. Chaplin. "Assessing the 'Rothstein Falsification Test': Does It Really Show Teacher Value-Added Models Are Biased?" CEDR working paper 2011-5. Seattle, WA: University of Washington, 2011.
- Goldhaber, D., and M. Hanson. "Using Performance on the Job to Inform Teacher Tenure Decisions." *American Economic Review: Papers & Proceedings*, vol. 100, no. 2, 2010, pp. 250–255.
- Hanushek, E. A., and S. G. Rivkin. "Do Disadvantaged Urban Schools Lose Their Best Teachers?" Washington, DC: The Urban Institute, National Center for Analysis of Longitudinal Data in Education Research, 2008.
- Hill, C. J., H. S. Bloom, A. R. Black, and M. W. Lipsey. "Empirical Benchmarks for Interpreting Effect Sizes in Research." *Child Development Perspectives*, vol. 2, no. 3, 2008, pp. 172–177.
- Jacob, B. A., and L. Lefgren. "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education." *Journal of Labor Economics*, vol. 25, no. 1, 2008, pp. 101–136.
- Kane, T. J., J. E. Rockoff, and D. O. Staiger. "What Does Certification Tell Us About Teacher Effectiveness? Evidence from New York City." *Economics of Education Review*, vol. 27, no. 6, 2008, pp. 615–631.
- Kane, T. J., and D. Staiger. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper No. 14607. Cambridge, MA: National Bureau of Economic Research, 2008.

- Kane, T. J., and D. O. Staiger. "Gathering Feedback for Teaching: Combining High-Quality Observations with Student Surveys and Achievement Gains." Seattle, WA: Bill & Melinda Gates Foundation, Measures of Effective Teaching Project, 2012.
- Koedel, C., and J. Betts. "Does Student Sorting Invalidate Value-Added Models of Teacher Effectiveness? An Extended Analysis of the Rothstein Critique." *Education Finance and Policy*, vol. 6, no. 1, 2011, pp. 18–42.
- Lane, S., and C. Horner. "Pennsylvania Teacher and Principal Evaluation Pilot Report. Final Report Submitted to the Team Pennsylvania Foundation." Pittsburgh, PA: University of Pittsburgh, 2011.
- Lipscomb, S., B. Gill, K. Booker, and M. Johnson. "Estimating Teacher and School Effectiveness in Pittsburgh: Value-Added Modeling and Results, Second Edition." Draft report submitted to the Pittsburgh Public Schools Office of Research, Assessment, and Accountability. Cambridge, MA: Mathematica Policy Research, 2010a.
- Lipscomb, S., B. Teh, B. Gill, H. Chiang, and A. Owens. "Teacher and Principal Value-Added: Research Findings and Implementation Practices." Final report submitted to the Team Pennsylvania Foundation. Cambridge, MA: Mathematica Policy Research, 2010b.
- Milanowski, A. T., S. M. Kimball, and B. White. "The Relationship Between Standards-Based Teacher Evaluation Scores and Student Achievement: Replication and Extensions at Three Sites." Working paper. Madison, WI: Consortium for Policy Research in Education, 2004.
- Pennsylvania Clearinghouse for Education Research. "Teacher Effectiveness: The National Picture and Pennsylvania Context." Philadelphia: Research for Action, 2011.
- Rockoff, J. E., and C. Speroni. "Subjective and Objective Evaluations of Teacher Effectiveness." *American Economic Review: Papers & Proceedings*, vol. 100, no. 2, 2010, pp. 261–266.
- Rothstein, J. "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement." *Quarterly Journal of Economics*, vol. 125, no. 1, 2010, pp. 175–214.
- Sartain, L., S. R. Stoelinga, and E. R. Brown. "Rethinking Teacher Evaluation in Chicago." Chicago: Consortium on Chicago School Research, 2011.
- Schochet, P., and H. Chiang. "Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains." Report submitted to the National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences. Princeton, NJ: Mathematica Policy Research, 2010.
- Team Pennsylvania Foundation. "Teacher Evaluations Highlighted in Governor's Education Agenda." Published online October 11, 2011. Harrisburg, PA: Team Pennsylvania Foundation, 2011. Available at <http://teampa.com>.
- Tyler, J. H., E. S. Taylor, T. J. Kane, and A. L. Wooten. "Using Student Performance Data to Identify Effective Classroom Practices." *American Economic Review: Papers & Proceedings*, vol. 100, no. 2, 2010, pp. 256–260.

APPENDIX A
TECHNICAL SPECIFICATIONS OF THE VAMS

THIS PAGE LEFT BLANK FOR DOUBLE-SIDED PRINTING

TECHNICAL SPECIFICATIONS OF THE VAMS

In this appendix, we provide a technical description of the value-added models (VAMs). In Section A, we describe the empirical specification for the teacher VAMs and discuss strategies that we employ for enhancing the validity and reliability of results. In Section B, we describe the principal transitions model and its key limitations for usability, and then we describe the school VAM.

A. VAMs for Teachers

The following statistical equation describes the primary teacher VAMs:

$$(1) \quad A_{i,j,g,y} = A_{i,j,g-1,y-1}\beta_1 + \beta_2 A_{i,j,g-2,y-2} + X_{i,y}\gamma + C_{i,j,g,y}\pi + T_{i,j,g,y}\delta + Y_y + e_{i,j,g,y}.$$

In the model, $A_{i,j,g,y}$ is an assessment score for student i in subject j (that is, math, reading, or science) in grade g in year y . For example, the elements of $A_{i,j,g,y}$ could be Pennsylvania System of School Assessment (PSSA) math scaled scores for 5th-grade students across the state or Dynamic Indicators of Basic Early Literacy Skills (DIBELS) scores for 2nd-grade students in Allentown and Cornell. The teacher VAMs use only assessments as outcome measures because available nonassessment measures such as student attendance are indicators of school-wide performance.

$A_{i,j,g-1,y-1}$ is a vector of baseline scores for student i from the prior grade.⁴³ We use all available prior-grade scores, which vary in number by grade. For example, 5th graders have three scores from 4th grade (that is, PSSA scores in math, reading, and science); however, 7th graders have only two scores from 6th grade (that is, PSSA scores in math and reading). We use prior-grade scores for grade repeaters as well, except that their prior-grade scores come from two years previously. For the outcomes in which a statewide comparison is not possible, $A_{i,j,g-1,y-1}$ can be modified to $A_{i,j,g,y}$ in Equation (1) if a fall baseline score from grade g is used instead of an end-of-year score from the prior grade. Baseline scores are measured with error. This can bias their coefficient estimates but need not create substantial bias for other coefficient estimates, especially when multiple baseline scores are used as control variables (Bollinger 2009).

When feasible, we also control for a student's same-subject score from two prior grades ago (indicated by $A_{i,j,g-2,y-2}$). Such scores cannot be included in 4th-grade models because state testing begins in 3rd grade. When subjects are not assessed in consecutive grades, as in science and writing, we use a different subject score for $A_{i,j,g-2,y-2}$ instead. For example, the 8th-grade science VAM controls for 6th-grade math and the 8th-grade writing VAM controls for 6th-grade reading. Finally, in VAMs for high school teachers, we use all available 8th-grade PSSA scores as $A_{i,j,g-2,y-2}$ because of the unavailability of assessment data in grades 9 and 10. In all cases, we include linear and quadratic functions of the baseline scores in $A_{i,j,g-1,y-1}$ and $A_{i,j,g-2,y-2}$ to allow for nonlinear relationships between current and prior achievement. This can help address potential issues related to test ceiling effects.

$X_{i,y}$ is a set of control variables for observable student characteristics, $C_{i,j,g,y}$ is a set of classroom-level characteristics, Y_y is a set of year indicator variables, $e_{i,j,g,y}$ is the error term, and $T_{i,j,g,y}$ is a set of

⁴³ Our VAMs let the data determine the persistence of a teacher's (or a school's) effect on the performance of students in a subsequent year. If the model included only one baseline score control, the coefficient estimate on the baseline score variable would be the degree of empirically determined persistence.

teacher indicator variables. The teacher variables indicate whether a student was taught by a teacher in a given grade, subject, and year, according to Pennsylvania’s longitudinal student data. Teachers can be linked to students only if they are listed as the teacher for a course that a student is listed as taking during the school year. We represent the teacher variables as binary 0/1 indicators. Ordinarily, we would use a dosage model that allows for fractional values that sum to one for each student in the event that students are taught by multiple teachers in a subject during the year. Due to the amount of data that enter the statewide teacher VAMs, computational limitations inhibited us from using a dosage model in Phase 1.⁴⁴ These constraints are less intensive in the school VAMs—because there are fewer principals than teachers—and, consequently, we use a dosage model there.⁴⁵

The coefficients in β_1 , β_2 , γ , π , and δ are the estimated relationships between students’ assessment scores and each respective variable, controlling for the other factors in the model.⁴⁶ There is one δ coefficient for each teacher in the VAM, where each identifies a teacher’s contribution to student learning—the extent to which the actual achievement of students tends to be above or below what is expected for the average teacher. We define the average VAM score to be a zero value, but this does not mean that student learning is zero for the teacher with the average VAM score. Rather, it means that a positive VAM score represents above-average teacher performance and a negative VAM score represents below-average teacher performance.

The reference point for determining the average teacher contribution depends on the sample of teachers in the model. If the data include students and teachers across Pennsylvania (for example, as in PSSA math grade 5), the VAM estimates would be calculated relative to the contribution of the average teacher in Pennsylvania in that grade and subject. If the data include only students and teachers from a particular grade in pilot districts (for example, 2nd-grade DIBELS in Allentown and Cornell), the VAM scores would be calculated relative to the contribution of the average teacher in that grade and subject at the particular pilot district(s). Our VAMs include both statewide and Phase 1 district samples, with the primary determining factor being the level at which the outcome and baseline data are available. Appendix B contains detailed information on data elements and samples.

In estimating the VAMs, we take the following additional steps to enhance the accuracy and reliability of the results:

- **Convert assessment scores into standard scores.** VAM estimates reported in assessment units (for example, PSSA scaled score points) are not comparable across assessments, grades, subjects, or years. Before estimating a VAM, we map assessment scores to a standard measure, called a z-score, by subtracting the average annual score from individual scores (by grade and school year) and then dividing by the standard

⁴⁴ For example, the statewide teacher VAM for 5th-grade math takes 14 hours to run on a Windows 7 personal computer using Stata 11 2-core Multi-Processor (MP) Edition when storing the teacher variables as binary indicators. The memory needed to estimate the VAM would quadruple if teacher variables were allowed to take fractional values.

⁴⁵ In the school models, the dosage is split evenly across principals leading all schools a student attended during the year. Because Pennsylvania’s longitudinal data system did not include attendance data for us to use in this report, we are not able to account for any time that students are enrolled outside Pennsylvania public schools during the academic year. The approximate run time for these models is one hour using Windows 7 and Stata 11 2-core MP Edition.

⁴⁶ The standard errors adjust for clustering of observations by student.

deviation of scores. Expressing scores in this way enables us to interpret above-average scores in terms of how close to the average most students tended to fall. Appendix C reports VAM results in *z*-score terms. Estimates are converted back to score points for the reporting of results in Chapters III and V.

- **Adjust estimates based on their precision.** Consistent with the research literature, the VAMs use a procedure known as empirical Bayes estimation or shrinkage to address the fact that, among teachers with the same level of true performance, those with fewer students in the estimation sample face a greater likelihood that their students happen, by chance, to have atypically high or low learning growth driven by other factors. Teachers with fewer students—that is, those with less precise estimates—will tend to be over-represented at both the high and low ends of the estimated performance distribution for reasons other than their effectiveness. Shrinkage adjustments account for the fact that estimates with greater precision carry greater strength of information about teachers’ true performance levels. The adjusted estimate is a weighted average of the individual initial estimate and the mean estimate across teachers, with more precise initial estimates receiving greater weight. In essence, teachers are assumed to be average in performance until evidence justifies a different conclusion. To further minimize the risk of making erroneous conclusions on the basis of imprecise estimates, we limit analyses to teachers who taught more than 10 students during the year.
- **Incorporate observations on students across multiple years.** Our primary teacher VAMs include students taught by a teacher in the past three years—that is, the number of current and prior student cohorts who contribute to the estimate—whenever three years of data are available. Multicohort VAM estimates are less prone to random fluctuations that stem from a teacher being assigned a few students with unusually high or low learning growth. The VAM estimates can therefore detect performance differences with greater reliability. Researchers have also found that multicohort VAMs are less prone to systematic fluctuations in scores as well, meaning that they might have greater validity, too. Multicohort VAMs also better distinguish teacher effects from the effects of students’ peers in the classroom, which is impossible to determine in a single-cohort model unless teachers teach in multiple classrooms during the year. We report estimates from multicohort VAMs as follows. First, we estimate the VAM based on all students and teachers in the included school years. We then restrict the resulting set of teacher estimates to those for teachers with students in the outcome subject and grade in the latest year covered by the VAM and with more than 10 students overall during the sample period. We next apply the empirical Bayes calculation to this subset of estimates and then center the resulting effectiveness measures on a zero value. We report estimates for teachers with students from all three prior years, rather than for all teachers, including those with fewer than three years of teaching data. In this report, we also provide estimates from several one-cohort VAMs, for the purpose of comparing the results.
- **Adjust some teacher effectiveness estimates for district and school factors.** In a diagnostic analysis, we adjust the teacher effectiveness estimates for district and school factors by subtracting the mean teacher effect in each district or school from each individual estimate. This adjustment—which has the same effect as adding district or school fixed effects to the VAM itself—has the potential to provide better controls for district- or school-level influences on teacher performance that are external to teachers. We do not make this adjustment to our primary teacher VAMs, however, because it

means that teachers are compared only with other teachers in their same district or school. It might also under-represent true differences in teacher effectiveness to the extent that highly effective (or ineffective) teachers tend to cluster together.

B. VAMs for Principals

1. Principal Transitions Model

The principal transitions model is based on leadership transitions between principals within the same school (Branch et al. 2011; Dhuey and Smith 2011; Coelli and Green, forthcoming). Thus, this model is fundamentally based on comparing principals who lead the same school (at different times) during the analysis period. As long as the mix of existing abilities within a school's teaching staff remains constant during the analysis period, comparing principals who have led the same school will effectively remove the influence of existing teacher abilities from the estimate of a principal's value-added.⁴⁷ A related advantage of this method is that it also removes the influence of other school-specific factors that remain constant during the analysis period, such as neighborhood quality.

The principal transitions model can be represented by the following equation:

$$(2) A_{i,j,g,y} = A_{i,j,g-1,y-1}\beta_1 + \beta_2 A_{i,j,g-2,y-2} + X_{i,y}\gamma + P_{i,j,g,y}\varphi + S_{i,j,g,y}\delta + Y_y + e_{i,j,g,y},$$

where $P_{i,j,g,y}$ is a set of principal variables, $S_{i,j,g,y}$ is a set of school variables, and all other variables are defined as in Equation (1). The coefficients, φ , on the principal variables are the estimates of principal effectiveness. Importantly, the school variables control for any school-specific influences on student achievement—potentially including existing teacher abilities—that do not change during the analysis period.⁴⁸ As discussed in Chapter V, the principal transitions model is likely to have a high degree of internal validity, but estimates from this model can be compared only among principals who have led schools connected by leadership transitions during the analysis period. For this reason, we do not consider this model to be a viable method for generating effectiveness measures in actual principal evaluations.

We use the principal transitions model as a point of comparison to an alternative model that bundles together the combined value-added of educators at the school. This latter model, which we call the school VAM, does not control for the mix of existing teacher abilities or other school factors outside a principal's control but can be applied to all principals in Pennsylvania.

⁴⁷ Even if the composition of the teaching staff changes concurrently with the leadership transition, the existing abilities of teachers remaining at the school will be removed from the principal effectiveness estimates.

⁴⁸ An alternative way to control for existing teacher abilities is to include teacher variables, rather than school variables, in Equation (2). This approach estimates principals' effects by examining how teacher value-added changes when a teacher transitions between principals. We chose not to use this approach because teachers' transitions between principals often occur as a result of teacher transfers between schools; a change in the value added of transferring teachers could be due to differences in school-specific factors other than just the change in principal.

2. School Model

The school VAM that we estimate is conceptually similar to the teacher VAM discussed previously and has the same basic empirical specification:

$$(3) A_{i,j,g,y} = A_{i,j,g-1,y-1}\beta_1 + \beta_2 A_{i,j,g-2,y-2} + X_{i,y}\gamma + P_{i,j,g,y}\delta + Y_y + e_{i,j,g,y}.$$

Most of the variables are defined in the same way as in Equation (1). For the school VAM, $P_{i,j,g,y}$ now denotes a set of principal variables, and hence the δ terms identify total contributions to student learning by educators at a principal's school(s), including the principal's own contribution. Most school VAMs control for schools rather than for principals. We chose to use principal variables because the pilot's focus is on principal evaluation; thus, the estimates measure the school value-added at a principal's school or schools. A conceptual distinction between school and teacher VAMs is that the baseline score in a teacher VAM is measured before each student entered the teacher's class, whereas the baseline score in a school VAM typically comes from the preceding year regardless of whether the student attended a school led by the same or a different principal in that year.

In applying school VAMs, we also use several nonassessment outcomes, such as attendance rates and rates of staying in school (see Appendix B). The VAMs are generally the same for nonassessment outcomes as for assessment outcomes.⁴⁹ However, one difference is that the baseline achievement measures might capture a somewhat different aspect of achievement than the outcome measure. For instance, when evaluating high schools' impacts on students' rates of staying in school, we control for students' assessment scores from 8th grade. Although the elements of student performance captured by these two variables are not identical, 8th-grade test scores might nevertheless be strongly predictive of staying in school, and controlling for these scores will reduce the likelihood that comparisons across schools could be biased by differences in their students' incoming risk of dropping out.

We apply all of the aforementioned steps for enhancing the validity and reliability of VAMs, including standardization of assessment scores, use of dosage to apportion responsibility for the outcomes of mobile students, adjusting VAM estimates for their level of precision, and exploring the implications of using multiple years of student growth data. For the school VAM approach, if we incorporate multiple years of student growth data, then there are principals who work in multiple schools during the analysis period. The principals' final VAM scores will be averaged over all of the relevant school–principal combinations.⁵⁰

Despite the similarities between school and teacher VAMs, there are two main substantive differences between these models. The first was mentioned earlier: the baseline scores used for most grades in the school models are not pretreatment measures—that is, measures from a period before students' enrollment in a specified school. Except in entry grades for middle and high schools,

⁴⁹ Although most VAMs in practice have focused on assessment outcomes, the methodology is very similar for nonassessment outcomes, so we also apply the terminology of *value-added models* to these latter methods as well.

⁵⁰ Several schools are led by more than one principal at the same time. If a group of principals is observed to have led the same school at the same time in all years of the VAM, then they are assigned a single effectiveness estimate.

students are generally served by the same principal both in the current grade (that is, the grade to which a set of VAM estimates apply) and in the prior grade, when baseline scores are measured. In contrast, baseline scores in the teacher models are typically pretreatment, given that teachers generally do not teach the same students in multiple grades. In prior analysis for the Pittsburgh Public Schools, we found that school VAM estimates were similar regardless of whether prior-grade baseline scores or pre-entry baseline scores were used (Lipscomb et al. 2010a). Our models therefore use prior-grade baseline controls because this approach can be applied consistently across grades and schools and allows the inclusion of a larger number of students.

A related limitation of the school VAM applies specifically to elementary schools. Because PSSA assessments begin in grade 3, VAMs relying solely on PSSA scores for both baseline and outcome measures provide no information about the value-added produced from kindergarten entry through the end of 3rd grade. In other words, unless school districts administer additional assessments in the early elementary grades (as some do), four years of schooling are invisible to VAMs that rely on tests beginning in grade 3.

APPENDIX B

DATA SOURCES AND SAMPLE CHARACTERISTICS

THIS PAGE LEFT BLANK FOR DOUBLE-SIDED PRINTING

DATA SOURCES AND SAMPLE CHARACTERISTICS

In this appendix, we describe the data that are used in the report. In Section A, we list the source for each data element. In Section B, we compare the characteristics of students in the pilot districts to those of students attending other school districts. In Section C, we provide the baseline and analysis sample sizes for students, teachers, and principals.

A. Data Sources

Nearly all data for this report come from the statewide Pennsylvania Department of Education (PDE) data and from the four Phase 1 pilot districts. In Table B.1, we summarize data elements by source. PDE's Pennsylvania Information Management System (PIMS) is the source for student characteristics and most information on linkages between students and their teachers, principals, and schools. The exception is for student–teacher links for students attending Pittsburgh Public Schools, which are largely missing in PIMS. For Pittsburgh, we used the district's own records instead.⁵¹ State assessment data come from PDE's Bureau of Assessment and Accountability (BAA) and PIMS. The remaining variables were obtained from pilot districts directly.

⁵¹ The missing data pertain to the student–course records (template 490). We identified the missing data problem for Pittsburgh through our involvement in Pittsburgh's Empowering Effective Teachers project and were able to use the district's own records because Mathematica already had access to them.

Table B.1. Data Sources

Agency	Data Element	School Years
Pennsylvania (PIMS)	Student background (template 0320)	2007-2008 to 2010-2011
	Student-course links (template 0490)	2008-2009 to 2010-2011
	Teacher-course links (template 0410)	2008-2009 to 2010-2011
	Principal-school links (template 0630)	2008-2009 to 2010-2011
	Course description (template 0310)	2008-2009 to 2010-2011
	PSSA scaled scores (all subjects)	2007-2008 to 2009-2010
Pennsylvania (BAA)	PSSA scaled scores (all subjects)	2010-2011
	PSSA-M scaled scores (all subjects)	2009-2010 to 2010-2011
Allentown	Student attendance	2009-2010 to 2010-2011
	Core courses attempted and passed	2009-2010 to 2010-2011
	4Sight scores	May 2010, Sept. 2010
	DIBELS scores	May 2010, Sept. 2010, May 2011
	Progress assessment scores	May 2010, Oct. 2010, May 2011
	Teacher observation rubric scores	Spring 2011
Cornell	Core courses attempted and passed	2009-2010 to 2010-2011
	4Sight scores	May 2010, Sept. 2010
	DIBELS scores	April 2010, Sept. 2010, April 2011
	Teacher observation rubric scores	Spring 2011
Mohawk	Student attendance	2009-2010 to 2010-2011
	Core courses attempted and passed	2009-2010 to 2010-2011
	4Sight scores	May 2010, Sept. 2010
	Teacher observation rubric scores	Spring 2011
Northwest Tri-County	Student attendance	2009-2010 to 2010-2011
	Teacher observation rubric scores	Spring 2011

BAA = Bureau of Assessment and Accountability; DIBELS = Dynamic Indicators of Basic Early Literacy Skills; PIMS = Pennsylvania Information Management System; PSSA = Pennsylvania System of School Assessment; PSSA-M = PSSA-Math.

B. Descriptive Statistics

In Table B.2, we report the baseline sample means for several student characteristics that are used in the VAMs. These data come from the PIMS for 2010–2011. The first two columns show averages across all students in nonpilot and pilot districts. The remaining four columns apply to individual pilot districts.

Table B.2. Descriptive Statistics on Student Characteristics, 2010-2011

Variable	Nonpilot PA Average	Pilot District Average	Pilot Districts			
			Allentown	Cornell	Mohawk	Northwest
Math PSSA, Grade 5 (scaled score)	1,468.5	1,402.8	1,395.3	1,420.6	1,520.1	*
Math PSSA, Grade 11 (scaled score)	1,373.8	1,210.6	1,193.9	1,261.5	1,372.7	1,110.5
Reading PSSA, Grade 8 (scaled score)	1,505.7	1,322.2	1,289.3	1,451.9	1,470.2	1163.2
Science PSSA, Grade 4 (scaled score)	1,452.9	1,329.6	1,316.4	1,375.7	1,506.0	*
Female (%)	48.4	47.0	47.5	50.3	47.7	31.0
White (%)	71.3	26.2	16.4	70.7	97.3	68.4
African American (%)	15.8	16.2	17.3	14.2	*	24.3
Hispanic (%)	8.0	54.7	63.6	2.1	*	4.6
Asian and Pacific Islander (%)	3.2	1.4	1.5	*	0.6	*
Multiracial or Other Race/Ethnicity (%)	1.7	1.6	1.1	13.1	1.2	1.9
Free Lunch Eligibility (%)	34.0	70.3	75.3	52.4	28.9	54.6
Reduced-Price Lunch Eligibility (%)	5.6	8.9	9.3	9.6	5.9	5.2
English-Language Learner (%)	2.8	10.1	11.7	*	*	1.4
Special Education (%)	16.0	15.5	13.0	17.2	11.4	78.0
Grade Repeater (%)	2.6	5.1	5.2	1.9	1.0	14.1
Number of Students (1,000's)	1,617.6	22.1 ^a	18.9	0.7	1.6	0.9

Source: Mathematica calculations based on Pennsylvania student data.

Note: **Bold** indicates a statistically significant mean difference between the nonpilot district average and the pilot district average at the 5 percent level. Descriptive statistics may differ for analysis samples based on the characteristics of the students included in each model. Pilot districts account for 1.3 percent of all students in the state.

* Indicates that a sample mean is withheld because it includes 10 or fewer students.

^a The number 22.1 refers to the sum of students across the four pilot districts (in thousands).

As indicated in the table, the sample characteristics of students in the four pilot districts differ from the characteristics of students in other Pennsylvania districts in terms of most of the observable demographic variables. That is, the Phase 1 districts are not representative of the state in terms of their student populations. The pilot district averages (column 2) are primarily influenced by Allentown, given its size compared with the other pilot districts, but each pilot district has important differences with nonpilot districts in terms of the characteristics of its students. Given these differences in baseline characteristics, we recommend interpreting findings based on analyses of pilot data as suggestive for future work involving larger Pennsylvania samples but not as representative of Pennsylvania students, teachers, or schools.

C. Baseline and Analysis Sample Sizes

In Table B.3, we describe the baseline and analysis samples for students in the teacher and school VAMs. By baseline sample, we mean the number of students that have a nonmissing value of the outcome variable for a particular VAM. The analysis sample includes the subset of those students with nonmissing data on prior scores, student characteristics, and teacher–principal links.

Table B.3. Baseline and Analysis Student Sample Sizes for Teacher and School VAMs, by Outcome

Outcome	Description of the Sample	Number of Cohorts in the Main Model	Number of Students with Nonmissing Values of the Outcome Measure	Number of Students in the Analysis Sample for Teacher VAMs	Number of Students in the Analysis Sample for School VAMs
Math PSSA, Grade 3	A,C,M	1	1,676	1,291	1,309
Math PSSA, Grade 4	PA	3	361,916	304,013	337,017
Math PSSA, Grade 5	PA	3	364,180	287,889	320,734
Math PSSA, Grade 6	PA	3	364,192	295,247	317,807
Math PSSA, Grade 7	PA	3	367,897	311,298	322,251
Math PSSA, Grade 8	PA	3	373,308	311,991	326,470
Math PSSA, Grade 11	A,C,M (T)	1	1,450	725	--
Math PSSA, Grade 11	PA (P)	2	242,819	--	196,976
Reading PSSA, Grade 3	A,C,M	1	1,666	1,291	1,310
Reading PSSA, Grade 4	PA	3	361,376	305,736	336,932
Reading PSSA, Grade 5	PA	3	363,678	287,611	319,966
Reading PSSA, Grade 6	PA	3	363,717	300,605	317,023
Reading PSSA, Grade 7	PA	3	367,273	312,985	321,469
Reading PSSA, Grade 8	PA	3	372,676	315,056	325,645
Reading PSSA, Grade 11	A,C,M (T)	1	1,450	730	--
Reading PSSA, Grade 11	PA (P)	2	242,371	--	196,834
Writing PSSA, Grade 5	PA	3	361,100	286,229	318,422
Writing PSSA, Grade 8	PA	3	369,574	313,409	323,880
Writing PSSA, Grade 11	A,C,M (T)	1	1,382	718	--
Writing PSSA, Grade 11	PA (P)	2	237,615	--	194,537
Science PSSA, Grade 4	PA	3	360,596	290,214	336,015
Science PSSA, Grade 8	PA	3	370,052	307,544	324,493
Science PSSA, Grade 11	A,C,M (T)	1	1,359	693	--
Science PSSA, Grade 11	PA (P)	2	237,189	--	193,955
Math Progress Assess., Grade 2	A	1	1,176	865	870
Writing Progress Assess., Grade 1	A	1	1,225	899	903
Writing Progress Assess., Grade 2	A	1	1,212	894	900
Writing Progress Assess., Grade 3	A	1	1,126	870	874
DIBELS (NWF), Grade 1	A	1	1,434	968	987
DIBELS (PSF), Grade 1	A	1	1,434	968	987
DIBELS (ORF), Grade 2	A,C	1	1,055	831	836
Attendance Rate, Grades 4-12	A,M,N (P)	1	11,787	--	9,339
Holding Power, Grade 9	PA (P)	3	294,343	--	244,709
Holding Power, Grade 10	PA (P)	2	292,921	--	229,602
Holding Power, Grade 11	PA (P)	1	136,785	--	107,972

Source: Mathematica calculations based on Pennsylvania student data and student data from pilot districts.

Note: Sample sizes refer to student-school year observations. Students are counted more than once if they appear in a sample in multiple years. The analysis sample for an outcome measure is the sample that is used for estimating a VAM.

A = Allentown; C = Cornell; M = Mohawk; N = Northwest; P = principal; PA = Pennsylvania; T = teacher.

-- indicates that sample size information is not available because a model was not estimated.

As described in Appendix A, the samples differ by outcome because some VAMs can include students from across the Commonwealth, whereas others are limited to pilot districts. The sample sizes also differ based on the number of student cohorts (up to three) that can be included. On average, analysis samples for school VAMs are 11 percent smaller than baseline samples in grades 4 to 8. Analysis samples for teacher VAMs in grades 4 to 8 are 17 percent smaller than baseline samples, on average. The primary source of sample reduction in the school VAMs is students who are missing at least one prior test. In contrast, only a small number of students are excluded for

other reasons. Sample exclusions in the teacher VAMs reflect two primary factors: students with missing score history (that is, the same reason as in the school VAMs) and missing teacher links. As described earlier in this appendix, we have partially addressed this source of sample loss by using Pittsburgh’s own records on student–teacher links. However, as indicated by the difference in sample sizes between the last two columns in Table B.3, Pennsylvania can increase the number of students and teachers in the VAMs for future years by improving the quality of student–teacher links.

In Table B.4, we report the number of teachers and principals with VAM estimates by outcome and by whether a VAM includes a single cohort or multiple cohorts of students. Although multicohort VAMs include more students, fewer estimates are reported because we output results only for teachers or principals linked to students during the entire multiyear period.⁵² For example, in the three-cohort VAM for 5th-grade math based on the PSSA, we report estimates for teachers with 5th-grade math students in 2008–2009, 2009–2010, and 2010–2011. We do not report estimates from that VAM for other teachers, such as new 5th-grade math teachers in 2010–2011. In future years if desired, it would be possible to report all estimates from a multicohort VAM regardless of the number of cohorts of students it includes for a given teacher.

⁵² As described in Appendix A, we also report estimates only for teachers and principals who can be linked with more than 10 students in the analysis because estimates based on very small numbers of students are likely to have low precision. This requirement applies for both single-cohort and multicohort VAMs.

Table B.4. Number of Teachers and Principals with VAM Estimates Reported from Multicohort and Single-Cohort VAMs

Outcome	Description of the Sample	Number of Cohorts in the Main Model	Teachers with VAM Estimates		Principals with VAM Estimates	
			Single Cohort	Multi-Cohort	Single Cohort	Multi-Cohort
Math PSSA, Grade 3	A,C,M	1	69	--	16	--
Math PSSA, Grade 4	PA	3	--	3,075		1,167
Math PSSA, Grade 5	PA	3	4,103	2,836	1,336	1,079
Math PSSA, Grade 6	PA	3	--	1,994	--	758
Math PSSA, Grade 7	PA	3	--	1,403	--	581
Math PSSA, Grade 8	PA	3	1,685	1,471	--	580
Math PSSA, Grade 11	A,C,M (T)	1	29	--	--	--
Math PSSA, Grade 11	PA (P)	2	--	--	626	577
Reading PSSA, Grade 3	A,C,M	1	69	--	16	--
Reading PSSA, Grade 4	PA	3	--	3,126	--	1,167
Reading PSSA, Grade 5	PA	3	4,167	2,907	--	1,079
Reading PSSA, Grade 6	PA	3	--	2,446	--	758
Reading PSSA, Grade 7	PA	3	--	1,749	--	581
Reading PSSA, Grade 8	PA	3	1,916	1,717	755	580
Reading PSSA, Grade 11	A,C,M (T)	1	19	--	--	--
Reading PSSA, Grade 11	PA (P)	2	--	--	--	578
Writing PSSA, Grade 5	PA	3	--	2,908	--	1,077
Writing PSSA, Grade 8	PA	3	--	1,711	--	579
Writing PSSA, Grade 11	A,C,M (T)	1	19	--	--	--
Writing PSSA, Grade 11	PA (P)	2	--	--	--	574
Science PSSA, Grade 4	PA	3	4,187	2,854	1,427	1,166
Science PSSA, Grade 8	PA	3	1,313	1,035	--	581
Science PSSA, Grade 11	A,C,M (T)	1	17	--	--	--
Science PSSA, Grade 11	PA (P)	2	--	--	--	574
Math Progress Assess., Grade 2	A	1	46	--	12	--
Writing Progress Assess., Grade 1	A	1	50	--	13	--
Writing Progress Assess., Grade 2	A	1	47	--	13	--
Writing Progress Assess., Grade 3	A	1	48	--	14	--
DIBELS (NWF), Grade 1	A	1	53	--	13	--
DIBELS (PSF), Grade 1	A	1	53	--	13	--
DIBELS (ORF), Grade 2	A,C	1	44	--	15	--
Attendance Rate, Grades 4-12	A,M,N	1	--	--	29	--
Holding Power, Grade 9	PA	3	--	--	--	612
Holding Power, Grade 10	PA	2	--	--	--	612
Holding Power, Grade 11	PA	1	--	--	690	--

Source: Mathematica calculations based on Pennsylvania student data.

Note: Teachers and principals are included in multiple rows if they have students in multiple grades.

A = Allentown; C = Cornell; M = Mohawk; N = Northwest; P = principal; PA = Pennsylvania; T = teacher.

-- indicates that sample size information is not available because a model was not estimated.

APPENDIX C

TECHNICAL RESULTS FROM VALUE-ADDED ANALYSES

THIS PAGE LEFT BLANK FOR DOUBLE-SIDED PRINTING

TECHNICAL RESULTS FROM VALUE-ADDED ANALYSES

In this appendix, we provide the full technical results from value-added models (VAMs) applied for estimating teacher and principal effectiveness. The tables are sequenced to correspond with the presentation of findings in Chapters III and V.

Table C.1. Sample Characteristics of Outcome Measures and Teacher VAMs Based on State Samples

Outcome	Distribution of Student-Level Outcome Variable		Characteristics of Teacher VAMs and Estimates			
	Mean	Standard Deviation	Adjusted R-Squared	85th Minus 50th Percentile of VAM Estimates (in z-score units)	Mean Standard Error (in z-score units)	Percentage of VAM Estimates that Are Statistically Distinguishable from the Average
Math PSSA, Grade 4	1,475	221	0.68	0.23	0.07	51.6
Math PSSA, Grade 5	1,479	223	0.78	0.20	0.06	52.0
Math PSSA, Grade 6	1,501	237	0.79	0.19	0.05	58.6
Math PSSA, Grade 7	1,502	239	0.81	0.17	0.04	61.7
Math PSSA, Grade 8	1,454	220	0.81	0.16	0.04	59.2
Reading PSSA, Grade 4	1,387	212	0.68	0.16	0.07	38.7
Reading PSSA, Grade 5	1,351	209	0.73	0.16	0.06	38.9
Reading PSSA, Grade 6	1,397	222	0.74	0.11	0.06	32.7
Reading PSSA, Grade 7	1,428	214	0.74	0.11	0.06	32.2
Reading PSSA, Grade 8	1,519	247	0.75	0.09	0.05	30.5
Writing PSSA, Grade 5	1,342	258	0.53	0.30	0.09	57.6
Writing PSSA, Grade 8	1,412	262	0.55	0.21	0.07	48.7
Science PSSA, Grade 4	1,462	182	0.66	0.22	0.07	49.8
Science PSSA, Grade 8	1,324	198	0.73	0.14	0.04	57.2

Source: Mathematica calculations based on Pennsylvania student data.

Notes: Findings on PSSA scores are based on three-cohort models with statewide samples of teachers and students. For three-cohort models, the sample of teachers consists of those who served as teachers in every year from 2008-2009 to 2010-2011. Teachers' VAM estimates are based on students in their classrooms at any time during the specified analysis periods. One z-score unit is equal to one standard deviation of student outcomes.

PSSA = Pennsylvania System of School Assessment; VAM = value-added model.

Table C.2. Estimated Regression Coefficients from Selected Three-Cohort PSSA Teacher VAMs

	Math, Grade 5 (in z-score units)		Reading, Grade 8 (in z-score units)		Science, Grade 4 (in z-score units)	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Math, Prior Grade	0.4888	254.98	0.1877	114.21	0.2794	160.59
Math, Prior Grade ^2	-0.0070	-5.04	-0.0003	-0.25	-0.0393	-29.16
Reading, Prior Grade	0.0722	41.47	0.3767	195.30	0.4624	252.01
Reading, Prior Grade ^2	0.0311	25.07	-0.0287	-21.25	0.0270	18.74
Science, Prior Grade	0.1263	72.57				
Science, Prior Grade ^2	0.0184	15.00				
Outcome Subject, 2-Prior Grade	0.2244	133.96	0.2718	149.51		
Outcome Subject, 2-Prior Grade ^2	-0.0263	-21.15	-0.0130	-9.93		
Free Meals	-0.0197	-11.63	-0.0216	-11.36	-0.0460	-19.94
Reduced-Price Meals	-0.0069	-7.39	-0.0108	-11.36	-0.0148	-12.86
English-Language Learner	0.0055	2.53	-0.0003	-0.11	-0.0098	-3.10
Specific Learning Disability	-0.0350	-30.74	-0.0221	-17.15	0.0177	12.39
Speech or Language Impairment	-0.0012	-1.40	-0.0020	-2.29	-0.0103	-9.20
Emotional Disturbance	-0.0113	-9.20	-0.0096	-6.51	0.0041	1.79
Intellectual Disability	-0.0023	-1.13	-0.0076	-4.95	0.0003	0.12
Autism	-0.0078	-6.34	-0.0058	-4.66	0.0012	0.52
Physical/Sensory Disability	-0.0036	-1.56	-0.0052	-3.01	-0.0045	-1.84
Other Impairment	-0.0178	-17.89	-0.0115	-11.53	-0.0041	-3.39
Mobility	-0.0069	-7.07	-0.0044	-4.45	-0.0016	-1.37
Grade Repeater	0.0045	3.39	0.0076	5.53	0.0162	12.20
Behind	-0.0006	-0.55	-0.0054	-4.92	-0.0049	-3.68
Age	-0.0247	-24.69	-0.0029	-2.78	-0.0013	-1.03
PSSA-Modified (outcome)	0.0711	77.03	0.0534	59.75		
PSSA-Modified (baseline)	-0.0242	-25.92	-0.0034	-3.44		
Female	-0.0064	-6.83	0.0701	73.12	-0.0698	-62.06
Asian/Pacific Islander	0.0263	25.57	0.0099	9.69	0.0057	4.78
African American	-0.0053	-4.07	0.0079	6.06	-0.0486	-30.12
Hispanic	0.0010	0.82	0.0047	4.01	-0.0229	-15.94
Other Race/Ethnicity	-0.0012	-1.28	0.0009	0.88	-0.0044	-3.74
Class Avg.: Free Meals	-0.0059	-1.74	-0.0414	-15.37	-0.0236	-5.30
Class Avg.: Reduced-Price meals	0.0015	1.05	-0.0089	-7.24	-0.0027	-1.48
Class Avg.: English-Language Learner	0.0047	2.14	-0.0088	-5.11	0.0015	0.51
Class Avg.: Special Education	-0.0004	-0.25	-0.0232	-12.66	0.0016	0.79
Class Avg.: Female	0.0027	2.10	0.0107	9.46	0.0016	1.01
Class Avg.: Asian/Pacific Islander	-0.0015	-0.71	0.0126	7.61	-0.0072	-2.84
Class Avg.: African American	0.0009	0.20	-0.0029	-0.90	-0.0408	-7.52
Class Avg.: Hispanic	-0.0029	-0.77	-0.0027	-1.00	-0.0234	-4.99
Class Avg.: Other Race/Ethnicity	0.0016	1.04	-0.0002	-0.20	-0.0093	-4.59
Class Size	-0.0091	-2.92	0.0097	4.82	-0.0065	-1.56
Class Size x Emotional Disturbance	-0.0018	-2.13	-0.0002	-0.19	-0.0045	-2.11
Class Size x Intellectual Disability	-0.0007	-0.32	-0.0010	-0.65	-0.0045	-1.61
Class Size x Autism	-0.0006	-0.66	0.0018	1.67	-0.0046	-2.04
Class Size x Physical/Sensory	0.0007	0.29	0.0029	1.76	0.0020	0.87
Class Size x Free Meals	-0.0036	-2.23	-0.0060	-3.11	-0.0026	-1.14
Class Size x English-Language Learner	0.0001	0.03	0.0026	0.93	0.0018	0.60

Source: Mathematica calculations based on Pennsylvania student data.

Notes: T-statistics that exceed 1.96 in absolute value are statistically significant at the 5 percent level. All variables are expressed in standard deviation units with a mean of zero. These regressions include indicator variables for each teacher and school year, and no intercept. One z-score unit is equal to one standard deviation of student outcomes.

Table C.3. Effect Sizes for Three-Cohort Teacher VAMs Expressed in Terms of One Year of Learning

Grade	85th Minus 50th Percentile of VAM Estimates in Z-score Units		Average Annual Gain on Nationally Normed Tests in Z-score Units (from Hill et al. 2008)		85th Minus 50th Percentile of VAM Estimates in Terms of One Year of Learning for a Typical Student	
	Math	Reading	Math	Reading	Math	Reading
4	0.23	0.16	0.52	0.36	0.44	0.44
5	0.20	0.16	0.56	0.40	0.36	0.40
6	0.19	0.11	0.41	0.32	0.46	0.34
7	0.17	0.11	0.30	0.23	0.57	0.48
8	0.16	0.09	0.32	0.26	0.50	0.35

Source: Table C.1 and Hill et al. (2008) Table 1.

Note: One z-score unit is equal to one standard deviation of student outcomes. A difference in VAM estimates expressed in terms of one year of learning equals the difference expressed in z-score units divided by the average annual gain in z-score units.

VAM = value-added model.

Table C.4. Sample Characteristics of Outcome Measures and Teacher VAMs Based on Phase 1 Samples

Outcome	Distribution of Student-Level Outcome Variable		Characteristics of Teacher VAMs and Estimates			
	Mean	Standard Deviation	Adjusted R-Squared	85th Minus 50th Percentile of VAM Estimates (in z-score units)	Mean Standard Error (in z-score units)	Percentage of VAM Estimates that Are Statistically Distinguishable from the Average
DIBELS (NWF), Grade 1 ^a	39	22	0.42	0.36	0.16	41.5
DIBELS (PSF), Grade 1 ^a	40	15	0.68	0.50	0.12	56.6
DIBELS (ORF), Grade 2 ^a	85	37	0.81	0.13	0.09	18.2
Writing Progress Assessment, Grade 1 ^b	76	14	0.52	0.34	0.15	38.0
Writing Progress Assessment, Grade 2 ^b	76	15	0.59	0.35	0.13	48.9
Math Progress Assessment, Grade 2 ^b	73	16	0.68	0.29	0.12	34.8
Writing Progress Assessment, Grade 3 ^b	75	19	0.56	0.26	0.14	27.1
Math PSSA, Grade 3 ^c	1,301	186	0.69	0.30	0.12	43.5
Reading PSSA, Grade 3 ^c	1,286	158	0.71	0.23	0.12	18.8
Math PSSA, Grade 11 ^c	1,280	211	0.79	0.08	0.09	3.4
Reading PSSA, Grade 11 ^c	1,277	235	0.74	0.05	0.08	0.0
Writing PSSA, Grade 11 ^c	1,451	238	0.52	0.10	0.12	0.0
Science PSSA, Grade 11 ^c	1,189	89	0.71	0.03	0.08	0.0

Source: Mathematica calculations based on Pennsylvania student data and student data from pilot districts.

Note: Findings are based on one-cohort models in which the sample of teachers consists of those who served as teachers in 2010-2011. One z-score unit is equal to one standard deviation of student outcomes.

^a Findings are based on students and teachers in Allentown and Cornell.

^b Findings are based on students and teachers in Allentown.

^c Findings are based on students and teachers in Allentown, Cornell, and Mohawk.

DIBELS = Dynamic Indicators of Basic Early Literacy Skills; NWF = nonsense word frequency; ORF = oral reading fluency; PSF = phoneme segmentation fluency; PSSA = Pennsylvania System of School Assessment; VAM = value-added model.

Table C.5. Sample Characteristics of Outcome Measures and School VAMs Based on State Samples

Outcome	Distribution of Student-Level Outcome Variable		Characteristics of School VAMs and Estimates			
	Mean	Standard Deviation	Adjusted R-Squared	85th Minus 50th Percentile of VAM Estimates (in z-score units)	Mean Standard Error (in z-score units)	Percentage of VAM Estimates that Are Statistically Distinguishable from the Average
Math PSSA, Grade 4	1,473	221	0.66	0.20	0.04	64.4
Math PSSA, Grade 5	1,477	223	0.76	0.19	0.04	66.4
Math PSSA, Grade 6	1,499	238	0.78	0.20	0.04	72.0
Math PSSA, Grade 7	1,500	240	0.80	0.18	0.03	72.6
Math PSSA, Grade 8	1,452	221	0.80	0.15	0.03	67.4
Math PSSA, Grade 11 ^a	1,402	256	0.71	0.21	0.04	68.3
Reading PSSA, Grade 4	1,386	213	0.67	0.14	0.04	56.8
Reading PSSA, Grade 5	1,350	210	0.71	0.16	0.04	54.2
Reading PSSA, Grade 6	1,397	222	0.73	0.12	0.04	52.5
Reading PSSA, Grade 7	1,426	214	0.74	0.13	0.04	59.7
Reading PSSA, Grade 8	1,517	248	0.74	0.11	0.03	58.3
Reading PSSA, Grade 11 ^a	1,399	260	0.67	0.17	0.05	58.1
Writing PSSA, Grade 5	1,339	256	0.50	0.31	0.05	75.8
Writing PSSA, Grade 8	1,410	262	0.54	0.29	0.05	75.3
Writing PSSA, Grade 11 ^a	1,535	281	0.48	0.30	0.06	67.9
Science PSSA, Grade 4	1,461	182	0.64	0.24	0.05	69.0
Science PSSA, Grade 8	1,322	199	0.73	0.18	0.04	71.4
Science PSSA, Grade 11 ^a	1,255	92	0.68	0.23	0.05	69.9
Holding Power, Grade 9 ^a	91	25	0.40	0.58	0.05	92.2
Holding Power, Grade 10 ^a	91	24	0.41	0.60	0.06	92.0
Holding Power, Grade 11 ^b	88	33	0.73	0.93	0.05	90.0

Source: Mathematica calculations based on Pennsylvania student data.

Note: Unless otherwise noted, findings are based on three-cohort models with statewide samples of schools, principals, and students and a 95 percent confidence interval. The sample of principals consists of those who served as principals in every year from 2008-2009 to 2010-2011. One z-score unit is equal to one standard deviation of student outcomes.

^a Findings are based on a two-cohort model.

^b Findings are based on a one-cohort model.

Table C.6. Sample Characteristics of Outcome Measures and School VAMs Based on Phase 1 Samples

Outcome	Distribution of Student-Level Outcome Variable		Characteristics of School VAMs and Estimates			
	Mean	Standard Deviation	Adjusted R-Squared	85th Minus 50th Percentile of VAM Estimates (in z-score units)	Mean Standard Error (in z-score units)	Percentage of VAM Estimates that Are Statistically Distinguishable from the Average
DIBELS (NWF), Grade 1 ^a	39	22	0.36	0.28	0.11	53.8
DIBELS (PSF), Grade 1 ^a	40	15	0.59	0.38	0.09	61.5
DIBELS (ORF), Grade 2 ^a	84	37	0.79	0.11	0.07	13.3
Math Progress Assessment, Grade 2 ^b	73	16	0.64	0.23	0.08	66.7
Writing Progress Assessment, Grade 1 ^b	76	14	0.47	0.40	0.10	53.8
Writing Progress Assessment, Grade 2 ^b	76	15	0.53	0.26	0.09	53.8
Writing Progress Assessment, Grade 3 ^b	75	19	0.50	0.16	0.10	21.4
Math PSSA, Grade 3 ^c	1,298	188	0.64	0.23	0.07	37.5
Reading PSSA, Grade 3 ^c	1,284	158	0.69	0.18	0.07	37.5
Attendance Rate, Grades 4-12 ^d	93	6	0.48	0.08	0.10	13.8

Source: Mathematica calculations based on Pennsylvania student data and student data from pilot districts.

Note: Findings are based on a 95 percent confidence interval and one-cohort models with samples of schools, principals, and students from the pilot districts. The sample of principals consists of those who served as principals in 2010-2011. One z-score unit is equal to one standard deviation of student outcomes.

^a Findings are based on students and teachers in Allentown and Cornell.

^b Findings are based on students and teachers in Allentown.

^c Findings are based on students and teachers in Allentown, Cornell, and Mohawk.

^d Findings are based on students and teachers in Allentown, Mohawk, and Northwest.

DIBELS = Dynamic Indicators of Basic Early Literacy Skills; NWF = nonsense word frequency; ORF = oral reading fluency; PSF = phoneme segmentation fluency; PSSA = Pennsylvania System of School Assessment; VAM = value-added model.

MATHEMATICA
Policy Research

www.mathematica-mpr.com

Improving public well-being by conducting high-quality, objective research and surveys

Princeton, NJ ■ Ann Arbor, MI ■ Cambridge, MA ■ Chicago, IL ■ Oakland, CA ■ Washington, DC

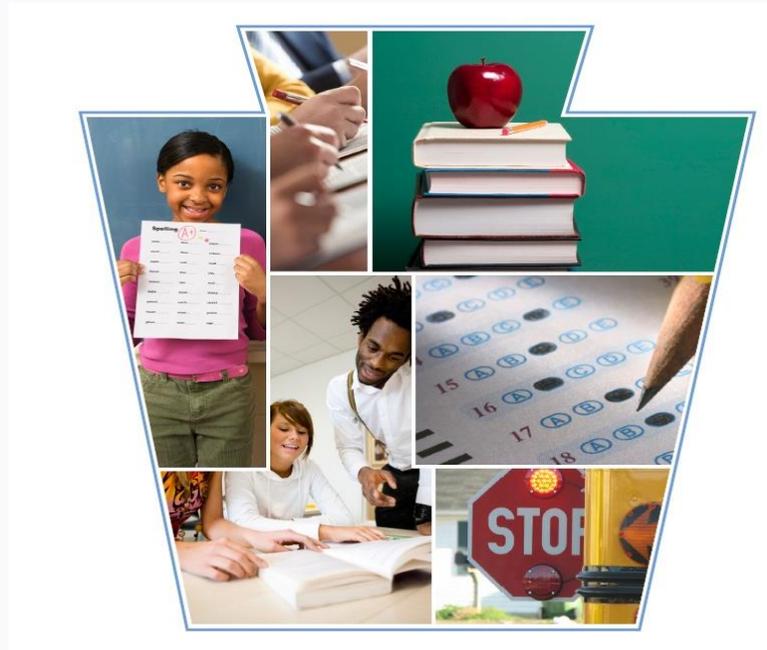
Mathematica® is a registered trademark of Mathematica Policy Research

TEACHER AND PRINCIPAL EFFECTIVENESS STAKEHOLDER GROUP

NAME	ORGANIZATION	STAKEHOLDER TYPE
[REDACTED]	Bucks County Intermediate Unit #22	Intermediate Unit
Sherri Smith	Superintendent, Lower Dauphin School District	Superintendent
Dr. Michael Leichter	Superintendent, Penn Manor School District	Superintendent
[REDACTED]	Marywood University	Higher Education - Private
[REDACTED]	Middle School Math, Philadelphia	Tested Teacher
[REDACTED]	6-12 grade English (currently 11-12), California Area	Tested Teacher
[REDACTED]	School Psychologist –Derry Township	Education Specialist
[REDACTED]	Music, Hollidaysburg School District	Non-tested teacher
[REDACTED]	Middle School Health and Phys Ed, Quakertown SD	Non-tested teacher
[REDACTED]	Literacy Specialist, IU 10	Teaching Specialist
Mike Reed	HS Principal, Williamsport	Principals
Jacqueline Cubberly	Elementary Principal, Bristol Township	Principals
[REDACTED]	Curriculum Director, Clarion Area SD	Administrators, other
[REDACTED]	Executive Director, Lehigh Career and Technical Institute	Career and Tech Ed
David Rossi	Esperanza Charter School, Philadelphia	Charter School CEO
Jim Smith	Bear Creek Community Charter School, Lehigh Valley	Charter School CEO
[REDACTED]	Central York Parent	Parent of School Age Children
[REDACTED]	Pennsylvania Partnerships for Children	Other
[REDACTED]	Executive Director, Office of Teacher Effectiveness Pittsburgh	Other
[REDACTED]	Harrisburg School District	Supervisor
[REDACTED]	Harrisburg School District	Parent
[REDACTED]	Warren Co. SD	Nurse
[REDACTED]	Keystone College	Professor and Chair Division of Education

Principle 3 – Appendix F

Measuring Educator Effectiveness



Teacher Effectiveness System in Act 82 of 2012

Observation/Evidence

Effective 2013-2014 SY

Danielson Framework Domains

1. Planning and Preparation
2. Classroom Environment
3. Instruction
4. Professional Responsibilities

School Building Data

Effective 2013-2014 SY

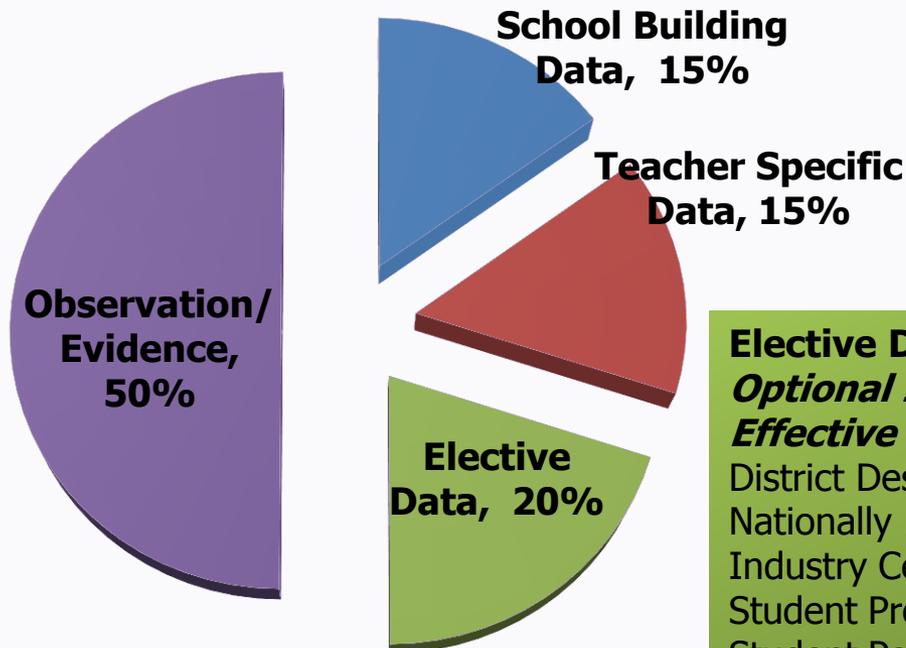
Indicators of Academic Achievement
Indicators of Closing the Achievement Gap, All Students
Indicators of Closing the Achievement Gap, Subgroups
Academic Growth PVAAS
Other Academic Indicators
Credit for Advanced Achievement

Teacher Specific Data

PVAAS / Growth 3 Year Rolling Average

1. 2013-2014 SY
2. 2014-2015 SY
3. 2015-2016 SY

Other data as provided in Act 82



Elective Data/SLOs

Optional 2013-2014 SY

Effective 2014-2015 SY

District Designed Measures and Examinations
Nationally Recognized Standardized Tests
Industry Certification Examinations
Student Projects Pursuant to Local Requirements
Student Portfolios Pursuant to Local Requirements

Teacher Effectiveness System in Act 82 of 2012

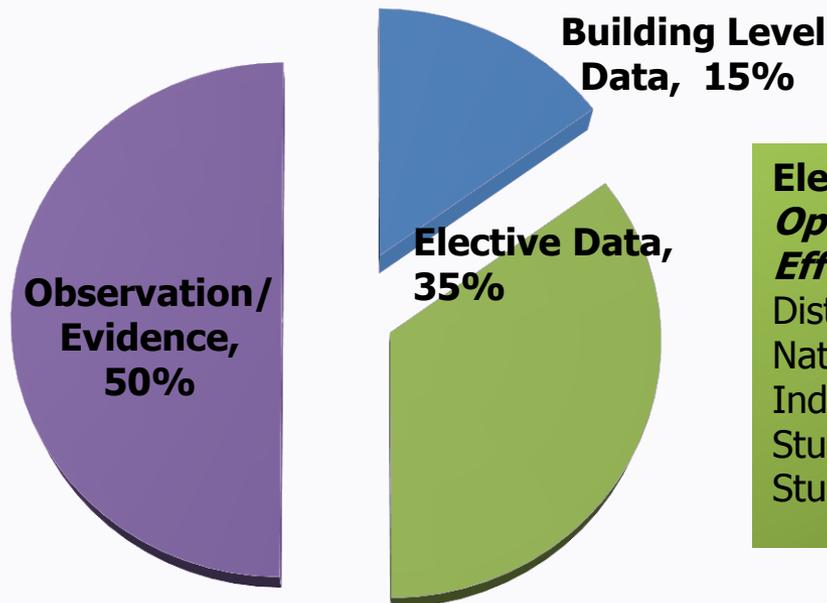
Observation/Evidence *Effective 2013-2014*

Danielson Framework Domains

1. Planning and Preparation
2. Classroom Environment
3. Instruction
4. Professional Responsibilities

Building Level Data *Effective 2013-2014 SY*

Indicators of Academic Achievement
Indicators of Closing the Achievement Gap, All Students
Indicators of Closing the Achievement Gap, Subgroups
Academic Growth PVAAS
Other Academic Indicators
Credit for Advanced Achievement



Elective Data/SLOs *Optional 2013-2014 SY* *Effective 2014-2015 SY*

District Designed Measures and Examinations
Nationally Recognized Standardized Tests
Industry Certification Examinations
Student Projects Pursuant to Local Requirements
Student Portfolios Pursuant to Local Requirements

Non Teaching Professional Employee Effectiveness System in Act 82 of 2012 *Effective 2014-2015 SY*

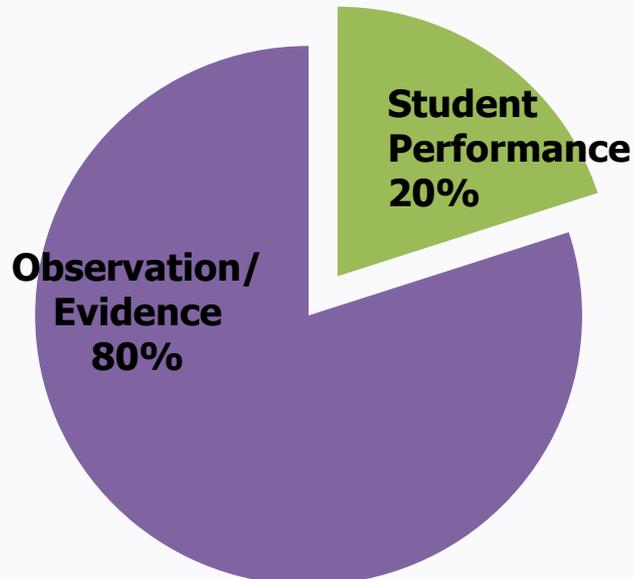
Observation/Evidence

Danielson Framework Domains

1. Planning and Preparation
2. Educational Environment
3. Delivery of Service
4. Professional Development

Student Performance of All Students in the School Building in which the Nonteaching Professional Employee is Employed

District Designed Measures and Examinations
Nationally Recognized Standardized Tests
Industry Certification Examinations
Student Projects Pursuant to Local Requirements
Student Portfolios Pursuant to Local Requirements



Measuring Educator Effectiveness

Tom Corbett, Governor • Ronald Tomalis, Secretary of Education

www.education.state.pa.us

Principal Effectiveness System in Act 82 of 2012 *Effective 2014-2015 SY*

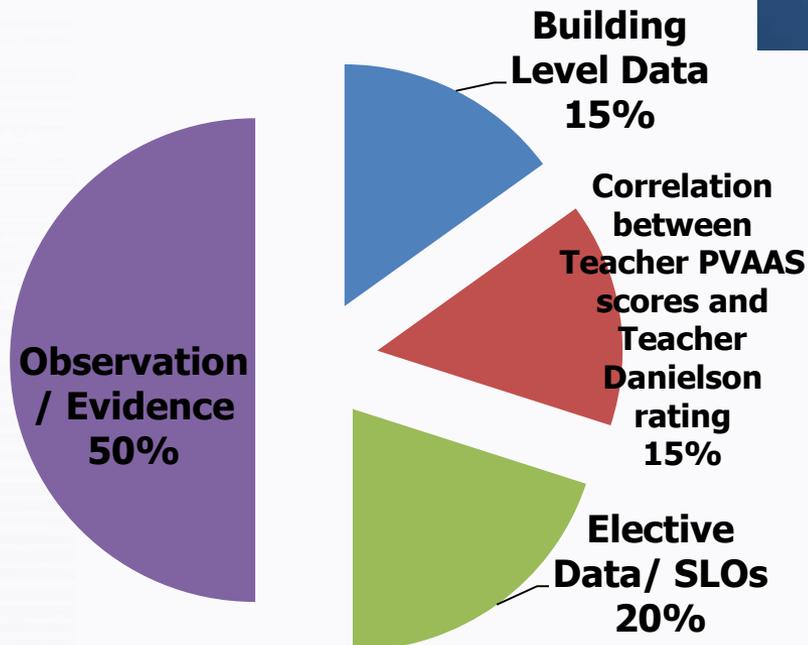
Observation/ Evidence

Domains

1. Strategic/Cultural Leadership
2. Systems Leadership
3. Leadership for Learning
4. Professional and Community Leadership

Building Level Data

Indicators of Academic Achievement
Indicators of Closing the Achievement Gap, All Students
Indicators of Closing the Achievement Gap, Subgroups
Academic Growth PVAAS
Other Academic Indicators
Credit for Advanced Achievement



Correlation Data Based on Teacher Level Measures PVAAS

Elective Data/SLOs

District Designed Measures and Examinations
Nationally Recognized Standardized Tests
Industry Certification Examinations
Student Projects Pursuant to Local Requirements
Student Portfolios Pursuant to Local Requirements

Measuring Educator Effectiveness

Tom Corbett, Governor ▪ Ronald Tomalis, Secretary of Education

www.education.state.pa.us



ESEA Waiver: Appendix: PVAAS Teacher Specific Reporting

PVAAS Methodology

- PVAAS uses the EVAAS methodology from SAS EVAAS for K-12
- SAS EVAAS for K-12 has been providing teacher level reporting since 1996
 - PVAAS is based on a sound statistical method that have been around for 2+ decades
 - Based on an abundance of research over the last 20 years
- EVAAS methodology has been published, research-based, and peer-reviewed
 - Published since 1997
 - Reviewed/Validated by:
 - Four separate US Department of Education Peer Review Committees
 - US Government Accountability Office (US GAO)
 - RAND Corporation
 - WestEd
- SAS EVAAS has 4 statewide implementations of teacher level reporting (PA, OH, TN, NC) + 24 regional/district implementations across the US
 - PA has been learning from other states, districts in the US
- PVAAS provides a measure of student growth that is aggregated to the district, school or teacher level.
- PVAAS adequately accounts for the entering achievement of students to fairly measure growth(high achieving, low achieving, ELL, IEP)
- PVAAS provides fair, valid and reliable information to educators on the students they serve for a variety of educational issues.

Purpose of PVAAS Teacher Reporting:

- Provide a teacher specific growth measure to be used as 15% of Pennsylvania’s Educator Effectiveness System
- Provide feedback to teachers regarding their influence on the academic growth of students
- Provide data for teachers and administrators to guide discussions about the academic progress of groups of students
 - Discussions between teachers
 - Discussions between teachers and administrators

Benefits of PVAAS Teacher Specific Reporting:

- PVAAS focuses on progress, rather than achievement.
- Educators are held accountable for the things that they can control, like their students’ progress during the school year, and they are not held accountable for the things they cannot change, like their students’ entering achievement and their socioeconomic/ demographic background.
- PVAAS focuses on all students - high and low achieving, making sure they all have an effective schooling experience, not just those around the “bubble.”

- Educators serving very high-achieving students can achieve positive results with PVAAS.
- Educators serving low-achieving students can achieve positive results with PVAAS.
- The reliability of these estimates makes them suitable for high stakes decisions, such as accountability, evaluation, etc.
- PVAAS Teacher Specific Reporting will provide reliable information to PA educators on the students they serve for a variety of educational issues.

PVAAS: Fair, Valid and Reliable

- Fair because they do not depend on the type of students a teacher receives each year
- Valid because they are unbiased measures of progress
 - Inherent challenges in student testing are accommodated
 - Students serve as their own control
- Reliable because they are repeatable and consistent
 - Multiple year estimates protect teachers from misclassification & are even more reliable
- Suitable for high stakes decisions, and as a component of accountability and evaluation systems

Stability of PVAAS Measures for Teacher Specific Reporting:

- Highly effective teachers remain highly effective regardless of years of experience.
 - A teacher with just three years of experience that is very highly effective (Level 5) using a three-year average has more than a 70% chance of remaining very effective (level 4 or 5).
 - This probability stays in between 70-80% regardless of the years of experience for highly effective teachers. It is even higher in Math.
- The reverse is also true but more related to years of experience.
 - A teacher with just three years of experience that is defined as level 1 using a three-year average has about a 50% chance of remaining a level 1 or 2 teacher three years into the future.
 - After having 6 years of experience, the probability rises to about 65% and after 10 years of experience that probability rises to above 70%.

What Will PVAAS Teacher Specific Reporting Provide to PA Educators:

- District, school and teacher value-added reporting
 - The reporting is available for a specific subject/grade/year as well as for composites across subjects/grades/years.
- District, school and teacher diagnostic reporting, as a less formal measure of growth for students in different achievement and demographic subgroups
- Student-level projections, which provide a probability of success on future academic assessments such as PSSA and Keystones

- Key point: EVAAS reporting goes beyond a single estimate of effectiveness and provides a variety of information and support at the state, district, school, teacher and student levels to inform educational practices and policies.

How Can PA Educators Use PVAAS Teacher Specific Reporting:

- Superintendents of school districts can...
 - Identify highly effective schools that are similar in demographic, socioeconomic makeup to less effective schools to share best practices (scatterplots and VA summary reports)
 - Identify highly effective teachers and principals based on student growth and progress that to guide PD efforts for struggling areas (teacher summary and VA summary reports)
 - Strategically allocate scarce resources based on evidence of program effectiveness as linked to student learning gains (student pattern reports)
 - Identify the number of students who will need immediate support or intervention for resource allocation (projection summary reports)
 - Empower district leaders with reliable analyses and evidence to support decision making and school improvement planning (all district reports)
- Principals of schools can...
 - Identify students who are candidates for advanced or remedial coursework (student projections and custom student reports)
 - Bring clarity to strategic planning (school VA and projection summary reports)
 - Provide data-driven basis to plans that ensure all students reach their potential (diagnostic reporting)
 - Understand and leverage the strengths of effective teachers (teacher reports)
 - Drive professional development of teachers in struggling areas by leveraging effective teaching within the school (teacher summary reports)
 - Become true instructional leaders when armed with these data-driven tools to benefit as many teachers and students as possible (all school, teacher and student reports)

Teachers Can:

- When the reporting is released, a teacher may view their overall value-added in each subject and grade as well as multi-year trends
- To gain a better understanding of the value-added reports, they can look at the diagnostic reports.
- This will allow a teacher to see if they are being more or less effective with different entering achievement students or possibly different demographic subgroups.
- Looking at projections helps identify those similar students that they currently have to help make informed decisions with students in the coming year.
- Teachers can do this alone or with their principal to build a plan for increasing effectiveness with all students they serve.
- Professional Learning Communities can also leverage and share this type of information to share best practices and improve instruction and outcomes

STUDENT LEARNING OBJECTIVE (SLO) TEMPLATE

A tool to provide a measure of educator effectiveness based on student achievement of content standards. SLOs are a part of Pennsylvania's multiple-measure, comprehensive system of Educator Effectiveness authorized by Act 82 (HB 1901).

1. Teacher Information

1a.	Teacher Name	
1b.	School Name	
1c.	District Name	

2. Targeted Content Areas

2a.	Course Title(s)/ Targeted Content Areas		2b.	Grade Level(s)	
2c.	PA Standards Subject Areas				
2d.	Classroom Setting				
2e.	Total Number of students (all classes/sessions)		2f.	Frequency of classes/sessions	
			2g.	Total Students used for SLO	

3. Student Learning Objective (SLO)

3a.	SLO Statement	
3b.	SLO Rationale	
3c.	Date SLO is due to principal/evaluator	
3d.	Date(s) for Assessment and Data Collection	
3e.	Date to complete Teacher Effectiveness Measure	

4. Data and Targets

4a.	Current Performance/ Baseline Data	
4b.	Targets	

5. Performance Measure(s)		
5a.	Name of the Performance Measure(s)	
5b.	Type of Performance Measure(s) (Check all being used)	<input type="checkbox"/> District-designed Measures and Examinations <input type="checkbox"/> Nationally Recognized Standardized Tests <input type="checkbox"/> Industry Certification Examinations <input type="checkbox"/> Student Projects <input type="checkbox"/> Student Portfolios
5c.	Performance Measure(s) Purpose Statement	
5d.	Metric Type (Check one)	<input type="checkbox"/> Growth (change in student performance across two or more points in time) <input type="checkbox"/> Mastery (attainment of a defined level of achievement) <input type="checkbox"/> Growth and Mastery

6. Administration		
6a.	Frequency of Administration	
6b.	Resources/Equipment Required	
6c.	Adaptations/Accommodations	
6d.	Personnel	

7. Student Performance Data		
7a.	Rubrics/Scoring	
7b.	Data Collection	
7c.	Scoring Student Mastery	
7d.	Scoring Student Growth	
7e.	Data Reporting	

8. Teacher Effectiveness Measure				
8a.	Classroom Objective How will the aggregated scores of the “Student Performance Data” results be used to define teacher effectiveness?			
8b.	<u>Failing</u> : few students achieve content mastery or growth	<u>Needs Improvement</u> : less than a significant number of students achieve content mastery or growth	<u>Proficient</u> : A significant number of students achieve content mastery or growth	<u>Distinguished</u> : An exceptional number of students achieve content mastery or growth
8c.				
8d.	Targeted Student Population Objective How will the mastery or growth of targeted student populations be described and used to define teacher effectiveness?			
8e.	<u>Failing</u> : Did not meet goal, little to no student mastery or growth	<u>Needs Improvement</u> : Did not fully meet goal but showed some student mastery or growth	<u>Proficient</u> : Met goal or otherwise demonstrated significant student mastery or growth	<u>Distinguished</u> : Surpassed goal otherwise demonstrated significant student mastery or growth
8f.	<u>Targeted Population:</u> <u>Mastery and/or growth goal:</u>			

9. Teacher Effectiveness Ratings	
What were the results of the assessments/tasks and how do they relate to the classroom and targeted objectives?	
9a.	Classroom Objective <input type="checkbox"/> Failing <input type="checkbox"/> Needs Improvement <input type="checkbox"/> Proficient <input type="checkbox"/> Distinguished
9b.	Targeted Objective <input type="checkbox"/> Failing <input type="checkbox"/> Needs Improvement <input type="checkbox"/> Proficient <input type="checkbox"/> Distinguished
	<u>Notes/Explanation</u>

Disclaimer:

CLASSROOM TEACHER / TEMPORARY CLASSROOM TEACHER RATING FORM

PDE 82-2 (9/12) – Teachers without Eligible PVAAS Score

Last Name	First	Middle
District/LEA	School	
Rating Date:	Evaluation: (Check one)	<input type="checkbox"/> Semi-annual <input type="checkbox"/> Annual

(D) Teacher Practice

Domain	Title	*Rating* (A)	Factor (B)	Earned Points (A x B)	Max Points
I.	Planning & Preparation		20%		0.60
II.	Classroom Environment		30%		0.90
III.	Instruction		30%		0.90
IV.	Professional Responsibilities		20%		0.60
(1) Total Teacher Practice Rating					3.00

Domain Rating Assignment 3 Point Scale (A)	
Rating	Value
Failing	0
Needs Improvement	1
Proficient	2
Distinguished	3

(E) Student Performance - Building Level Data, Teacher Specific Data, and Elective Data

Building Level Score (0 – 105)	
(2) Building Level Score Converted to 3 Point Rating	

(3) Teacher Specific Rating	
(4) Elective Rating	

(F) Final Teacher Effectiveness Rating – All Measures

Measure	Rating (C)	Factor (D)	Earned Points (C x D)	Max Points
(1) Total Teacher Practice Rating		50%		1.50
(2) Building Level Rating		15%		0.45
(3) Teacher Specific Rating	NA	NA	NA	NA
(4) Elective Rating		35%		1.05
Total Earned Points				3.00

Conversion to Performance Rating	
Total Earned Points	Rating
0.00-0.49	Failing
0.50-1.49	Needs Improvement
1.50-2.49	Proficient
2.50-3.00	Distinguished
Performance Rating	

Rating: Classroom Teacher, *OR* **Rating: Temporary Classroom Teacher**

I certify that the above-named employee for the period beginning _____ and ending _____ has received a performance rating of:
(month/day/year)
(month/day/year)

DISTINGUISHED **PROFICIENT** **NEEDS IMPROVEMENT** **FAILING**
 resulting in a FINAL rating of:
 SATISFACTORY **UNSATISFACTORY**

An overall performance rating of Distinguished, Proficient or Needs Improvement shall be considered satisfactory, except that the second Needs Improvement rating issued by the same employer within 10 years of the first final rating of Needs Improvement where the employee is in the same certification shall be considered unsatisfactory. A rating of Failing shall be considered unsatisfactory.

 Date Designated Rater / Position: Date Chief School Administrator

I acknowledge that I have read the report and that I have been given an opportunity to discuss it with the rater.
 My signature does not necessarily mean that I agree with the performance evaluation.

Date

Signature of Employee

DRAFT

Supervision Models

Pennsylvania Department of Education

February 13, 2013

Table of Contents

Supervision and Evaluation	3
Supervision Models.....	4
Formal Observation	4
Performance Improvement Plan.....	4
Intensive Supervision	5
Differentiated Supervision	5

Supervision and Evaluation

The Pennsylvania Department of Education has identified a supervision model consisting of two modes that will result in the professional development of educators: **Formal Observation** and **Differentiated Supervision**. Formal observation of the teachers' practice is accomplished through formal and informal observations measured by research supported best practices: Danielson's *Framework for Teaching*. The assessment supported by Danielson's *Framework for Teaching* and other observational data is **formative**. The collaborative reflections of the observational data will focus the efforts of the teacher on a professional development plan to improve instructional practices and student achievement. Differentiated Supervision recognizes the level of experience, the effectiveness, and professionalism of teachers as well as the intensity and time commitment to formal observation. Professional employees will develop an action plan for professional development based on the guidelines found on page 5.

The summative evaluation complements the formative supervision process. Act 82 states that all professional employees must be evaluated once a year and temporary professional employees must be evaluated twice a year. The data from Danielson's *Framework for Teaching*, other observational data, and the multiple measures data will be used to determine a performance rating (Summative Evaluation). Act 82 requires that all teachers will be rated as Distinguished, Proficient, Needs Improvement or Failing. An overall performance rating of either Proficient or Distinguished shall be considered satisfactory. An overall performance rating of Needs Improvement shall be considered as satisfactory, except that any subsequent overall rating of Needs Improvement issued by the same employer within ten (10) years of the first overall performance rating of Needs Improvement where the employee is in the same certification shall be considered unsatisfactory. An overall performance rating of failing shall be considered unsatisfactory. Unsatisfactory overall performance ratings require a **Performance Improvement Plan** with **Intensive Supervision** that is enacted for evaluation purposes based upon the **Performance Improvement Plan** (refer to page 4).

Supervision Models

Formal Observation (also referred to as clinical supervision)

Formal observation is based on the Danielson *Framework for Teaching* unless an alternative rating system has been approved by the Pennsylvania Department of Education (PDE). This requires a pre-observation conference, observation, a reflective session of the observed lesson and a post-observation conference to determine areas for growth (formative assessment) and to assist the data collection process for the summative assessment. Formal observations will be supplemented by walkthroughs, informal observations, conversations, etc., that may occur at any time. Professional employees must be evaluated at least once a school year, while temporary professional employees must receive two summative evaluations.

Each local LEA should create a **Cycle of Supervision** for all professional employees based upon their experience and professionalism. It is recommended that a cycle of supervision should be established based upon the number of teachers required to be supervised through formal observation. It is recommended that during the cycle of supervision the following classifications of professional employees will participate in the Formal Observation Mode:

1. Tenured professionals will be assigned to the Formal Observation Mode for 1 year during the Cycle of Supervision.
2. Tenured professionals new to the district will be placed in the Formal Observation Mode for 1 year prior to being eligible for Differentiated Supervision.
3. Level I certificated professionals and Long Term Substitutes will be assigned to the Formal Observation Mode. These professionals will be evaluated twice a year until tenure is granted.
4. Professionals who have been identified as **Needs Improvement** or **Failing** as an overall rating. It is recommended that teachers should receive a professional development plan in the deficient domain/component until a proficient rating has been achieved.

Performance Improvement Plans

Teachers who receive an overall performance rating of **Needs Improvement or Failing** are required by Act 82 to participate in a **Performance Improvement Plan**. A **Performance Improvement Plan** shall be designed with the professional employee's input addressing the area(s) of concern, recommendations for professional development, types of data (evidence) that will be collected to determine improvement, and an observation schedule with **Intensive Supervision**. PDE recommends that an intensive supervision timeline is established to enact the Performance Improvement Plan. At the conclusion of the allotted period to demonstrate a Proficient level of performance, the data will be analyzed and used to make a determination of the employee's employment status. It is recommended that the administrator recruit a colleague such as an assistant principal or the administrator's immediate supervisor in this process to provide additional reliability to the final determination of the professional employee's continuation of employment.

When the Performance Improvement Plan has been successfully completed by the at-risk employee, it is recommended that tenured professionals should be placed in the Formal Observation Mode for at least a full school year and non-tenured professionals remain in the Formal Observation Mode until tenure is granted

Intensive Supervision

Intensive supervision is initiated by an employee's overall Unsatisfactory performance rating and is accompanied by a Performance Improvement Plan. The purpose of intensive supervision is to ensure that the Performance Improvement Plan is enacted with fidelity and data is collected to determine the employment status of the employee.

Differentiated Supervision

Differentiated supervision recognizes the level of experience, the effectiveness, and professionalism of teachers as well as the intensity and time commitment of the formal observation process using the Danielson *Framework for Teaching*. Differentiated supervision provides a framework for professional growth designed to improve teacher effectiveness, instructional practices, and student achievement. Differentiated Supervision Modes should be included in the district's Supervision manual.

PDE recommends that professional employees who have received a Satisfactory summative rating in the previous two years should be eligible to participate in Differentiated Supervision. Prior to the 2013 - 2014 school year, a Satisfactory performance rating using a previously approved rating tool e.g., PDE 5501, PDE 427 or PDE 428 may be used to qualify for participation in differentiated supervision. It is also recommended that professional employees newly hired by a district should be eligible to participate in differentiated supervision after successfully completing their first year in the Formal Observation Mode.

LEA's should create a cycle of supervision based on the number of teachers requiring formal observations (temporary professional employees, professional employees new to a district, employees assigned to their required year of formal observation and employees assigned to a performance improvement plan). Professional employees should be assigned to differentiated supervision for the length of the cycle of supervision except for the required year of **formal observation** e.g., If a district has a three year cycle of supervision and a teacher is assigned to the Formal Observation Mode in the second year of the cycle, the teacher would be placed in the differentiated supervision in years one and three of the cycle. A cycle of supervision usually last for three (3) or four (4) years: however, this is a local decision.

LEAs should collaboratively create a timeline to ensure the successful completion of the professional's Differentiated Supervision Action Plan. The professional employee will be required to complete a mid-year review and an end-of-the-year self-reflection report with respect to his/her goal setting, planning, progress, and results. It is also recommended that the professional employee report the findings of their action plan to a Professional Learning Community (faculty meeting, in-service gathering, PTA/PTO). However, this is a local decision.

The professional should select a differentiated supervision mode in collaboration with the supervising administrator. All differentiated supervision modes should be aligned to the Danielson *Framework for Teaching* or a PDE approved alternative system and/or is related to a district or school initiative designed to improve instructional practices.

Additionally, while formal observations may not occur in differentiated supervision, it is recommended that informal observations should occur throughout the school year. PDE recommends that the principal reserves the right to remove a teacher from differentiated supervision at any time and place the teacher in the **Formal Observation Mode** or assign the teacher to a Performance Improvement Plan with **Intensive Supervision**. The professional employee should remain in the differentiated supervision mode for the length of the **cycle of Supervision** except for the required year of **formal observation**.

The rating tool will require principals/supervisors to provide a score in the 4 domains for all teachers every year regardless of their mode of supervision. Therefore, principals should collect evidence in each of the four domains for teachers in differentiated supervision through walkthroughs, informal observations, conversations, etc. Resources employed by the professional, data collection instruments, and the results of the reflective sessions will be used in formative and summative assessments. In the absence of data, the score assigned to a domain for the employee would revert to their most recent summative evaluation.

While the nomenclature applied to the various differentiated supervision modes may be germane to each LEA, they are generally grouped by common subject matter. Districts are not limited to the following categories as long as the mode meets the requirements and rigor of the PDE's Teacher Effectiveness Instrument.

The following descriptions of Differentiated Supervision Modes are to serve as examples of acceptable modes:

1. **Peer Coaching** - professional employees work in dyads or triads to discuss and observe their own or another professional employee's pedagogy, student learning, curriculum, Common Core alignment, and other pertinent issues in a collaborative manner. The professionals will work together to define their professional needs and develop plans to assist them in the successful completion of the identified tasks including focusing on: specific target area(s), the evidence to be collected, establishing observation dates and a reflective session meeting. Meeting notes, data collection tools, results of the observations and the results of reflective sessions should be shared with the principal and used in formative and summative assessments.
2. **Self-Directed Model/Action Research**-professionals employees will develop a structured, on-going reflection of a practice-related issue (Danielson *Framework for Teaching* or a PDE approved alternative system). Professionals may work individually or in small groups, dyads or triads, to complete the action research project. Meeting notes, resources, data collection tools, and the results of the reflective sessions should be shared with the principal and used in formative and summative assessments.

Principle 3 – Appendix J

3. **Portfolios** - professional employees will examine their own practice in relation to the Danielson *Framework for Teaching* or a PDE approved alternative system and reflect on their portfolio in a written report and/or documented discussions with colleagues. Portfolios may be developed according to criteria established by the administrator or by the teacher based upon their interests or needs. Resources, data collection tools, and the results of the reflective sessions should be shared with the principal and used in formative and summative assessments.

**Book/research reviews would not be acceptable as a separate Differentiated Supervision mode. They may be used as the basis to develop the research for the action plan



Approval of Locally-Developed Alternative Teacher Evaluation System

PDE-5501D (1/13)

LEA Name:	Contact Information:
------------------	-----------------------------

INSTRUCTIONS: 1. This form is to accompany an LEA’s request to obtain PDE approval of a locally-developed alternative teacher evaluation system. 2. Requirements of Section 1123 or Pa.’s State Board of Education Regulations (listed in column (a) on page 4) that cannot be revised have been blacked-out in column (b) below. 3. An LEA must insert a check mark (√) in column (b) for each requirement it is changing in its alternative evaluation system. 4. An LEA must insert page reference(s) in column (c) to identify location of its documentation and evidence. 5. PDE will insert into column (d) the date each requirement is approved. 6. Each LEA approved to implement an alternative teacher evaluation system is required to resubmit its alternative system for approval every five years. 7. Chief School Administrator will certify compliance with regulations by signing and dating the certification statement in Section VIII. of this document.

This form is divided into the following 10 sections:

GENERAL OVERVIEW OF PENNSYLVANIA’S ALTERNATIVE TEACHER EVALUATION SYSTEM			
Section Number	Section Title	Page Number(s)	Requirements May Be Revised
Section I.	Description of Process for Developing Local Alternative Teacher Evaluation System	4-6	No
Section II.	Summative Teacher Rating Form	7-9	LEA may develop a summative teacher rating form as long as the form (1) includes the four observation domains (Planning and Preparation; Classroom Environment; Instruction; and

GENERAL OVERVIEW OF PENNSYLVANIA’S ALTERNATIVE TEACHER EVALUATION SYSTEM			
Section Number	Section Title	Page Number(s)	Requirements May Be Revised
			Professional Responsibilities); (2) retains four rating categories of Distinguished; Proficient; Needs Improvement; and Failing; and (3) provides the formulae used in the final teacher performance rating.
Section III.	Teacher Practice Rubrics – Cross-walked to 2007 or Later Edition of Danielson Framework and Limited to Four Domains Developed by PDE	10-11	Teacher practice must equal 50% of final teacher performance rating. LEA may develop rubrics for four domains established by PDE that are aligned with Summative Teacher Rating Form; locally-developed rubrics will be submitted as part of the alternative teacher evaluation system.
Section IV.	Uses Multiple Student Performance Measures	12-14	No
Section V.	Building-Level Specific Data	15-18	LEA may submit locally-developed performance measures as long as this element of teacher evaluation is greater than 0% and meets or exceeds measures of effectiveness established by §1123 of the Public School Code; Performance measures also subject to review by Technical Advisory Committee before PDE approval.
Section VI.	Teacher-Specific Data	19-21	LEA may submit locally-developed performance measures as long as this element of teacher evaluation is at least 15% and meets or exceeds measures of effectiveness established

GENERAL OVERVIEW OF PENNSYLVANIA’S ALTERNATIVE TEACHER EVALUATION SYSTEM			
Section Number	Section Title	Page Number(s)	Requirements May Be Revised
			by §1123 of the Public School Code; Performance measures also subject to review by Technical Advisory Committee before PDE approval.
Section VII.	Elective Data Measures – submitted via the Student Learning Objective (SLO) process developed by PDE	22-24	If an LEA’s locally-developed teacher effectiveness system reduces the weight (percentage) of elective data measures so they are less than 20%, the LEA must submit its adjustments for PDE consideration; in order for elective-data measures to meet or exceed the measures of effectiveness established by §1123 of the Public School Code LEA adjustments, LEA must submit its elective measures via the Student Learning Objective (SLO) process developed by PDE for non-tested subjects; elective-data measures must be greater than 0% and selected from PDE’s pre-approved list published in the <i>Pennsylvania Bulletin</i>.
Section VIII.	Chief School Administrator Certification and Signature	25	
Section IX.	Date Alternative Teacher Evaluation System Returned to LEA for Correction	26	
Section X.	Date of Final PDE Approval of LEA’s Alternative Teacher Evaluation System	27	



Approval of Locally-Developed Alternative Teacher Evaluation System

PDE-5501D (1/13)

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (√) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval ^{1, 2} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section I. Description of Process for Developing Local Alternative Teacher Evaluation System			
✓ Description of Procedures for Developing Alternative Teacher Evaluation System			
<ul style="list-style-type: none"> • List of Names and Titles of Individuals and Groups Involved in Development 			
<ul style="list-style-type: none"> • Dates of Meetings 			

¹ When cell is blackened, the LEA may not alter this requirement in its locally-developed alternative teacher evaluation system.

² An LEA is to insert a check mark (√) in Column (b) for each requirement it is changing in its alternative teacher evaluation system.

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (√) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{1, 2} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section I. Description of Process for Developing Local Alternative Teacher Evaluation System			
✓ Objectives of Alternative Rating Tool			
✓ Excerpt from Minutes Documenting Local Board of Directors' Approval			
○ School Year When Alternative Form Will Be Implemented			
✓ Description of How Alternative Rating Tool Will Be Used to Improve Effectiveness of Teachers			
✓ Provide Method for Determining Final Performance Rating and Ranges for Determining Distinguished, Proficient, Needs Improvement, and Failing Performances			
✓ Describe How Overall Teacher's Final Performance			

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (√) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{1, 2} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section I. Description of Process for Developing Local Alternative Teacher Evaluation System			
Rating is Determined to be "Satisfactory" or "Unsatisfactory"			



Approval of Locally-Developed Alternative Teacher Evaluation System

PDE-5501D (1/13)

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (√) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval ^{3, 4} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section II. Summative Teacher Rating Form			
✓ Appearance of Summative Rating Form			
<ul style="list-style-type: none"> • Maintains four domains of Planning and Preparation; Classroom Environment; Instruction; and Professional Responsibilities 			
<ul style="list-style-type: none"> • Retains rating categories of Distinguished, 			

³ When cell is blackened, the LEA may not alter this requirement in its locally-developed alternative teacher evaluation system.

⁴ An LEA is to insert a check mark (√) in Column (b) for each requirement it is changing in its alternative teacher evaluation system.

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (√) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{3, 4} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section II. Summative Teacher Rating Form			
Proficient, Needs Improvement, and Failing			
<ul style="list-style-type: none"> • Results in Final Rating of “Satisfactory” or “Unsatisfactory” 			
<ul style="list-style-type: none"> • Presents formula(ae) different from PDE’s approved Summative Teacher Rating Form for: 			
<ul style="list-style-type: none"> ○ Teacher Practice (Observation) 			
<ul style="list-style-type: none"> ➤ Weight Factor for Teacher Practice is Equal to 50% 			
<ul style="list-style-type: none"> ○ Building-level Performance Measures 			
<ul style="list-style-type: none"> ➤ Weight Factor for Building-level is > 0% 			
<ul style="list-style-type: none"> ○ Elective Data 			

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (√) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{3, 4} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section II. Summative Teacher Rating Form			
Performance Measures			
➤ Weight Factor for Elective Data is > 0%			
✓ Provide Method for Determining Final Performance Rating and Ranges for Determining Distinguished, Proficient, Needs Improvement, and Failing			
✓ Describe How Overall Teachers’ Final Performance Rating is Determined to be Satisfactory or Unsatisfactory			



Approval of Locally-Developed Alternative Teacher Evaluation System

PDE-5501D (1/13)

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval ^{5, 6} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section III. Teacher Practice Rubrics – Cross-walked to 2007 or Later Edition of Danielson Framework and Limited to Four Domains Developed by PDE			
✓ 50% of a Classroom Teacher’s Overall Rating is Based on Multiple Student Performance Measures			
<ul style="list-style-type: none"> • Teacher-specific Performance Measures Comprise at Least 15% of 			

⁵ When cell is blackened, the LEA may not alter this requirement in its locally-developed alternative teacher evaluation system.

⁶ An LEA is to insert a check mark (✓) in Column (b) for each requirement it is changing in its alternative teacher evaluation system.

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (√) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{5, 6} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section III. Teacher Practice Rubrics – Cross-walked to 2007 or Later Edition of Danielson Framework and Limited to Four Domains Developed by PDE			
Multiple Student Performance Measures			
<ul style="list-style-type: none"> • Building-level Performance Measures are greater than 0% 			
<ul style="list-style-type: none"> • Elective Data Performance Measures are greater than 0% 			



Approval of Locally-Developed Alternative Teacher Evaluation System

PDE-5501D (1/13)

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval ^{7, 8} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section IV. Uses Multiple Student Performance Measures			
✓ Maintains Four Domains of Planning and Preparation; Classroom Environment; Instruction; and Professional Responsibilities			
✓ Includes Rating Categories that Are Aligned with Locally-			

⁷ When cell is blackened, the LEA may not alter this requirement in its locally-developed alternative teacher evaluation system.

⁸ An LEA is to insert a check mark (✓) in Column (b) for each requirement it is changing in its alternative teacher evaluation system.

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (√) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{7, 8} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section IV. Uses Multiple Student Performance Measures			
Developed Summative Teacher Rating Form			
<ul style="list-style-type: none"> • Rubrics Are Cross-Walked to Danielson Framework (2007 or later edition) 			
<ul style="list-style-type: none"> ✓ Describe the Data Elements to Be Used to Calculate the Final Teacher Evaluation Rating (i.e., Evidence Submitted to Document LEA’s Alternate Student Performance Measures Meet or Exceed Effectiveness Established by §1123 of the Pennsylvania School Code) 			
<ul style="list-style-type: none"> ✓ Describe the Weights and Scaling Used to Produce Final Teacher Evaluation Rating 			

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (√) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{7, 8} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section IV. Uses Multiple Student Performance Measures			
(i.e., Evidence Submitted to Document LEA’s Alternate Student Performance Measures Meet or Exceed Effectiveness Established by §1123 of the Pennsylvania School Code)			
✓ Elective Data Measures for Non-Tested Subjects were Developed Using PDE’s Student Learning Objective (SLO) process			



Approval of Locally-Developed Alternative Teacher Evaluation System

PDE-5501D (1/13)

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval ^{9, 10} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section V. Building-Level Specific Data -- must be greater than 0% of Overall Teacher Effectiveness (other documentation required by PDE's Technical Advisory Committee are included in Appendix ____)			
✓ Locally-developed Building-Level Data is greater than 0%			
✓ Describe in Detail How the Building-level Data are Used in Determining the Final Evaluation Score for Building-level Data (i.e.,			

⁹ When cell is blackened, the LEA may not alter this requirement in its locally-developed alternative teacher evaluation system.

¹⁰ An LEA is to insert a check mark (✓) in Column (b) for each requirement it is changing in its alternative teacher evaluation system.

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{9, 10} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section V. Building-Level Specific Data -- must be greater than 0% of Overall Teacher Effectiveness (other documentation required by PDE’s Technical Advisory Committee are included in Appendix ____)			
Evidence Submitted to Document LEA’s Alternate Student Performance Measures Meet or Exceed Effectiveness Established by §1123 of the Pennsylvania School Code)			
✓ Describe the Methodology Used to Produce the Building-level Data, Including Rationale for Data in the Proposed Model; How Other Data Will be Used; and How Each Data Element Will be Weighted/ Calculated in			

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{9, 10} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section V. Building-Level Specific Data -- must be greater than 0% of Overall Teacher Effectiveness (other documentation required by PDE's Technical Advisory Committee are included in Appendix ____)			
Alternative Model			
✓ Describe the Data Elements to Be Used to Calculate the Final Teacher Evaluation Rating for Building-level Data (i.e., Evidence Submitted to Document LEA's Alternate Student Performance Measures Meet or Exceed Effectiveness Established by §1123 of the Pennsylvania School Code)			
✓ Technical Advisory Committee Will Review Building-level Performance Measures			

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{9, 10} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section V. Building-Level Specific Data -- must be greater than 0% of Overall Teacher Effectiveness (other documentation required by PDE’s Technical Advisory Committee are included in Appendix ____)			
Submitted by LEA and Make Recommendation to PDE about Whether the Performance Measures, Formulae, and Other Features of the Alternative Teacher Evaluation System Meet or Exceed the Requirements of §1123			



Approval of Locally-Developed Alternative Teacher Evaluation System

PDE-5501D (1/13)

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval ^{11, 12} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section VI. Teacher-Specific Data - must be at Least 15% of Overall Teacher Effectiveness (other documentation required by PDE’s Technical Advisory Committee are included in Appendix ____)			
✓ Locally-developed Teacher Specific Data is at least 15%			
✓ Describe in Detail How the Teacher-specific Data are Used in			

¹¹ When cell is blackened, the LEA may not alter this requirement in its locally-developed alternative teacher evaluation system.

¹² An LEA is to insert a check mark (✓) in Column (b) for each requirement it is changing in its alternative teacher evaluation system.

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{11, 12} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section VI. Teacher-Specific Data - must be at Least 15% of Overall Teacher Effectiveness (other documentation required by PDE’s Technical Advisory Committee are included in Appendix _____)			
Determining the Final Evaluation Score for Teacher			
✓ Describe the Methodology Used to Produce the Teacher-specific Data, Including Rationale for Data in the Proposed Model; How Other Data Will be Used; and How Each Data Element Will be Weighted/ Calculated in Alternative Model			
✓ Describe the Data Elements to Be Used to Calculate the Final Teacher Evaluation Rating for Teacher-specific Data (i.e., Evidence			

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{11, 12} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section VI. Teacher-Specific Data - must be at Least 15% of Overall Teacher Effectiveness (other documentation required by PDE’s Technical Advisory Committee are included in Appendix _____)			
Submitted to Document LEA’s Alternate Student Performance Measures Meet or Exceed Effectiveness Established by §1123 of the Pennsylvania School Code)			
✓ Technical Advisory Committee Will Review Teacher-specific Performance Measures Submitted by LEA			



Approval of Locally-Developed Alternative Teacher Evaluation System

PDE-5501D (1/13)

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval ^{13, 14} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section VII. Elective Data Measures - must be greater than 0% of Overall Teacher Effectiveness (other documentation required by PDE’s Technical Advisory Committee are included in Appendix ____)			
✓ Elective Data Measures must be greater than 0% and selected from PDE’s pre-approved list published in the <i>Pennsylvania Bulletin</i>			
✓ If an LEA’s locally-developed teacher evaluation system			

¹³ When cell is blackened, the LEA may not alter this requirement in its locally-developed alternative teacher evaluation system.

¹⁴ An LEA is to insert a check mark (✓) in Column (b) for each requirement it is changing in its alternative teacher evaluation system.

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{13, 14} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section VII. Elective Data Measures - must be greater than 0% of Overall Teacher Effectiveness (other documentation required by PDE’s Technical Advisory Committee are included in Appendix ____)			
reduces the weight (percentage) of elective data measures so they are less than 20%, the LEA must submit its adjustments for PDE consideration			
✓ For elective-data measures to meet or exceed the measures of effectiveness established by §1123 of the Public School Code, LEA must submit its elective measures via the Student Learning Objective (SLO) process developed by PDE for non-tested			

Requirements of Section 1123 or State Board of Education Regulations (a)	Insert a Check Mark (✓) to Indicate Requirement Is Revised and Submitted for Pennsylvania Department of Education (PDE) Approval^{13, 14} (b)	Page Reference(s) in Materials Submitted to PDE (c)	Date of PDE Approval (d)
Section VII. Elective Data Measures - must be greater than 0% of Overall Teacher Effectiveness (other documentation required by PDE’s Technical Advisory Committee are included in Appendix ____)			
subjects			
✓ Describe in Detail How the Elective Data Measures Are Used in Determining the Final Teacher Evaluation			
✓ Technical Advisory Committee Will Review Elective Data Measures Submitted by LEA			



Approval of Locally-Developed Alternative Teacher Evaluation System

Accuracy Certification Statement

PDE-5501D (1/13)

Section VIII. Chief School Administrator Certification and Signature	
I certify that the information provided by my LEA accurately describes the alternative teacher evaluation system that was developed collaboratively by teachers and approved by the local board of directors for implementation. I further certify that the guidance provided by PDE was utilized in incorporating multiple student performance measures that meet or exceed the effectiveness established by §1123 of the Pennsylvania School Code.	
Chief School Administrator's Signature:	Date



Approval of Locally-Developed Alternative Teacher Evaluation System

PDE-5501D (1/13)

Section IX. Date Alternative Teacher Evaluation System Returned for Revision	
Section(s) of LEA's Alternative Teacher Evaluation System Being Returned	Reason(s) Why Section(s) Are/Were Returned
✓ Section I. Description of Process for Developing Local Alternative Teacher Evaluation System	
✓ Section II. Summative Teacher Rating Form	
✓ Section III. Teacher Practice Rubrics	
✓ Section IV. Uses Mandatory Multiple Student Performance Measures	
✓ Section V. Building-Level Specific Data	
✓ Section VI. Teacher-Specific Data	
✓ Section VII. Elective Data Measures	



Approval of Locally-Developed Alternative Teacher Evaluation System

PDE-5501D (1/13)

Section X. Date of Final PDE Approval of LEA's Alternative Teacher Evaluation System	
Date of Final PDE Approval	
Approved by:	
Title	
LEA must resubmit and reapply for approval of a locally-developed alternative teacher evaluation system within <i>five (5) years</i> of date of final approval or earlier if revisions are made.	

2011

Effects of the National Institute for School Leadership's Executive Development Program on School Performance in Pennsylvania: 2006-2010 Pilot Cohort Results

John A. Nunnery, Ed.D.

Cherng-Jyh Yen, Ph.D.

Old Dominion University

Steven M. Ross, Ph.D.

Johns Hopkins University

March, 2011



OLD DOMINION UNIVERSITY

DARDEN COLLEGE OF EDUCATION

THE CENTER FOR EDUCATIONAL PARTNERSHIPS

Contents

EXECUTIVE SUMMARY	2
Research Questions	2
Summary Method.....	3
Summary Results	3
Summary Findings	5
Introduction.....	7
The NISL Executive Development Program	7
Methods	8
Research Questions.....	9
Sample	9
Program participant selection process.....	9
Study sample inclusion	10
Elementary school matching procedure	10
Secondary school matching procedure.....	11
Measures	13
Analyses.....	13
Factorial analyses of variance	13
Hierarchical linear models	13
Results.....	16
Reading: Factorial Analyses of Variance.....	16
Reading: Hierarchical Linear Model Results.....	17
Mathematics: Factorial Analysis of Variance	19
Mathematics: Hierarchical Linear Model Results for All Schools	20
Hierarchical Linear Modeling Results by School Level in Reading and Mathematics	22
Discussion.....	24
References.....	29

EXECUTIVE SUMMARY

The National Institute of School Leadership's (NISL's) Executive Development Program (EDP) was established to provide professional development to school leaders to drive their schools to high performance. The program emphasizes the role of principals as strategic thinkers, instructional leaders, and creators of a just, fair, and caring culture in which all students meet high standards. Its primary goal is to ensure that the participating school leaders have the knowledge, skills, and tools to effectively set direction for teachers, support their staffs, and design an efficient organization. The curriculum is organized into four courses: *World-Class Schooling* (Principal as a Strategic Thinker and School Designer, Standards-Based Instruction), *Teaching and Learning*, *Developing Capacity and Commitment*, and *Driving for Results*. Training sessions are designed to be highly interactive through the use of simulations and assignment of “pre-work” and “homework” to participants.

This study examined the impact of EDP on student achievement in Pennsylvania schools from 2006-2010. It updates and extends a prior evaluation (Nunnery, Ross, & Yen, 2010a) study of this same cohort from 2006-2009, which found that elementary, middle, and high schools served by EDP principals had significantly larger gains in the percentages of students achieving proficiency in reading and mathematics.

Research Questions

The research questions addressed by the present study were:

1. How do the trends in school level performance in reading and English/Language Arts (ELA) differ between schools served by EDP-trained principals and matched comparison schools overall, and at the elementary, middle, and high school levels?

2. How do the trends in school level performance in mathematics differ between schools served by EDP-trained principals and matched comparison schools overall, and at the elementary, middle, and high school levels?

Summary Method

Participation by principals in the present EDP sample occurred through a multi-step process. Initially, the State Education Agency (SEA) publicized NISL to district superintendents. The latter, in turn, identified principals and assistant principals in their districts and encouraged them to apply. Actual applicants were selected by regional coordinators using an evaluation rubric. During the first two years of the program, participation was limited to principals or assistant principals in their first three years on the job. The highest weighting on the rubric evaluation was given to candidates from the lowest performing schools. All principals in the present sample completed EDP in 2008 or 2009.

On the basis of statistical analyses of school characteristics (e.g., percentage of students who are economically disadvantaged, special needs, LEP) and prior achievement in mathematics and reading/English Language Arts (ELA), 36 NISL elementary schools were individually matched to a highly similar comparison school in the same school district. Due to lacking appropriate within-district matches, an additional 32 elementary schools, all 19 NISL middle schools, and all 14 NISL high schools were matched to out-of-district comparison schools. Although this study was based on a rigorous *ex post facto* design with a carefully matched comparison group, selection effects cannot be ruled out given that the findings are not based on a fully randomized experiment.

Summary Results

Statistical analyses were conducted to compare the relationship between principals' participation in EDP and school-level performance in ELA and mathematics from 2006 (baseline)

through 2010¹. In both subjects, schools led by EDP principals improved at a greater rate than matched comparison schools. Specifically, all NISL schools improved in ELA performance at an average rate of about +0.5% greater than the comparison schools *each year*. In other words, by the end of the four years post-baseline, the typical NISL school had 2.16% more students achieving proficiency in than otherwise would be expected. This translates to about 1,225 more students between 2008 and 2010 who achieved reading/ELA proficiency in the NISL study schools that would have been expected otherwise. Likewise, in mathematics, NISL schools also improved at an average rate of about +0.5% proficient each year above the rate of improvement in comparison schools. By the end of the four years post-baseline, the typical NISL school had about 1.92% more students achieving proficiency than otherwise would be expected, which translates to a projected increase of 1,089 students across all NISL schools included in the study.

EDP effects were smaller (though still statistically significant) in elementary schools than in middle and high schools. Specifically, average annual increases in ELA proficiency over comparison schools were 0.18%, 1.04%, and .81% for elementary, middle, and high schools, respectively. For mathematics, the respective average annual increases were 0.17%, 0.51%, and 2.37%. Cumulative four-year gains (2007-2010) in the percentage of students achieving proficiency in reading/English language arts and mathematics are presented in Figure 1.

¹ Analyses consisted of factorial analysis of variance and two-level hierarchical linear modeling.

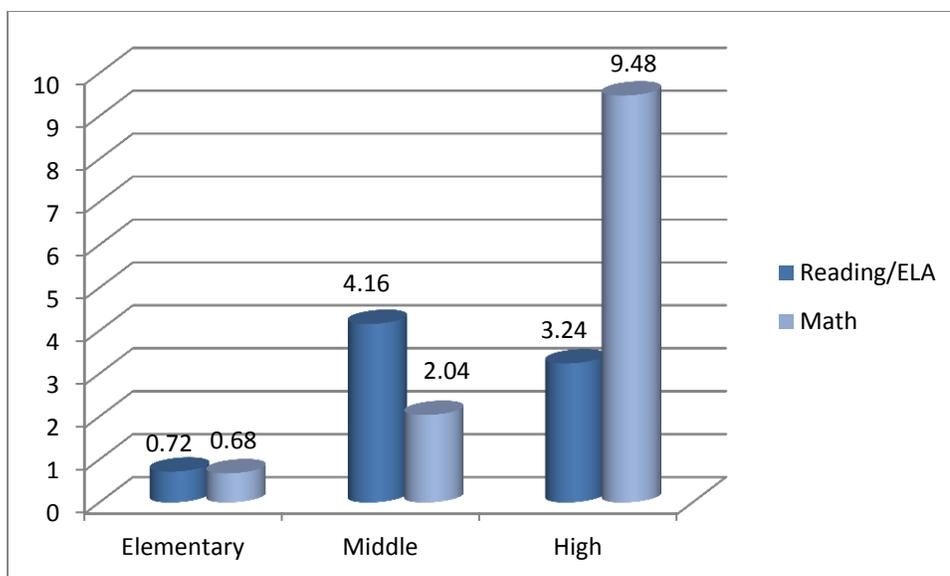


Figure 1. *Cumulative Four-Year Gains in the Percentage of Students Achieving Proficiency in Reading/English Language Arts and Math: Pennsylvania Executive Development Program Pilot Cohort 2007-2010.* Note. All gains statistically significant at $p < .001$.

Summary Findings

The results of this study show significantly larger gains by NISL schools relative to comparison schools in both reading/ELA and mathematics. In both subjects, NISL schools had lower percentages of students achieving proficiency in 2006 (the difference favoring comparison schools was statistically significant in mathematics). However, following principals' participation in EDP, the NSL schools were significantly superior in both subject areas in 2008, 2009, and 2010. Across all grade levels, the percentages of students achieving proficiency in NISL relative to comparison schools were +0.54 and +0.48 higher each year in reading/ELA and mathematics, respectively.

Compared to elementary schools, much stronger EDP effects were found in middle schools and high schools. This finding is particularly noteworthy, as there is little systematic evidence that any of the many high school reforms attempted to date have had a positive effect on student achievement (Fleischman & Heppen, 2009). In particular, very large gains in mathematics

performance were observed in EDP high schools. In 2006 (the baseline year), EDP high schools had only a +0.6% advantage in terms of percentage of students achieving proficiency in mathematics (46.3% versus 45.7%), whereas by 2010 EDP high schools had a 12% advantage (62.4% versus 50.5%). The greatest acceleration in mathematics performance in EDP schools occurred between 2009 and 2010, the first year following completion of the EDP program for all pilot cohort principals.

The results of this study represent highly promising evidence that the NISL Executive Development Program for school leaders results in statistically significant, substantial, and sustained improvements in student performance in reading and mathematics, particularly in the challenging context of secondary schools. This is particularly noteworthy given that the program is highly cost-effective, with current median participant costs of about \$4,000. In the context of the current study, this amounts to only about \$117 per additional student achieving proficiency: if the observed trends continued for one additional year, that cost would drop to about \$69 per additional student achieving proficiency in either reading or mathematics.

Introduction

This study examined the impact of the National Institute of School Leadership's (NISL) Executive Development Program (EDP) on student achievement in Pennsylvania schools from 2006-2010. The current study used an *ex post facto* design in which schools served by the first Pennsylvania cohort of EDP principal participants were matched individually to control schools with similar school performance and demographic profiles in the pre-program year of 2006. This report is an update of the Nunnery, Ross, & Yen (2010a) study of this same cohort from 2006-2009, which found that elementary, middle, and high schools served by principals who had completed the EDP program had significantly larger gains in the percentages of students achieving proficiency in reading and mathematics over that time span. This study extends and improves upon the Nunnery, Ross, & Yen (2010a) study by adding an additional year of results, by utilizing a more sophisticated statistical model to control for non-linear secular trend in school performance trajectories, and by implementing school-level weights based on the number of students tested each year in order to get more precise error estimates for statistical hypothesis testing and effect size estimation.

The NISL Executive Development Program

The National Institute of School Leadership's (NISL's) Executive Development Program was established to train school leaders to drive their schools to high performance. The program emphasizes the role of principals as strategic thinkers, instructional leaders, and creators of a just, fair, and caring culture in which all students meet high standards. Its primary goal is to ensure that the participating school leaders have the knowledge, skills, and tools to effectively set direction for teachers, support their staffs, and design an efficient organization. The curriculum, which was

designed by experts on leadership training across a number of fields, was developed with an \$11 million investment and five years of research and piloting.

Professional development goals of the EDP program are to provide high-quality instruction (both online and face-to-face), an advanced research-based curriculum, and an interactive approach to learning that includes simulations, case studies, school evaluations, and online activities.

Key expectancies for NISL-trained principals include:

- Formulating a clear vision to inspire others in the school communities,
- Implementing fully-aligned, standards-based instructional systems,
- Building effective instructional programs in the core academic subjects, particularly math, language arts and science,
- Using data to produce continuous improvements in instruction and student achievement,
- Providing effective training programs to build a professional learning community for school faculty and staff, and
- Creating integrated school improvement plans that reflect strategic and systemic thinking.

The curriculum is organized into four courses: *World-Class Schooling* (Principal as a Strategic Thinker and School Designer, Standards-Based Instruction), *Teaching and Learning*, *Developing Capacity and Commitment*, and *Driving for Results*. Professional development sessions are designed to be highly interactive through the use of simulations and assignment of “pre-work” and “homework” to participants.

Methods

This study is an extension of the Nunnery, Ross, & Yen 2010 study of EDP effects in Pennsylvania. It extends the study by incorporating an additional year of school performance data (2010). It is based on a carefully matched comparison-group ex post facto design in which schools served by principals participating in the program were individually matched to control

schools with similar school performance and demographic profiles in the baseline (pre-program) year of 2006. It also improves upon the prior study by utilizing more advanced hierarchical linear modeling techniques to get more precise estimates of program impacts. Although this study employed a rigorous *ex post facto* design with a carefully matched comparison sample, there are potential confounds, or alternative explanations, for the observed results that arise because the findings are not based on a fully randomized experiment. Thus, it is possible that selection effects may account for some of the observed differences between NISL and comparison schools. Also, as is the case with any statistical model, it also is possible that unmeasured variables might account for some of the observed differences reported in this study.

Research Questions

1. *How do the trends in school level performance in reading and English/Language Arts (ELA) differ between schools served by NISL-trained principals and matched comparison schools overall, and at the elementary, middle, and high school levels?*
2. *How do the trends in school level performance in mathematics differ between schools served by NISL-trained principals and matched comparison schools overall, and at the elementary, middle, and high school levels?*

Sample

Program participant selection process. Participation by principals occurred through a multi-step process. Initially, the State Education Agency (SEA) publicized NISL to district superintendents. The latter, in turn, identified principals and assistant principals in their districts and encouraged them to apply. The potential applicants were characterized by the SEA and superintendents as mixed in leadership potential (experiences, accomplishments, and skills), with some demonstrating strong promise and others regarded as needing professional development support to improve instructional leadership skills. Actual applicants were selected by regional coordinators using an evaluation rubric (see Appendix B). During the first two years of the program, participation was limited to principals or assistant

principals in their first three years on the job. The highest weighting on the rubric evaluation was given to candidates from the lowest performing schools. Starting on January 1, 2008, a new state (Act 45 of 2007) policy requiring a principals' induction program went into effect. Based on the law's requirement that all school and system leaders meet approved continuing education requirements, all applicants henceforth were accepted to the program on a first-come-first-served basis. This change in recruitment practices, however, did not affect the present sample of NISL participants.

Study sample inclusion. Data from all Pennsylvania elementary schools with complete test score data from 2005-2006 through 2008-2009 were initially considered for inclusion in the analyses. There were a total of 70 NISL elementary schools, 19 NISL middle schools, and 12 NISL high schools. As explained below, 36 of the NISL elementary schools were included in a within-district matched samples analysis, and 32 were included in a separate set of analyses based on an out-of-district matched comparison sample. In the middle school sample, 19 of the NISL principals completed the NISL program in 2009, and 17 completed it in 2008. In the middle school sample, 7 of the NISL principals completed the NISL program in 2009, and 12 completed it in 2008. In the high school sample, 6 of the NISL principals completed the NISL program in 2009, and 8 completed it in 2008. Thus, roughly half of NISL principals at any given grade level started the program in 2007 and completed in 2008, while half started in 2008 and 2009. The elementary within-district matched sample included an average of 4,565 students in comparison schools each year, and 5,898 students in NISL schools. Corresponding comparison and NISL average annual student sample sizes were 5,233 and 4,847 for elementary out-of-district matched samples, 8,916 and 7,498 for middle school out-of-district matched samples, and 3,017 and 2,552 for high school out-of-district matched samples.

Elementary school matching procedure. A principal components analysis was conducted using 2006 school performance index values in mathematics and reading, and the proportions of students who were economically disadvantaged, received special education services, or classified as having limited English proficiency (LEP). Results of the principal components analyses were used to construct a

regression-based factor score to use to identify matched pairs of schools for the analyses. Each NISL school was individually matched to a comparison school in the same school district that had the closest factor score. Matches were considered suitable only if the factor scores were within +/-0.25 standard deviation units. The matching process yielded a final sample for analysis of 36 NISL and 36 comparison schools at the elementary level that had matches within the same school district. Of the remaining 34 NISL elementary schools, two did not have test score data at all four time points. The remaining 32 were matched out-of-district, and these analyses were conducted separately. As shown in Table 1, this matching process yielded very closely matched samples. In 2006, the NISL elementary schools with within-district matches had a slightly lower percentage of economically disadvantaged students (15.8% versus 17.8%), slightly higher percentages of students with limited English proficiency (18.4% versus 14.8%), and a lower percentage of students who were proficient in math (78.0% versus 80.5%). For the out-of-district matches, all matching variables were within 0.1%, except percentage of limited English proficient students served (1.1% in comparison sites versus 0.6% in NISL sites).

Secondary school matching procedure. It was not possible to individually match middle and high school NISL schools to a comparison school within the same school district—in many cases, the NISL secondary school was the only school at that level within the district. At the middle and high school levels, an out-of-district match was made to each NISL school by matching the NISL school to a comparison school with the closest factor score. There were 19 NISL middle schools and 14 NISL high schools. As shown in Table 1, the matching process led to relatively well-matched samples for the middle school analyses, although the NISL sample was slightly more economically disadvantaged (27.6% versus 24.6%), and had somewhat lower initial scores in math (68.2% versus 72.0% proficient) and reading (69.9% versus 73.2%). Likewise, the high school samples were relatively well-matched, with comparison schools having somewhat higher percentages of economically disadvantaged students (27.9% versus 23.9%), higher proficiency rates in mathematics (47.8% versus 45.0%), and higher proficiency rates in Reading/ELA (63.8% versus 59.6%). NISL schools served higher percentages of special

education students (14.3% versus 10.9%). Independent samples *t*-tests conducted on school-level data showed that none of the NISL/comparison differences were statistically significant at $p < .05$. As shown in Table 1, proficiency levels at the high school level in 2006 were much lower than those observed at elementary or middle school, particularly in mathematics, where both NISL and comparison schools had less than 50% proficiency.

Table 1

Selected 2006 Characteristics of NISL and Comparison Schools

School Type	Economically Disadvantaged %	IEP %	LEP %	Math % Proficient	Reading % Proficient
Elementary ¹					
Comparison	17.8	6.41	14.8	80.49	69.20
NISL	15.8	6.61	18.4	78.04	68.10
Elementary ²					
Comparison	17.0	7.6	1.1	79.0	68.7
NISL	16.9	7.5	0.6	78.9	68.5
Middle ²					
Comparison	24.6	12.1	0.5	72.0	73.2
NISL	27.6	13.5	0.5	68.2	69.9
High ²					
Comparison	27.9	10.9	0.2	47.8	63.8
NISL	23.9	14.3	0.2	45.0	59.6

¹ $N = 36$ NISL and comparison schools for elementary within-district matches. ² $N = 32$ of each type for elementary out-of-district matches. $N = 19$ of each for middle schools. $N = 14$ of each for high schools. Note. No NISL/comparison differences on matching variables were statistically significant.

Measures

The proportions of students in tested grade levels (3-8 and 11) who were economically disadvantaged, received special education services as evidenced by the existence of an individualized education plan (IEP), or who were classified as having limited English proficiency (LEP) were computed for each school to utilize in the school matching procedure, and to provide for descriptive comparisons between EDP and comparison schools. Summary indices of school performance were constructed for both reading/ELA and mathematics by computing the proportion of students at all tested grade levels (grades 3 to 8 and grade 11) that scored proficient or higher on the Pennsylvania State Assessment. Note that for high schools, test scores were available only for eleventh grade.

Analyses

Factorial analyses of variance. The first stage of the analysis was to estimate 2 (comparison versus EDP) X 5 (Years) factorial analyses of variance for both reading and mathematics performance indices to provide a straightforward, unadjusted perspective on the actual results obtained in comparison schools and schools led by EDP program completers. Preliminary diagnostics indicated no *program X school level* interactions, so these analyses were performed on the total data set including schools across all grade levels. Scheffe's test was employed as the post hoc multiple comparisons criterion to account for unequal school-level numbers of students. Where the *program X year* interaction effect was statistically significant, pairwise comparisons between comparison and EDP means were conducted within years using Holme's sequential Bonferroni technique to control for experimentwise alpha inflation.

Hierarchical linear models. Two-level multilevel models were estimated to determine whether there were statistically significant relationships between EDP participation status at level 2, and the linear and quadratic components of longitudinal school-level achievement trajectories. Achievement trajectories were indicated as the proportion of students scoring at or

above proficiency in reading or English/language arts and mathematics. Indicators were available for 2006 through 2010.

Level-1 model. The level one (outcome and year within schools) was specified as:

$$Y_{ti} = \pi_{0i} + \pi_{1i}(\text{Year}-2008) + \pi_{2i}(\text{Year}-2008)^2 + \varepsilon_{ti}; \text{ where}$$

Y_{ti} is the proportion of students achieving at or above proficiency in year t at school i ;

π_{0i} is the proportion of students achieving at or above proficiency at Year-2008 = 0;

π_{1i} is the linear coefficient indicating magnitude and direction change in the outcome per year in school i ;

π_{2i} is the curvilinear rate of acceleration in the achievement trajectory curve or school i ;

and

ε_{ti} is a random level-1 error term assumed normally distributed with variance = 1.0.

The numbers of students tested in each subject each year were employed as level-1 weights.

Level 2 models. The level-2 (between schools) models were specified as:

$$\pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11}(\text{Program})_i + r_{1i}$$

$$\pi_{2i} = \beta_{20} + r_{2i}$$

where

π_{0i} is the proportion of students achieving at or above proficiency at Year-2008 = 0 in year i ;

β_{00} is the grand mean proportion of students achieving at or above proficiency at Year-2008 = 0;

r_{0i} is a random level 2 error term representing variation in mean proportions across years;

π_{1i} is the linear trend in performance in school i ;

β_{10} is the grand mean linear trend in performance across schools;

β_{11} is the coefficient associated with the interaction of EDP program status and the linear trend in school performance;

r_{1i} is a random level 2 error term representing variation in linear slopes across schools;

π_{2i} is the quadratic trend in performance in school i ;

β_{20} is the grand mean quadratic trend in performance across schools;

r_{2i} is a random level 2 error term representing variation in quadratic trends across schools.

Results

Reading: Factorial Analyses of Variance

Across all grade levels, 2 (Program) X 5 (Year) factorial analysis of variance revealed statistically significant effects for program ($F_{1,227566} = 1720.3, p < .001$), year ($F_{1,227566} = 387.1, p < .001$), and a statistically significant *program X year* interaction effect ($F_{4,227566} = 112.4, p < .001$). Due to the interaction effect, a graphic profile was constructed and pairwise comparisons within years were conducted to determine the precise nature of the interaction (see Figure 2).

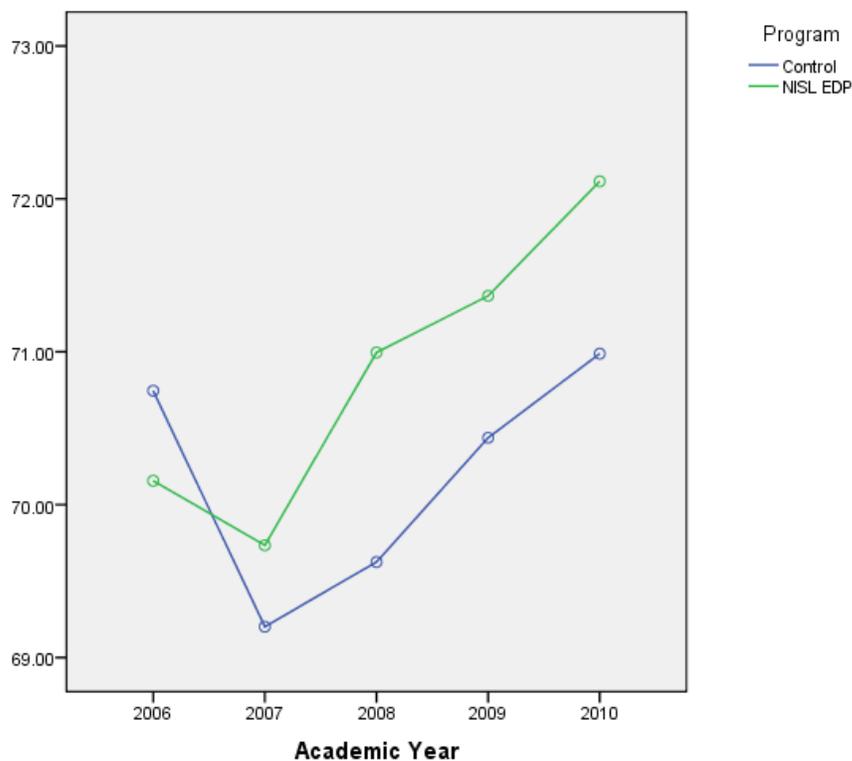


Figure 2. *Observed mean proportion of students achieving proficiency in reading/ English Language Arts by year: Control schools versus NISL EDP schools.* Note: Results include EDP schools that started in 2007 and in 2008.

Pairwise comparisons indicated no statistically significant differences between EDP and comparison schools in 2006 or 2007, whereas NISL schools had statistically significantly higher mean proportions of students reaching proficiency in 2007 (70.1% vs. 67.3%; $F_{1,43846} = 405.4, p < .001$), 2008 (70.9% vs. 68.6%; $F_{1,52555} = 345.1, p < .001$), 2009 (72.4% vs. 67.8%; $F_{1,33216} = 893.8, p < .001$), and 2010 (73.3% vs. 70.9%; $F_{1,53844} = 464.2, p < .001$).

Reading: Hierarchical Linear Model Results

The hierarchical linear model improves upon the analysis of variance model by accounting for the non-linear, or “curvy,” aspects of school performance trajectories to better isolate program effects. Table 2 provides estimates of the fixed effects estimated in the hierarchical linear model for reading. A statistically significant fixed effect was observed for year ($\beta = +0.78, t = 78.72, df = 228016, p < .001$), indicating that, on average, all schools improved the percent proficient in reading/ELA by about 8/10ths of one percent each year. The year-squared (quadratic) effect was also statistically significant ($\beta = +0.40, t = 68.40, df = 228016, p < .001$), which is illustrated by the S-shaped NISL curve and U-shaped comparison school curve from 2006-2010 (see Figure 3). Finally, the *program X year* was statistically significant, ($\beta = +0.54, t = 38.59, df = 228016, p < .001$), which indicates that NISL EDP schools improved at an average rate of about +0.5% greater than the comparison schools *each year*. In other words, by the end of the four years post-baseline, the typical NISL EDP school had about 2.16% more students achieving proficiency in reading/English Language Arts than otherwise would be expected. This translates to about 1,225 more students achieving reading/ELA proficiency from 2008-2010 in the NISL study schools than would have been expected otherwise.

Table 2

Estimates of Fixed Effects: Reading/ELA

Parameter	β	s.e.	df	t	p
Intercept	69.09	1.41	200.0	48.95	.000
Program	0.56	1.99	199.9	-.27	.781
Year	.78	.009	228016.5	78.72	.000
Year-squared	.40	.006	228016.3	68.40	.000
Program X Year	0.54	.014	228016.5	38.59	.000

$N = 228,221$.

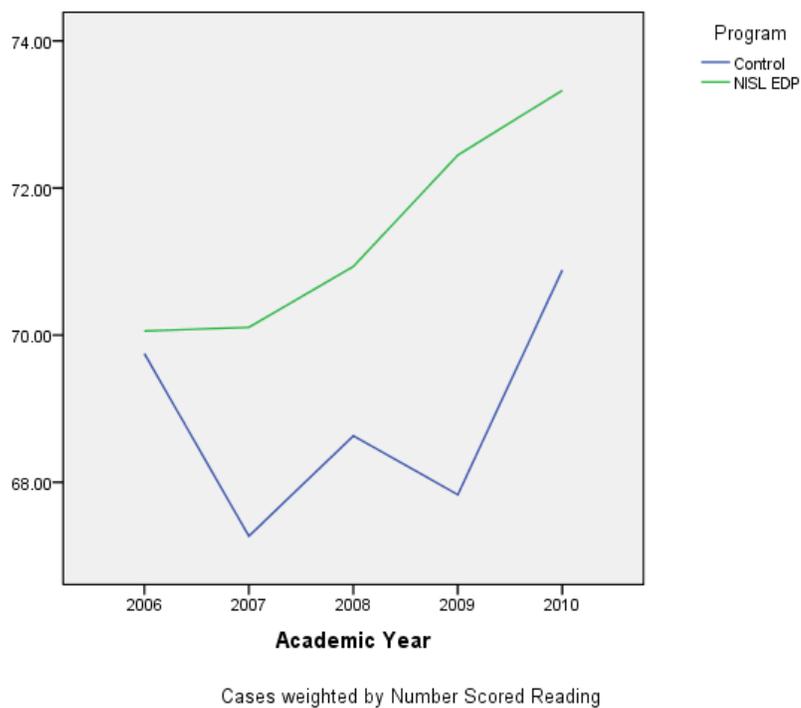


Figure 3. *Hierarchical linear modeling results in reading weighted by number of students at Level 1.*

Mathematics: Factorial Analysis of Variance

A 2 (Program) X 5 (Year) factorial analysis of variance revealed statistically significant effects for program ($F_{1,228211} = 3839.7, p < .001$), year ($F_{1,228211} = 940.1, p < .001$), and a statistically significant *program X year* interaction effect ($F_{4,228211} = 125.0, p < .001$). Due to the interaction effect, a graphic profile was constructed and pairwise comparisons within years were conducted to determine the precise nature of the interaction (see Figure 4).

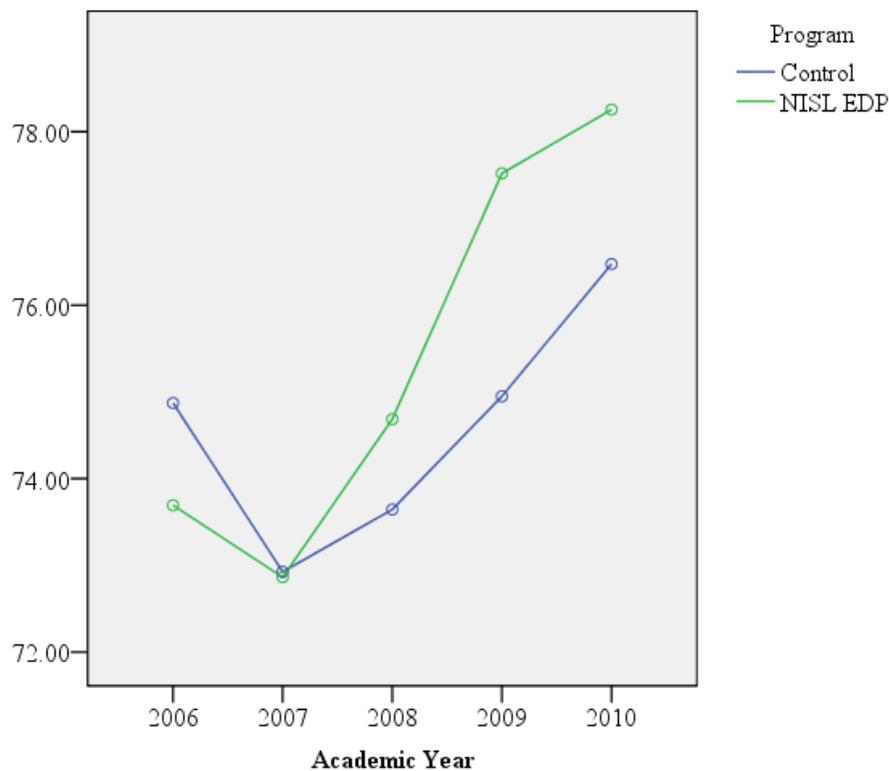


Figure 4. *Observed mean proportion of students achieving proficiency in mathematics by year: Control schools versus NISL EDP schools. Note: EDP program implementation began in 2007 and was completed in 2009.*

Pairwise comparisons indicated that comparison schools had a statistically higher mean proportion of students achieving proficiency in 2006 (74.9% versus 73.7%; $F_{1,44105} = 226.3$, $p < .001$) and no statistically significant difference in 2007. NISL schools had statistically significantly higher mean proportions of students reaching proficiency in mathematics in 2008 (76.2% vs. 71.7%; $F_{1,52555} = 978.7$, $p < .001$), 2009 (76.8% vs. 69.3%; $F_{1,33216} = 1598.6$, $p < .001$), and 2010 (79.6% vs. 75.9%; $F_{1,53844} = 821.5$, $p < .001$).

Mathematics: Hierarchical Linear Model Results for All Schools

Table 3 provides estimates of the fixed effects estimated in the hierarchical linear model for reading. A statistically significant fixed effect was observed for year ($\beta = +1.33$, $t = 126.39$, $df = 227371$, $p < .001$), indicating that, on average, all schools improved the percent proficient in mathematics by about 1.3% each year. The year-squared (quadratic) effect was also statistically significant ($\beta = +0.46$, $t = 74.07$, $df = 227371$, $p < .001$), which is illustrated by the accelerated growth in the NISL curve and U-shaped comparison school curve between 2006-2010 (see Figure 5). Finally, the *program X year* was statistically significant, ($\beta = +0.48$, $t = 32.83$, $df = 227371$, $p < .001$), which indicates that NISL EDP schools improved at an average rate of about +0.5% proficient each year over and above the rate of improvement in comparison schools. In other words, by the end of the four years post-baseline, the typical NISL EDP school had about 1.92% more students achieving proficiency in reading/English Language Arts than otherwise would be expected. This translates to about 1,089 more students achieving math proficiency from 2008-2010 in the NISL study schools than would have been expected otherwise.

Table 3

Estimates of Fixed Effects: Mathematics

Parameter	β	s.e.	df	t	p
Intercept	74.46	1.63	200.0	45.66	.000
Program	0.76	2.30	199.99	0.33	.744
Year	1.33	.011	227371.5	126.39	.000
Year-squared	0.46	.006	227371.3	74.07	.000
Program X Year	0.48	.015	227371.4	32.83	.000

$N = 227,576$.

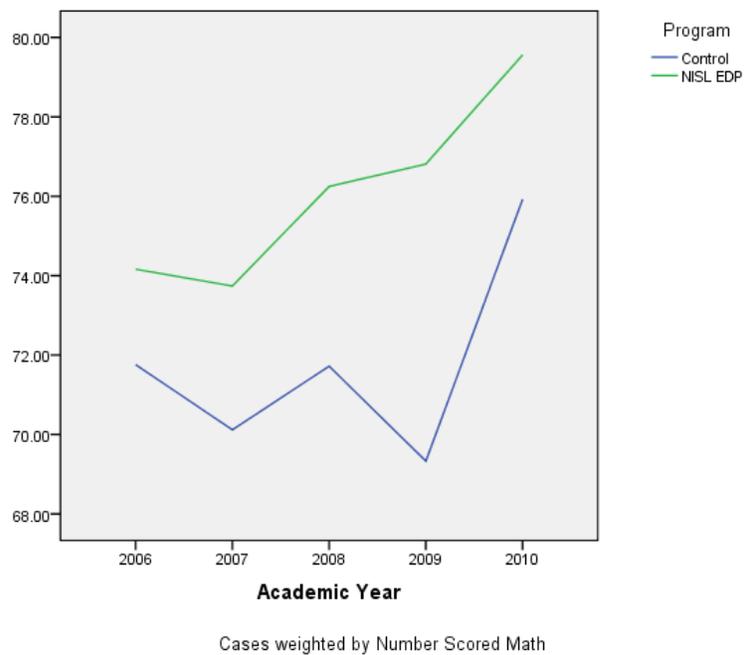


Figure 5. *Hierarchical Linear Modeling Results in Mathematics Weighted by Number of Students at Level 1.*

Hierarchical Linear Modeling Results by School Level in Reading and Mathematics

Table 4 provides tests of *program X year* interaction effects by school grade level configuration in both reading/ELA and mathematics. Statistically significant effects were observed in reading at all grade levels: elementary ($\beta = +0.18$; $t = 8.48$; $df = 115,447$; $p < .001$), middle ($\beta = +1.04$; $t = 121.6$; $df = 82,659$; $p < .001$); and high ($\beta = +0.81$; $t = 21.83$; $df = 29,904$; $p < .001$). Likewise, positive statistically significant program X year interaction effects were observed in mathematics at all grade levels: elementary ($\beta = +0.17$; $t = 8.46$; $df = 115,191$; $p < .001$), middle ($\beta = +0.51$; $t = 23.55$; $df = 82,395$; $p < .001$), and high ($\beta = +2.37$; $t = 47.84$; $df = 29,779$; $p < .001$). As shown in Figure 6, although all effects were statistically significant, much larger effects were observed for middle and high schools.

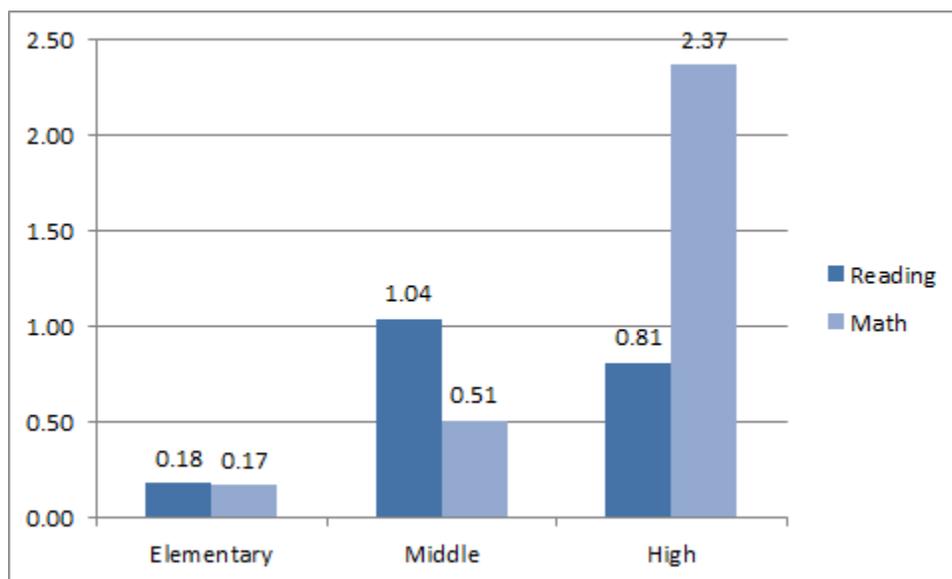


Figure 6. *Annual Acceleration in Percentage Proficient Gains in NISL versus Comparison Schools: Program X Year Interaction Effects by School Grade Level Configuration.*

Table 4. Hierarchical Linear Modeling *Program X Year* Interaction Effects by Grade Level.

Subject	Grade Level	β	t	df	p
Reading/ELA	Elementary	+0.18	8.48	115447	<.001
	Middle	+1.04	121.61	82659	<.001
	High	+0.81	21.83	29904	<.001
Mathematics	Elementary	+0.17	8.46	115191	<.001
	Middle	+0.51	23.55	82395	<.001
	High	+2.37	47.84	29779	<.001

Discussion

This study employed traditional analysis of variance to examine unadjusted longitudinal performance patterns in school performance in schools led by NISL EDP pilot cohort principals versus individually matched comparison schools. In both reading/ELA and mathematics, NISL schools had lower percentages of students achieving proficiency in these subject areas in 2006 (the difference favoring comparison schools was statistically significant in mathematics). As shown in Figures 5 and 6 above, NISL and comparison school performance trend lines “crossed over” between 2006 and 2008, with NISL schools having statistically significant higher levels of performance in both subject areas in 2008, 2009, and 2010. Although these simple models cannot account for non-linearity in trend lines and provide somewhat less precise estimates of annual program effects than the hierarchical linear models, the observed cross-over pattern on unadjusted outcomes is a powerful indication that implementation of the NISL program was associated with a substantial and sustained improvement in school performance during the study period from 2006-2010.

Hierarchical linear modeling analyses revealed statistically significant *program X year* interaction effects for both reading/ELA and mathematics. Across schools of all grade levels, NISL schools performance accelerated about +0.48 and +0.54 additional percent proficient each year in reading/ELA and mathematics, respectively. To put these statistics in perspective, Figure 7 illustrates the number of additional students estimated to have achieved proficiency in reading and mathematics between 2008 and 2010 in the NISL pilot cohort schools that met study inclusion criteria. These findings are consistent with previous state-wide studies conducted in Pennsylvania (Nunnery, Ross, Yen, 2010a), and Massachusetts (Nunnery, Ross, Yen,

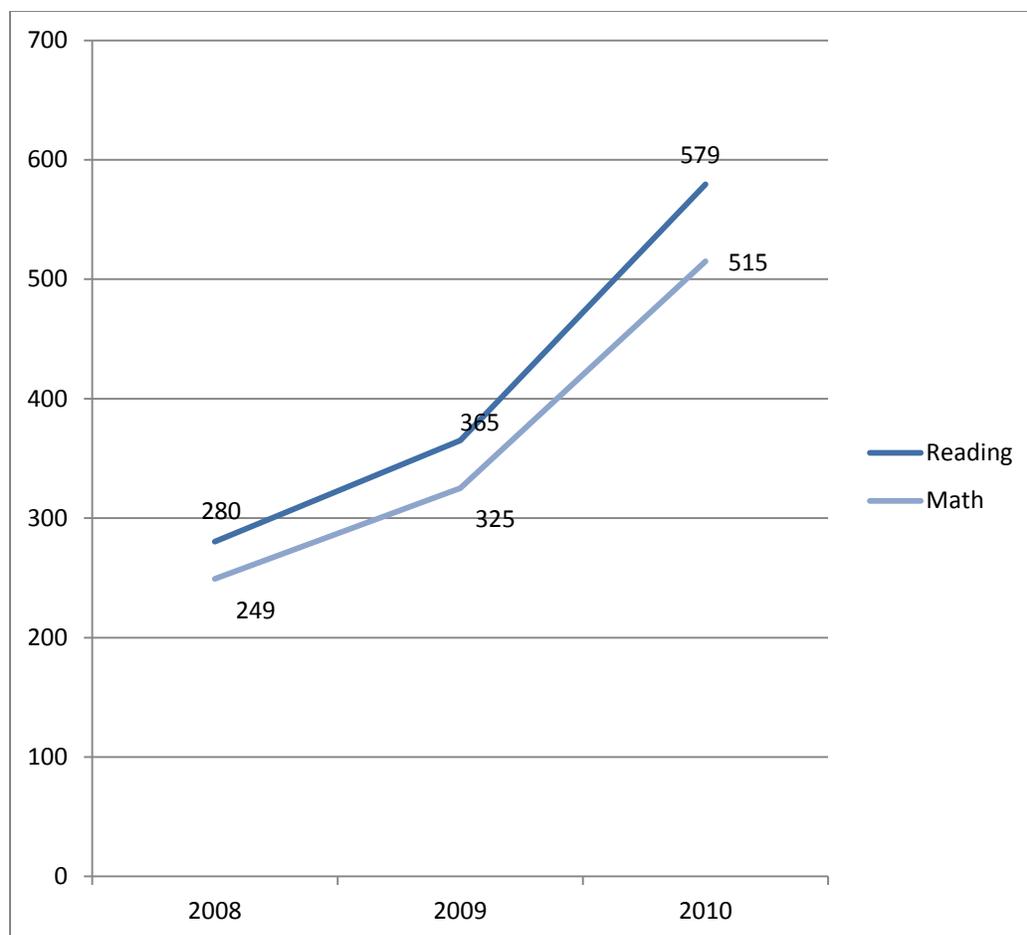


Figure 7. *Estimated Number of Additional Students Reaching Proficiency in Reading and Mathematics By Year in Pennsylvania NISL Schools: 2008-2010 Study Schools.*

2010b), which found positive associations between student achievement patterns and EDP program participation by school leaders. In Pennsylvania, NISL schools had statistically significantly higher proportions of students achieving proficiency in both reading and mathematics, with much larger proportions achieving proficiency in mathematics (see Figure 7 for Pennsylvania summary results from the previous study). As shown in Figure 8, schools led by EDP program completers achieved substantial and statistically significant gains in mathematics relative both to comparison schools ($d = +0.10$) and the Commonwealth as a whole ($d = +0.08$).

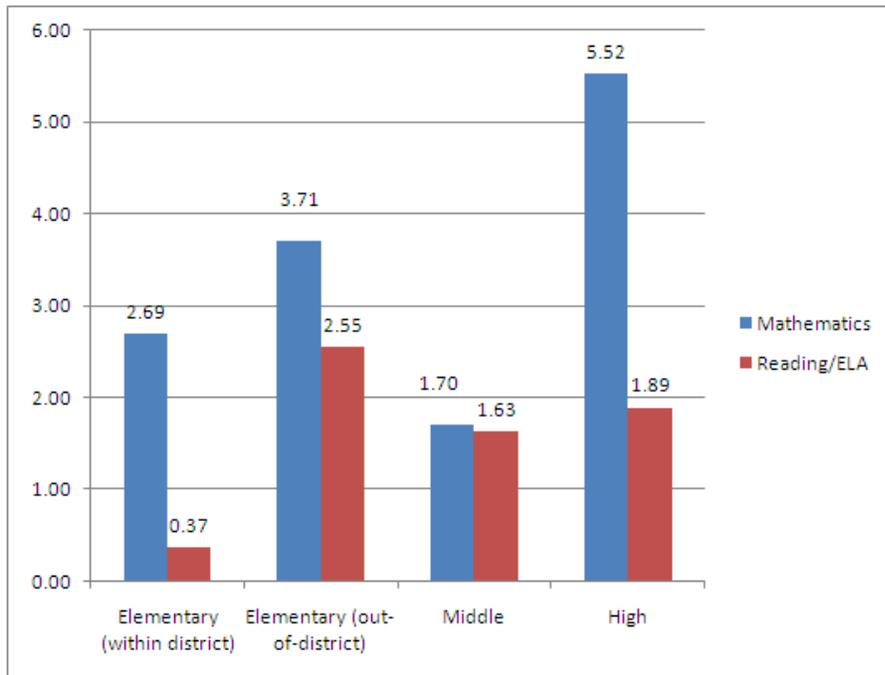


Figure 7. NISL Effects in Pennsylvania: Summary Results from Prior 2006-2009 Study of the Pennsylvania implementation of the EDP program: Differences in Percentages of Students Achieving Proficiency Relative to Matched Comparison Schools (from Nunnery, Ross, & Yen, 2010a).

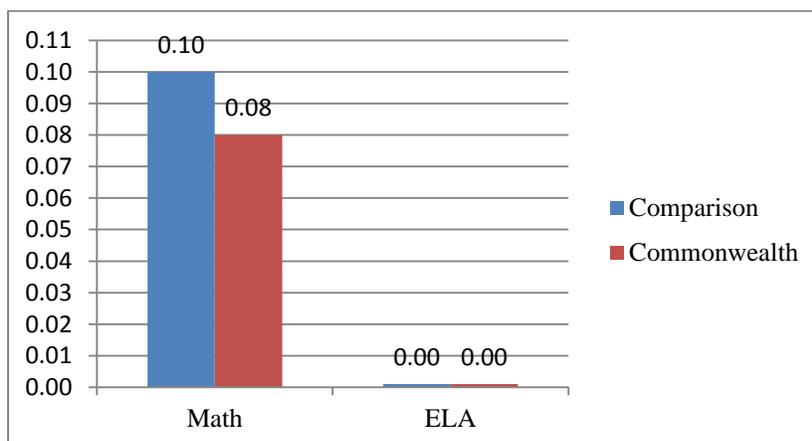


Figure 8. NISL Effects in Massachusetts: Summary Results from the 2006-2009 Study of the Massachusetts implementation of the EDP program: Effect Size Estimates Expressed as Cohen's d. (from Nunnery, Ross, & Yen, 2010b).

Although statistically significant effects were associated with the NISL program at all grade levels and in both subject areas, much stronger improvements in school-level performance were observed in NISL middle schools and high schools. This finding is particularly noteworthy, as there is little systematic evidence that any of the many high school reforms attempted to date have had a positive effect on student achievement (Fleischman & Heppen, 2009). In particular, very large gains in mathematics performance were observed in NISL high schools. As shown in Figure 9, NISL high schools had only a +0.6% advantage in terms of percentage of students achieving proficiency in mathematics in 2006 (46.3% versus 45.7%), whereas by 2010 NISL high schools had a 12% advantage (62.4% versus 50.5%). Larger gains in high school mathematics for both NISL and comparison schools might have been expected given the low levels of initial performance in 2006 due to regression to the mean, but the difference in comparative gains is striking.

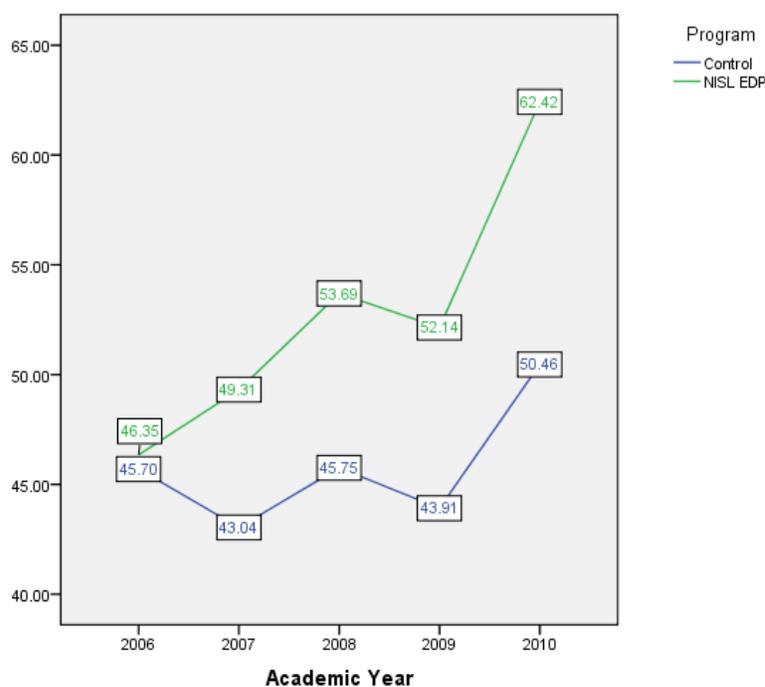


Figure 9. *Percentage Scoring Proficient in Mathematics from Pre-implementation Baseline (2006) through 2010 by Program: High Schools in Pennsylvania EDP Pilot Cohort Study.*

The greatest acceleration in math performance in NISL schools occurred between 2009 and 2010, the first year following completion of the EDP program for all pilot cohort principals. Given the staggered implementation (about half completing in 2008 and half completing in 2009), and the expectation of somewhat more modest effects during the two-year professional development sequence. NISL principals are expected to begin implementing strategies *during* the sequence, but full program effects would not be expected until post completion. The greater gains observed post-completion for all pilot cohort participants lends support to the notion that the effects are attributable to the program, although lack of randomization militates against a straightforward causal interpretation of the findings.

The results of this study represent highly promising evidence that the NISL Executive Development Program for school leaders may result in statistically significant, substantial, and sustained improvements in student performance in reading and mathematics, particularly in the challenging context of secondary schools. This is particularly noteworthy given that the program is highly cost-effective, with current median participant costs of about \$4,000. In the context of the current study, this amounts to about \$117 per additional student achieving proficiency already observed: if the observed trends continued for one additional year, that cost would drop to about \$69 per additional student achieving proficiency in either reading or mathematics. Prior research has also indicated that the EDP program can be implemented with high fidelity in a cost-effective manner (Meristem Group, 2009).

References

- Fleischman, S., & Heppen, J. (2009). Improving low-performing high schools: Searching for evidence of promise. *Future of Children, 19*, 105-133.
- Meristem Group (2009). *National Institute for School Leadership (NISL): Massachusetts program implementation 2005-2008*. Boston, MA: Meristem Group.
- Nunnery, J.A., Ross, S.M., & Yen, C. (2010a). *The effect of the National Institute for School Leadership's Executive Development Program on school performance trends in Pennsylvania*. Norfolk, VA: The Center for Educational Partnerships at Old Dominion University.
- Nunnery, J.A., Ross, S.M., & Yen, C. (2010b). *An examination of the effects of a pilot of the National Institute for School Leadership's Executive Development Program on school performance trends in Massachusetts*. Norfolk, VA: The Center for Educational Partnerships at Old Dominion University.

Date: _____
 Leader Self-Assessment
 Evaluator Assessment

Domain 1: Strategic/Cultural Leadership

The school leader will systematically and collaboratively develop a positive culture to promote continuous student growth and staff development. The leader articulates and models a clear vision of the school’s culture that involves students, families, and staff.

Component	Failing	Needs Improvement	Proficient	Distinguished
<p><i>1a: Creates an Organizational Vision, Mission, and Strategic Goals:</i></p> <p>The school leader plans strategically and creates an organizational vision, mission, and goals around personalized student success that is aligned to LEA goals.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to develop a school wide vision, mission, or strategic goals.</p> <p>Fails to demonstrate the involvement of staff and stakeholders in a strategic process that leads to the development of the school’s vision, mission, and goals.</p>	<p>Develops school wide vision, mission, and strategic goals based on his/her own individual beliefs regarding future needs of student performance, with limited evidence of stakeholder involvement.</p>	<p>Implements a process that includes stakeholders for developing a shared vision and strategic goals for student achievement that results in rigor and relevance for students and staff.</p> <p>Maintains a focus on the vision and strategic goals throughout the school year.</p> <p>Ensures that staff incorporates the school’s vision, mission, and strategic goals in their instructional plans to assure that students achieve expected outcomes.</p>	<p>... and</p> <p>Designs, initiates, and implements collaborative processes to collect and analyze data about the school’s progress for the periodic review and revision of the school’s vision, mission, and strategic goals.</p> <p>Systematically ensures that the school’s vision, mission, values, beliefs and goals drive decisions that positively influence the culture of the school.</p>

Framework for Leadership

<p><i>Ib: Uses Data for Informed Decision Making:</i></p> <p>The school leader analyzes and uses multiple data sources to drive effective decision-making.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to demonstrate the ability to analyze or use data to drive effective decision-making.</p>	<p>Infrequently uses data and assessments to monitor progress.</p> <p>Exhibits the inability to develop the capacity of staff and other stakeholders to use data for decision-making.</p>	<p>Collects, analyzes, monitors, and uses data systematically regarding the school's progress in driving informed decision-making for the attainment of strategic goals and objectives.</p> <p>Develops the capacity of staff and other stakeholders to use data for decision-making.</p>	<p>... and</p> <p>Activates and sustains a school wide system for monitoring and evaluating progress toward achieving school goals and student outcomes.</p> <p>Listens, evaluates, and considers staff and other stakeholders input regarding recommended activities and initiatives</p>
<p><i>Ic: Builds a Collaborative and Empowering Work Environment:</i></p> <p>The school leader develops a culture of collaboration, distributive leadership, and continuous improvement conducive to student learning and professional growth. The school leader empowers staff in the development and successful implementation of initiatives that better serve students, staff, and the school.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to demonstrate the involvement of staff and stakeholders in discussions and decisions regarding school issues.</p>	<p>Frequently makes unilateral decisions (uses distributive leadership infrequently).</p> <p>Inconsistently includes parents, families, and the larger school community in the decision-making processes.</p> <p>Articulates the importance of building a sense of empowerment among staff, but only sporadically incorporates activities, tools, and protocols to develop empowerment among staff.</p>	<p>Creates a collaborative work environment predicated upon cooperation among and between students, parents, staff, and the community.</p> <p>Consistently engages in shared decision-making and distributive leadership.</p> <p>Actively models behaviors that promote a sense of empowerment among staff and stakeholders.</p>	<p>... and</p> <p>Empowers staff and other stakeholders to assume responsibility for making decisions regarding the school culture and student achievement.</p> <p>Establishes an environment where staff and other stakeholders:</p> <ul style="list-style-type: none"> • Select and implement effective improvement strategies. • Assess and monitor progress towards achieving the vision, mission, and strategic goals. <p>Lead planning and monitoring efforts.</p>

<p><i>Id: Leads Change Efforts for Continuous Improvement:</i></p> <p>The school leader systematically guides staff through the change process to positively impact the culture and performance of the school.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to identify the importance of the change process with no provision for positively impacting the culture and performance of the school.</p>	<p>Articulates the importance of the change process; however, when change occurs, it is only through random processes.</p>	<p>Implements a change process to ensure continuous school improvement.</p>	<p>... and</p> <p>Drives major initiatives that help students become college and career ready.</p> <p>Systematically examines the status quo, identifies beneficial changes, and leads the change process to successful completion.</p>
<p><i>Ie: Celebrates Accomplishments and Acknowledges Failures:</i></p> <p>The school leader utilizes lessons from accomplishments and failures to positively impact the culture and performance of the school.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to utilize lessons from accomplishments and failures to positively impact the culture and performance of the school.</p>	<p>Inconsistently utilizes lessons from accomplishments and failures to positively impact the culture and performance of the school.</p>	<p>Recognizes individual and collective contributions in a systematic manner toward attainment of strategic goals.</p> <p>Utilizes failure as an opportunity to improve school culture and student performance.</p>	<p>... and</p> <p>Utilizes recognition, reward, and advancement as a way to promote the accomplishments of the school.</p>

Framework for Leadership

Domain 2: Systems Leadership

The school leader will ensure that the school has processes and systems in place for budgeting, staffing, problem solving, communicating expectations and scheduling that result in organizing the work routines in the building. The school leader must efficiently, effectively, and safely manage the building to foster staff accountability and student achievement.

Component	Failing	Needs Improvement	Proficient	Distinguished
<p>2a: Leverages Human and Financial Resources:</p> <p>The school leader establishes systems for marshaling all available resources to better serve students, staff, and the school.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to systematically allocate human and financial resources that support the vision, mission, and strategic goals of the school.</p>	<p>Utilizes systems for allocating human and financial resources that are not transparent.</p>	<p>Designs transparent systems to equitably manage human and financial resources.</p> <p>Ensures the strategic allocation and equitable use of human and financial resources to meet instructional goals and support teacher needs.</p>	<p>... and</p> <p>Integrates school, LEA, and community resources to maximize the efficiency of school operations.</p> <p>Uses data and feedback to assess the success of funding and program decisions.</p>
<p>2b: Ensures School Safety:</p> <p>The school leader ensures the development and implementation of a comprehensive safe schools plan that includes prevention, intervention, crisis response, and recovery.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to develop and implement a comprehensive safe schools plan that includes prevention, intervention, crisis response, and recovery.</p>	<p>Lacks a process for reviewing/revising the school safety plan.</p> <p>Lacks a process to collect data on the effectiveness of the school safety plan that includes prevention, intervention, crisis response, and recovery.</p>	<p>Reviews, analyzes and adjusts school safety and discipline plans based on school data, crisis feedback, and current regulations/mandates.</p> <p>Maintains and creates open communication processes that allow for proactive identification and intervention of potential incidents.</p> <p>Communicates to stakeholders regarding safety issues in a clear, appropriate, and timely manner.</p>	<p>... and</p> <p>Incorporates active involvement of various safety agencies in the development, implementation, and evaluation of the comprehensive safe schools plan.</p>

<p>2c: <i>Complies with Federal, State, and LEA Mandates:</i></p> <p>The school leader designs protocols and processes in order to comply with federal, state and LEA mandates.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to comply with policies, mandates, and contractual agreements in a timely and/or complete manner.</p>	<p>Inconsistently complies with federal, state, and LEA mandates and all contractual agreements in a timely and/or complete manner.</p>	<p>Designs protocols and processes in order to comply with federal, state and LEA mandates.</p> <p>Consistently complies with federal, state, and LEA mandates and all contractual agreements in a timely and complete manner.</p>	<p>...and</p> <p>Presents federal, state and LEA mandates so that such mandates are viewed as an opportunity for improvement within the school.</p> <p>Identifies opportunities for improvement to develop programs derived from the mandates.</p> <p>Implements related programs supported by the school community.</p>
<p>2d: <i>Establishes and Implements Expectations for Students and Staff:</i></p> <p>The school leader establishes and implements clear expectations, structures, rules, and procedures for students and staff.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to establish clear expectations, structures, rules, and procedures for students and staff.</p>	<p>Utilizes only school rules and procedures required by LEA administration and/or school policy.</p> <p>Inconsistently communicates and enforces expectations, rules, and procedures for students and staff.</p>	<p>Engages students and staff members in developing expectations for learning and improved performance.</p> <p>Creates and revises rules and procedures to maintain a safe and positive school culture conducive to student learning.</p> <p>Communicates and enforces clear expectations, structures, and fair rules and procedures for students and staff.</p>	<p>... and</p> <p>Empowers staff to monitor their own performance and exceed school-wide expectations.</p> <p>Encourages students to monitor their performance and strive to exceed expectations set by their teachers, parents and themselves.</p>

Framework for Leadership

<p>2e: <i>Communicates Effectively and Strategically:</i></p> <p>The school leader strategically designs and utilizes various forms of formal and informal communication with all staff and stakeholders.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to develop a coherent plan to effectively communicate with all staff and stakeholders.</p>	<p>Defines a communications plan for staff and stakeholders; however, actual communications lack purpose, clarity, consistency, or regularity.</p>	<p>Designs and utilizes a system of open communication that provides for the timely, responsible sharing of information to, from, and with staff and stakeholders.</p> <p>Provides information in various formats in multiple ways through different media in order to ensure communication with staff and stakeholder</p>	<p>... and</p> <p>Ensures that staff and stakeholders are aware of school goals for instruction and achievement, activities used to meet these goals, and progress toward meeting these goals.</p>
<p>2f: <i>Manages Conflict Constructively:</i></p> <p>The leader effectively and efficiently manages the complexity of human interactions and relationships, including those among and between parents/guardians, students, and staff.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to develop and implement conflict management processes to manage the complexity of human interactions and relationships.</p>	<p>Inconsistently implements processes to resolve problems and/or areas of conflict within the school.</p> <p>Interacts with students, staff and other stakeholders primarily on an as needed basis in order to defuse potentially stressful situations.</p>	<p>Consistently resolves school-based problems/conflicts in a fair, democratic way.</p> <p>Provides opportunities for affected stakeholders (students, staff, and parents) to express opinions and discusses options to address discordant issues.</p> <p>Implements and reviews solutions that address discordant issues.</p>	<p>... and</p> <p>Provides conflict management and relationship building training for students, staff, and other stakeholders.</p> <p>Empowers students, staff, and others to engage each other in relationship building activities designed to avoid conflict and maintain a positive school culture.</p> <p>Encourages staff and students to accept responsibility for their own actions by adhering to operational norms.</p>

Domain 3: Leadership for Learning

The school leader assures a Standards Aligned System is in place to address the linkage of curriculum, instruction, assessment, and data on student learning and teacher effectiveness based on research and best practices.

Component	Failing	Needs Improvement	Proficient	Distinguished
<p>3a: Leads School Improvement Initiatives:</p> <p>The school leader develops, implements, monitors, and evaluates a School Improvement Plan that provides the structure for the vision, goals, and changes necessary for improved student achievement.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to develop a School Improvement Plan that provides the structure for the vision, goals, and changes necessary for improved student achievement.</p>	<p>Develops a School Improvement Plan; however, the plan lacks clear and consistent processes and systems to improve student achievement.</p>	<p>Develops a School Improvement Plan, as well as establishes clear and consistent processes and systems to:</p> <ul style="list-style-type: none"> • Implement the School Improvement Plan. • Monitor and evaluate progress toward achieving school improvement goals and student outcomes. • Revise school improvement goals and outcomes based on data analysis. 	<p>... and</p> <p>Incorporates principles of continuous improvement into a School Improvement Plan, which positively impacts the school’s culture and exceeds expectations of student achievement.</p>

Framework for Leadership

<p>3b: Aligns Curricula, Instruction, and Assessments:</p> <p>The school leader ensures that the adopted curricula, instructional practices, and associated assessments are implemented within a Standards Aligned System. Data are used to drive refinements to the system.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to monitor that the LEA’s curricula are being implemented.</p> <p>Fails to engage staff in curricula planning and instruction.</p>	<p>Inconsistently monitors that the LEA’s curricula are implemented with fidelity throughout the school.</p> <p>Inconsistently engages staff in curricula planning and instruction.</p>	<p>Consistently ensures that the LEA’s curricula are implemented with fidelity throughout the school.</p> <p>Aligns curricula with assessments and instructional material.</p> <p>Engages staff in curricula planning and instruction based upon state and local assessments.</p> <p>Creates opportunities to collaboratively use data/assessments to drive instructional decisions and practices.</p>	<p>... and</p> <p>Engages staff to assesses curricula for strengths and weaknesses.</p> <p>Reports data and recommendations to curriculum committee for refinement of the LEA’s curricula.</p>
<p>3c: Implements High Quality Instruction:</p> <p>The school leader monitors progress of teachers and staff. In addition, the school leader conducts formative and summative assessments in measuring teacher effectiveness in order to ensure that rigorous, relevant, and appropriate instruction and learning experiences are delivered to and for all students.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to monitor the effectiveness of professional staff in the domains of:</p> <ul style="list-style-type: none"> • Planning and Preparation. • Classroom Environment. • Instruction. • Professional Responsibilities. 	<p>Inconsistently monitors the effectiveness of and timely feedback to professional staff in the domains of:</p> <ul style="list-style-type: none"> • Planning and Preparation. • Classroom Environment. • Instruction. • Professional Responsibilities. <p>Lacks participation in ongoing professional development activities to better monitor and coach the use of effective instructional and assessment practices.</p>	<p>Consistently monitors the effectiveness of and timely feedback to professional staff in the domains of:</p> <ul style="list-style-type: none"> • Planning and Preparation. • Classroom Environment. • Instruction. • Professional Responsibilities <p>Participates in professional development activities, including inter-rater reliability, to better monitor and coach the use of effective instructional and assessment practices.</p>	<p>...and</p> <p>Collaboratively works with staff members to:</p> <ul style="list-style-type: none"> • Identify professional development needs based upon observation data. • Plan short and long term professional development activities to address identified needs based upon observation data. • Monitor performance following professional development to ensure the application of lessons learned.

<p>3d: Sets High Expectations for All Students:</p> <p>The school leader holds all staff accountable for setting and achieving rigorous performance goals for all students.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to hold all staff accountable for setting and achieving rigorous performance goals for all students.</p>	<p>Inconsistently holds all staff accountable for setting and achieving rigorous performance goals for all students.</p>	<p>Articulates a belief in high measureable goals for all students and staff.</p> <p>Leads school efforts to set and monitor learning goals for all students and establish safety nets for struggling students.</p> <p>Holds every staff member responsible and accountable for ensuring that all students achieve the rigorous outcomes established for them.</p>	<p>... and</p> <p>Models high expectations for staff and other stakeholders by systematically pursuing performance goals for all students.</p>
<p>3e: Maximizes Instructional Time:</p> <p>The school leader creates processes which protect teachers from disruption of instructional and preparation time.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to protect teachers from disruption of instructional and preparation time.</p>	<p>Sporadically permits interruptions to instructional and planning time.</p>	<p>Implements processes and schedules in a systematic manner to protect instructional and planning time from interruptions.</p>	<p>... and</p> <p>Structures the school schedule to increase opportunities for teachers to have collaborative planning time.</p> <p>Systematically monitors the effect of the master schedule on collaborative planning and student achievement.</p>

Framework for Leadership

Domain 4: Professional and Community Leadership

The school leader promotes the success of all students, the positive interactions among building stakeholders, and the professional growth of staff by acting with integrity, fairness and in an ethical manner.

Component	Failing	Needs Improvement	Proficient	Distinguished
<p>4a: Maximizes Parent and Community Involvement and Outreach:</p> <p>The school leader designs structures and processes, which result in parent and community engagement, support, and ownership for the school.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to design structures and processes, which result in a lack of parent and community engagement, support, and ownership for the school.</p>	<p>Efforts for community outreach do not result in meaningful support for teaching and learning.</p> <p>Unilaterally designs structures and processes that result in limited involvement of parents and other stakeholders.</p>	<p>Creates systems and engages parents/ guardians and all community stakeholders in a shared responsibility for student and school success reflecting the community’s vision of the school.</p> <p>Collaboratively works to establish a culture that encourages and welcomes families and community members and seeks ways in which to engage them in student learning.</p>	<p>... and</p> <p>Proactively develops relationships with parents/guardians and the community so as to develop good will and garner fiscal, intellectual and human resources that support specific aspects of the school’s learning agenda.</p>
<p>4b: Shows professionalism:</p> <p>The leader operates in a fair and equitable manner with personal and professional integrity.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to display honesty in interactions with students, staff, and stakeholders.</p> <p>Fails to recognize student needs and contributes to school practices that result in some students being ill served.</p>	<p>Interacts honestly with students, staff, and stakeholders, but attempts to serve students are inconsistent.</p>	<p>Articulates and demonstrates a personal and professional code of ethics (e.g. AASA, NASSP, PAESSP).</p> <p>Displays high standards of honesty, integrity, and confidentiality in interactions with students, staff, and stakeholders.</p> <p>Actively serves students to ensure that all students receive a fair opportunity to succeed.</p>	<p>... and</p> <p>Holds the highest standards of honesty, integrity, and confidentiality.</p> <p>Proactively serves students, seeking out resources when needed.</p> <p>Makes a concerted effort to challenge negative attitudes or practices to ensure that all students, particularly those traditionally underserved, are honored in the school.</p>

<p>4c: Supports Professional Growth:</p> <p>The school leader supports continuous professional growth of self and others through practice and inquiry.</p>	<p>Fails to satisfy the component as defined.</p> <p>Fails to identify professional growth needs of self and others, which would positively impact the culture and performance of the school.</p>	<p>Implements professional development inconsistently which is not aligned with curricular, instructional, and assessment needs.</p>	<p>Targets professional development toward the improvement of learning experiences, including quality of classroom instruction and the ability of teachers to meet the needs of all students.</p> <p>Plans and routinely participates in professional development focused on improving instructional programs and practices.</p>	<p>...and</p> <p>Ensures that professional development within the school is aligned with curricular, instructional, and assessment needs, while recognizing the unique professional development needs of individual staff members and self.</p>
---	---	--	--	---

Framework for Leadership

The following provides alignment to Legislative Categories and PIL Standards

Domain	Component	Alignment with Legislative Categories	Alignment with the Pennsylvania Inspired Leadership (PIL) Program
Domain 1: Strategic/Cultural Leadership	1a: Creates an Organizational Vision, Mission, and Strategic Goals	<ul style="list-style-type: none"> • Planning and Preparation 	<ul style="list-style-type: none"> • Core Standards 1,3 • Corollary Standard 3
	1b: Uses Data for Informed Decision Making	<ul style="list-style-type: none"> • Planning and Preparation 	<ul style="list-style-type: none"> • Core Standard 3 • Corollary Standards 3, 6
	1c: Builds a Collaborative and Empowering Work Environment	<ul style="list-style-type: none"> • School Environment • Delivery of Service 	<ul style="list-style-type: none"> • Corollary Standards 3, 6
	1d: Leads Change Efforts for Continuous Improvement	<ul style="list-style-type: none"> • Planning and Preparation • School Environment 	<ul style="list-style-type: none"> • Core Standard 1 • Corollary Standards 1,2
	1e: Celebrates Accomplishments and Acknowledges Failures	<ul style="list-style-type: none"> • School Environment • Delivery of Service 	<ul style="list-style-type: none"> • Corollary Standard 1
Domain 2: Systems Leadership	2a: Leverages Human and Financial Resources	<ul style="list-style-type: none"> • Planning and Preparation • Delivery of Service 	<ul style="list-style-type: none"> • Corollary Standards 2,3, 4
	2b: Ensures School Safety	<ul style="list-style-type: none"> • Planning and Preparation • School Environment • Delivery of Service 	<ul style="list-style-type: none"> • Core Standard 3 • Corollary Standards 2, 3
	2c: Complies with Federal, State, and LEA Mandates	<ul style="list-style-type: none"> • Planning and Preparation 	<ul style="list-style-type: none"> • Corollary Standard 2
	2d: Establishes and Implements Expectations for Students and Staff	<ul style="list-style-type: none"> • School Environment 	<ul style="list-style-type: none"> • Corollary Standard 3
	2e: Communicates Effectively and Strategically	<ul style="list-style-type: none"> • Planning and Preparation • School Environment 	<ul style="list-style-type: none"> • Core Standard 1 • Corollary Standard 3
	2f: Manages Conflict Constructively	<ul style="list-style-type: none"> • School Environment 	<ul style="list-style-type: none"> • Corollary Standards 2, 3, 4
Domain 3: Leadership for Learning	3a: Leads School Improvement Initiatives:	<ul style="list-style-type: none"> • Planning and Preparation • Delivery of Service • Professional Development 	<ul style="list-style-type: none"> • Core Standard 1 • Corollary Standards 1, 2, 3, 4
	3b: Aligns Curricula, Instruction, and Assessments	<ul style="list-style-type: none"> • Planning and Preparation • Delivery of Service 	<ul style="list-style-type: none"> • Core Standards 2, 3 • Corollary Standards 1, 3
	3c: Implements High Quality Instruction	<ul style="list-style-type: none"> • Planning and Preparation • Delivery of Service • Professional Development 	<ul style="list-style-type: none"> • Core Standard 3 • Corollary Standards 1, 3, 6
	3d: Sets High Expectations for All Students	<ul style="list-style-type: none"> • School Environment • Delivery of Service 	<ul style="list-style-type: none"> • Core Standards 1, 2, 3 • Corollary Standards 1, 3
	3e: Maximizes Instructional Time	<ul style="list-style-type: none"> • Delivery of Service 	<ul style="list-style-type: none"> • Core Standard 3 • Corollary Standards 1, 2, 3
Domain 4: Professional and Community Leadership	4a: Maximizes Parent and Community Involvement and Outreach	<ul style="list-style-type: none"> • Planning and Preparation • School Environment • Delivery of Service 	<ul style="list-style-type: none"> • Corollary Standards 2, 3, 4, 5
	4b: Shows professionalism	<ul style="list-style-type: none"> • School Environment 	<ul style="list-style-type: none"> • Corollary Standards 2, 4, 5
	4c: Supports Professional Growth	<ul style="list-style-type: none"> • School Environment • Delivery of Service • Professional Development 	<ul style="list-style-type: none"> • Core Standard 2 • Corollary Standard 6

The following documents were used as reference in the development of this document:

Colorado Department of Education. (November 2011). *Rubric for Evaluating Colorado's Principals and Assistant Principals*. Denver, Co.

Danielson, C. (2011). *Framework for Teaching Evaluation Instrument*. The Danielson Group

Delaware Department of Education. (August 2008). *Delaware Performance Appraisal System*. Dover, DE.

North Carolina Department of Public Instruction. (May 2008). *Principal and Assistant Principal Evaluation Process*. Raleigh, NC.

Pittsburgh Public Schools. (2009). *Administrator Performance Standard Rubric Revised 09-10*. Pittsburgh, PA.

State of Washington: Office of Superintendent of Public Instruction (July 2011). *Teacher and Principal Evaluation Pilot*. Olympia, WA

Tennessee Department of Education. (September 2011). *Tennessee's Principal Evaluation System*. Nashville, TN

Principal Effectiveness Framework for Leadership Stakeholder Information

Stakeholder	Position	Organization
Toni Arnold	Assistant Director	Dauphin County Technical School
Susan Bigger	Director of Federal Programs	Williamsport Area School District
Karen Boardman	Principal, Dodd Elementary School	Allentown School District
Joel Boyd	Assistant Superintendent, Academic Division 6	Philadelphia School District
Gail Cooper	Principal, Pottstown Middle School	Pottstown School District
Jackie Cullen	Director	Pennsylvania Association of Career and Technical Administrators
Della Gentile	Director of School Improvement Services	Northwest Tri-County Intermediate Unit
Beth Haldeman	Principal	Reamstown Elementary School
Michael Healy	Superintendent of Schools	Tunkhannock School District
Bill Jimenez	Principal, McCaskey High School	School District of Lancaster
Josh Keene	Principal, Lincoln Middle School	School District of Lancaster
Jerri Lippert	Chief Academic Officer	School District of Pittsburgh
John Mastillo	Superintendent of Schools	Blacklick Valley School District
Dean Maynard	Assistant Executive Officer	Northwest Tri-County Intermediate Unit
Cindy Muehlbauer	Principal	School District of Pittsburgh
Lawrence Mussoline	Superintendent of Schools	Downingtown Area School District
Jen Ramos	Deputy Chief, Office of Leadership	Philadelphia School District
Rosemary Sheridan	Principal, Pfeiffer-Burleigh Elementary School	Erie School District
Edward Small	Principal, Academy Park High School	Southeast Delco School District
Wanda Suarez	Principal, Lafayette Elementary School	School District of Lancaster
Cathy Tashner	Assistant Superintendent	Susquehanna School District
Brian Toth	Superintendent of Schools	Bellwood-Antis School District
Brian White	Superintendent of Schools	Chartiers Valley School District
Michael Reed	Principal	Williamsport Area School District
Bi Vuong	Senior Associate: Data Quality	Philadelphia School District
David Warren	Director	Lancaster County CTC
Paula Wilson	Principal	School District of Lancaster