



Research Report

Overview of Research on Teacher Professional Development and Effectiveness Evaluation

August 14, 2012

**Prepared for the Senate and House
Committees on Education**

Table of Contents

Introduction	1
Purpose of This Report	1
Indicator of Need for Professional Development	2
Research on Professional Development Opportunities	3
Research on Effective Professional Development	4
Key Principles in Designing Professional Development	4
Content of Professional Development	5
Collaborative Professional Development	5
Job-embedded Professional Development	6
Professional Learning Communities.....	6
Professional Development’s Effect on Student Achievement	7
Evaluating the Effectiveness of Teachers	8
Methods of Teacher Evaluation	9
Arkansas Teacher Evaluation and Support System.....	9
Portfolio as a Measure of Teacher Effectiveness	10
Value-Added Modeling Measure of Teacher Effectiveness	11
Comparison of Methods of Evaluating Teacher Effectiveness	13
Discussion of Strengths and Weaknesses of Each Evaluation Approach	14
Conclusions and Policy Implications	16
References	19
Appendix A	25
Data on Arkansas Professional Development and Teacher Evaluation	25
Table 1. Percentage of Professional Development Provided by Source	25
Table 2. Number of Formative Evaluations in a School Year	26
Table 3. Type of Summative Evaluation Used in Arkansas	26
Table 4. Total Minutes of Classroom Observation in a Year For Summative Evaluations	27
Table 5. Number of Teacher Separations from school Due to Performance	27
PD and Evaluation Issues Arising from Surveys, Site Visits and ADE Officials	27
Appendix B	29

INTRODUCTION

In the present education policy environment, improving quality of instruction and teaching effectiveness has become a high priority in U. S. schools (Blank & de las Alas , 2009; Darling-Hammond et al., 2008, 2009; Obama, 2009). Standards-based educational improvement requires teachers to have comprehensive knowledge of their subject(s) and the pedagogy that is most effective for teaching their subject(s). States and school districts are charged with establishing and leading professional development (PD) programs, some with federal funding support, which will address major needs for improved preparation of teachers (Blank & de las Alas, 2009). The increasing emphasis on teacher quality, including teacher preparation, mentoring, PD, and evaluation of classroom instruction, is at the heart of efforts to improve the academic performance of our public schools (Darling-Hammond et al., 2008, 2009; Wei et al., 2009).

The Council of Chief State School Officers (CCSSO) (Blank & de las Alas, 2009) and the National Staff Development Council (NSDC) at Stanford University (Wei et al., 2009), among others, have led recent initiatives designed to identify, analyze, and disseminate important findings from research and evaluation studies of teacher PD. By examining information about the nature of PD opportunities currently available to teachers across the United States and in a variety of contexts, education leaders and policymakers can begin both to evaluate the needs of the systems in which teachers learn and do their work and to consider how teachers' learning opportunities can be facilitated (Blank & de las Alas, 2009; Darling-Hammond et al., 2008, 2009).

Large-scale synopses of existing research are essential to elevating student performance because several national studies on what distinguishes high-performing, high-poverty schools from their lower-performing counterparts consistently identify effective school-wide collaborative professional learning as critical to student achievement gains. And yet, as a nation we have failed to leverage this evidence to ensure that every educator and student benefits from highly effective professional learning (Wei et al., 2009). To meet federal requirements and public expectations for school and student performance, the nation needs to bolster skills and knowledge to ensure that every teacher possesses the ability to teach increasingly diverse learners, who have differing capacities and challenges (Fulton & Britton, 2011; Thoonene et al., 2011)

PURPOSE OF THIS REPORT

The primary purpose of this report is to summarize empirically-based, effective professional learning strategies and programs from the literature, with a secondary aim of providing a succinct review of evaluating effective teaching (Glazerman et al., 2011; MET Project, 2010; The New Teacher Project, 2010; Weisberg et al. 2009). A prevailing observation in the literature is that PD can enhance teachers' knowledge and teaching skills, which in turn leads to effective instruction that promotes student learning gains (Darling-Hammond et al., 2008, 2009). Effective instruction requires a comprehensive, in-depth knowledge of content taught, a thorough understanding of different learning styles, and an array of teaching skills to present complex ideas to a diverse group of learners (Darling-Hammond et al., 2008; Fulton & Britton, 2011; Marzano, 1998; Thompson & Goe, 2009; U. S. Department of Education, 2008). Complete mastery in teaching presupposes the flexibility to match instruction to learning styles and abilities of students (Darling-Hammond et al., 2008).

In an effective professional learning system, teachers and school leaders learn from outside experts, academic coaches, and highly-skilled peers. Leaders work with coaches and other

teachers to create a culture, structures, and dispositions that promote continuous incremental PD aimed at identifying individual teacher and student needs, instructional strategies to address those needs, and data-driven evaluations of teaching (Wei et al., 2009). The continual deepening of knowledge, skills, and application is an integral part of the responsibilities of teaching (Wei, Darling-Hammond, & Adamson, 2010; U. S. Department of Education, 2008).

INDICATOR OF NEED FOR PROFESSIONAL DEVELOPMENT

Although the impact of PD on teaching and student achievement has been debated, the preponderance of evidence suggests a significant positive relationship between PD and student achievement (Blank & de las Alas, 2009; Darling-Hammond et al., 2008, 2009). Research also indicates that effective PD presupposes a sequence of developmental learning, consisting of individually-tailored instruction, modeling, practice teaching and observational feedback from peers and coaches, and regular evaluation based on a variety of indicators (Baker et al., 2010; Daley & Kim, 2010).

Recent reviews of the condition of education in this country by the U. S. Department of Education (2012) reveal a serious need for rigorous PD. For example, during most of the 20th century, the United States possessed peerless mathematical prowess, measured not only by the depth and number of the mathematical specialists, but also by the scale and quality of its engineering, science, financial leadership, and even by the extent of mathematical education in the broader population (U. S. Department of Education, 2008).

However, the average U.S. mathematics literacy score (487) on the Program for International Student Assessment (PISA) in 2009 was lower than the average scores of the 34 OECD (Organization for Economic Co-operation and Development) countries (U.S. Department of Education, 2011). The National Assessment of Educational Progress (NAEP) math exam shows that in 2011 40% of 4th graders in the country scored proficient (grade-level) or above, whereas 35% of 8th graders scored at that level. In 2011, 45% of the high school graduates who took the American College Testing (ACT, Inc. 2011) met the Benchmark for college readiness in math; 25% met the Benchmark for all four exams (i.e., math, science, English, and reading).

The lack of proficiency in math led to the creation of a National Mathematics Advisory Panel in April 2006, which concluded:

- Differences in teachers account for 12% to 14% of total variability in students' mathematics achievement gains during an elementary school year.
- When teachers are ranked according to their ability to produce student achievement gains, there is a 10 percentile point difference across the course of a school year between achievement gains of students of top-quartile teachers versus bottom-quartile teachers.
- Teachers must know the mathematical content they are responsible for teaching in detail, and its connections to other important elements of math, both prior to and beyond the level they are assigned to teach.
- Teachers, especially below high school, do not know enough math to teach it. A college major, or even a concentration, in math is not required to teach math in virtually all states.

The mathematics preparation of elementary and middle school teachers must be strengthened as one means for improving teachers' effectiveness in the classroom. This includes pre-service teacher education, early career support, and professional development programs (Department of Education, 2008).

RESEARCH ON PROFESSIONAL DEVELOPMENT OPPORTUNITIES

In her preface to a report from the National Staff Development Council (NSDC), Stephanie Hirsh, Executive Director, writes:

“At a time when the nation is moving quickly to adopt common core standards and schools are challenged to accelerate gains in student achievement to meet federal mandates, states and districts need to move more aggressively to provide continuous professional development. Effective professional learning — which enables teachers to work regularly together to improve their practice and implement strategies to meet the needs of their students — must be a key ingredient in any effort to bolster student achievement and ensure that all students complete high school ready for college and careers. Without ratcheting up support for effective educator learning, the ability of teachers and school leaders to meet these new challenges will be diminished. As consensus among researchers indicates, the quality of teaching students experience is highly correlated with their academic success. Professional development is a key strategy available to schools and school systems for improving teaching quality. To ensure effective teaching in every classroom, educators must have opportunities each day to refine and expand their practice, reflect on how their practice impacts student learning, and engage in ongoing improvement to address learning challenges in the school. States and school systems have the authority and responsibility to establish policies to guide effective professional learning and to monitor its implementation and impact. Unfortunately, implementation as well as the impact on students is inconsistent state to state.” (Wei et al., 2010, p. 2).

The NSDC (Wei et al., 2009) found that opportunities for sustained, collegial PD of the kind that produces changes in teaching practice and student outcomes were much more limited in the United States than in most high-achieving nations abroad. In a follow-up report (Wei et al., 2010), the NSDC used several datasets to analyze the progress of PD efforts in the states over the last decade. They found that more teachers had access to PD directly related to the content they teach and they spent more time on these subjects than did teachers a decade earlier. However, there had been a decline in the intensity of PD in all other areas of professional learning. Teachers nationwide had fewer opportunities to engage in sustained professional learning opportunities (i.e., 8 hours or more) than they had earlier. In areas like reading instruction, uses of computers, teaching of English language learners and special education students, U.S. investments in teacher learning appear to be increasingly focused on the least effective models of professional development—the short-term workshops that research suggests are unlikely to influence practice and student achievement (Wei et al., 2010).

Of special concern, well under half of teachers reported access to PD on teaching students with disabilities (42%) and teaching ELL students (27%). In addition, only a third of teachers agreed that their schools provide support for teaching students with special needs, a lower proportion than in 2004. Examining several sources of data, including the MET data (MET Project, 2010) and the Schools and Staffing Survey (U. S. Department of Education, 2011), the NSDC finds that about three-quarters of teachers indicate that they have some formal opportunities for collaborative planning with other teachers. However, the opportunities are limited, averaging only about 2.7 hours a week, and rarely translate into a school climate that is perceived to

enable significant cooperative effort. Teachers were only half as likely to report collaborative efforts in their schools (16%) as reported a decade earlier (34%).

Data from the first Teaching and Learning International Survey indicate that American teachers—compared with teachers in high-achieving Organization for Economic and Co-operative Development (OECD) nations — have much less time in their regular work schedules for cooperative work with colleagues. “These findings suggest that while an increasing number of American teachers have opportunities for collaborative work in schools, the current structures (e.g., work schedules) rarely allow for deep engagement in joint efforts to improve instruction and learning.” (Wei et al., 2010, p. 6).

Contrary to national trends, there are some states that, over the past decade, are providing more intensive, sustained learning experiences. The NSDC report (Wei et al., 2010) rates state PD efforts on 11 separate indicators, noting that some states have clearly made significant strides in offering induction opportunities for beginning teachers (e.g., Colorado, Delaware, Iowa, Pennsylvania, South Carolina) and professional learning opportunities for veteran teachers (e.g., Arkansas, Colorado, Oregon, Utah).

RESEARCH ON EFFECTIVE PROFESSIONAL DEVELOPMENT

KEY PRINCIPLES IN DESIGNING PROFESSIONAL DEVELOPMENT

In their first report, the NSDC (Wei et al., 2009) provided a narrative review of the research on effective PD as the basis for evaluating the current status of professional learning in the United States. These researchers defined “effective” as PD that results in improvements in teachers’ content knowledge and teaching skills, as well as increases in student achievement.

Their review contains studies that use a range of research methodologies, including experimental and quasi-experimental designs, surveys, and case studies. In the latter case, they note that the inferences that can be drawn from such research should be treated as suggestive rather than inferential or conclusive. The research on effective PD has begun to create a consensus about key principles in the design of learning experiences that can impact teachers’ knowledge and practices (Blank & de las Alas, 2009; Darling-Hammond et al., 2008, 2009; Wei et al., 2009, 2010). For example, existing research clearly affirms that single-session, fragmented workshops have little if any positive impact on teaching or student achievement, whereas individually-tailored, developmental PD plans, consisting of modeling, practice teaching, and classroom feedback from peers and academic coaches, are effective in enhancing instruction and student performance (Baker et al., 2010; Darling-Hammond et al., 2008, 2009; Wei et al., 2010). Evidence indicates that the following features of PD are related to knowledge accumulation, enhanced teaching skills, and increased student achievement:

- PD focused on specific curriculum content and pedagogies needed to teach that content effectively.
- PD offered as a coherent part of a whole school reform effort, with assessments, standards, and professional development seamlessly linked.
- PD designed to engage teachers in active learning that allows them to make sense of what they learn in meaningful ways.
- PD presented in an intensive, sustained, and continuous manner over time.
- PD linked to analysis of teaching and student learning, including the use of formative assessments.
- PD supported by coaching, modeling, classroom observation, and feedback.
- PD that is connected to teachers’ collaborative work in school-based professional learning communities and learning teams.

Succinctly stated, there is an emerging consensus about both the content of and contexts for professional learning, as well as approaches to designing learning experiences (Garet et al., 2001; Wei et al., 2009).

CONTENT OF PROFESSIONAL DEVELOPMENT

The NSDC (Wei et al., 2009) writes,

“The content of the professional development is most useful when it focuses on concrete tasks of teaching, assessment, observation and reflection” rather than abstract discussions of teaching. Studies find strong effects of professional development on practice when it focuses on enhancing teachers’ knowledge of how to engage in specific pedagogical skills and how to teach specific kinds of content to learners. Equally important is a focus on student learning, including analysis of the conceptual understanding and skills that students will be expected to demonstrate.” (p. 3)

As one example, Saxe, Gearheart, and Nasir (2001) compared three types of support for teacher learning, and found that student achievement improved most when teachers were engaged in sustained, collaborative PD that specifically focused on deepening teachers’ content knowledge and instructional practices. The three teacher learning options included: traditional single-session PD workshops, a professional community-based activity which offered support to teachers using new curriculum units, and the Integrated Mathematics Assessment (IMA) approach, which directly engaged teachers in learning the mathematics in the new curriculum as well as facilitating discussion around pedagogical content knowledge necessary to teach the units. The researchers found that students whose teachers had participated in the IMA program showed the greatest gains in conceptual understanding. The study’s findings underscore the need for learning opportunities that focus on specific content knowledge and content pedagogy and “point to the problems with reform curriculum when such curriculum are not accompanied by focused supports for teachers’ subject matter knowledge, knowledge of children’s mathematics and implementation of reform-oriented pedagogical practices” (Saxe et al., 2001, p. 70).

COLLABORATIVE PROFESSIONAL DEVELOPMENT

Research also finds that PD is more effective when it is not approached in isolation (e.g., single-session workshops) but as a coherent part of the school’s educational reform or approach (Wei et al., 2009). For substantial change to occur, curriculum, assessment, standards, performance evaluations (teachers and leaders), and professional learning should be seamlessly linked in order to avoid disconnections (confusion) between what teachers learn in professional development and what they implement in their classrooms and schools.

Furthermore, direct training of principals and teacher leaders, distributed leadership, and explicit protocols for leading grade-level learning teams are critical for effective PD, or positive changes in teaching and student achievement. Wei et al. (2010) observe that good teaching occurs when educators collaboratively articulate student and teacher learning goals based on assessments of progress, and together with academic coaches create lesson plans that are designed to meet the needs of their particular students. The NSDC (Wei et al., 2010) concludes that professional learning should tap the expertise of educators in the school and at the district office, with support from universities and other external experts who can help local educators address needs specific to their students and school improvement goals.

JOB-EMBEDDED PROFESSIONAL DEVELOPMENT

In tandem with collaborative PD in professional learning communities (Dufour & Marzano, 2011), there is growing interest in job-embedded PD, such as mentoring, coaching, collaborative teaching, and study groups. Some of the many advantages and benefits of job-embedded PD include, but are not limited to, learning occurring at least in part during the process of classroom instruction or during regularly scheduled teacher planning periods. By locating opportunities for PD within a teacher's regular work day, job-embedded learning is more likely to be ongoing and cumulative (Blank & de las Alas, 2009; Garet et al., 2001).

Job-embedded PD typically involves modeling, practice teaching and classroom observation and feedback from peers and academic coaches, individually tailored instruction, and sequentially developed PD plans that span more than one year. The opportunity to observe skilled peers and coaches, practice teaching skills, and receive immediate feedback within the scope of a long-range PD plan leads to more effective teaching and student achievement than fragmented, single-session workshops that provide no follow-up or continuity in learning (Darling-Hammond et al., 2008, 2009). Job-embedded PD provides assistance to teachers immediately, when they need it most.

Opportunities for active learning or “application experiences” are critical to augmenting abilities to teach (Snow-Renner & Lauer, 2005). These application experiences typically entail teachers practicing skills immediately following modeling of the same skills by peers and coaches, and receiving feedback on their performance. These feedback consultations provide teachers with the opportunity to discuss subtle nuances in content and skills that, added together, can transform individual teaching and contribute to altering the culture of instruction in a school (Garet et al., 2008; Wei et al., 2009). Loosely organized, single-session workshops, required for all teachers, irrespective of individual needs, cannot provide the specific details and practice experiences needed to enhance teaching knowledge and skill (Garet et al., 2008; Wei et al., 2009).

PROFESSIONAL LEARNING COMMUNITIES

As studies reveal more about how teachers learn, many researchers and practitioners have begun to place greater emphasis on collaborative learning in professional learning communities (PLCs) (Darling-Hammond & McLaughlin, 1995; Dufour & Marzano, 2011; Dufour et al., 2006). The literature increasingly describes how teachers learn by working with their colleagues in PLCs, engaging in continuous dialog and examination of their practice and student performance to develop and enact more effective instructional practices. In ongoing opportunities for collegial collaboration, teachers have an opportunity to learn about, try out, and reflect upon new practices in their specific setting, sharing individual knowledge and expertise with one another. To characterize what she observed occurring in productive teacher learning communities, Little (1990, 2003) developed a construct she termed “joint work,” which requires norms of mutual aid over privacy and “thoughtful, explicit examination of practices and their consequences” (p.520).

Joint work can be found in shared planning activities and collaboration on curriculum, when teachers work in grade-level teams that share students or content goals, and when teachers observe and critique each other's instruction based on a shared understanding of effective teaching and goals for student learning. Interdependence between teachers is cultivated through these activities. Collaborative or joint work promotes mutual problem-solving, and the creation of a shared technical language and agreement on sound practices.

A number of large-scale studies have demonstrated that collaborative, job-embedded, professional learning that is focused on student performance has resulted in changed practices and improved student achievement (Wei et al., 2009, p.11). At the same time, few existing studies have been designed in such a way as to allow for causal inferences about the impact of particular interventions on student learning. However, a longitudinal quasi-experimental study of the impact on student achievement of grade-level teams provides empirical evidence of the effectiveness of professional learning communities for increasing student achievement (Saunders, Goldenburg, & Gallimore, 2009). This study found that in the nine Title I schools in which a grade-level teaming strategy was implemented, students outperformed their peers in six matched schools in the same large, urban district on standardized achievement tests, with effect sizes improving over time to an effect size of 0.88 in the fifth year.

PROFESSIONAL DEVELOPMENT'S EFFECT ON STUDENT ACHIEVEMENT

Professional development has emerged as a major forum (or vehicle) for teaching content and instructional skills needed to effectively enhance student achievement (Baker et al., 2010; Blank & de las Alas, 2009; Newton et al., 2010). In one of the most comprehensive reviews of studies on the effectiveness of teachers' in-service PD, Yoon et al. (2007) analyzed the findings from over 1,300 studies and evaluation reports, and identified only nine experimental or quasi-experimental studies using control groups with pre- and post-test designs that could evaluate teacher PD impacts on student achievement. Their review of these nine studies concluded that sustained and intensive PD was related to student achievement gains. Specifically, five of six studies that offered substantial contact hours of PD (ranging from 30 to 100 hours in total) spread out over six to 12 months showed a positive significant effect on student achievement gains. The remaining three studies that involved limited amount of PD (ranging from 5 to 14 hours in total) showed no statistically significant effect on student learning. Across the nine studies, an average of 49 hours in a year increased student achievement by approximately 21 percentile points.

In 2006, the Council of Chief State School Officers (CCSSO) (Blank & de las Alas, 2009) was awarded a grant by the National Science Foundation to conduct a meta-analysis of the effects of teacher PD on student achievement. The two-year meta-analysis was designed to assess and summarize consistent, systematic findings across multiple studies that show significant effects of teacher PD on student achievement gains in K-12 mathematics or science. Meta-analyses examine "effect" sizes of rigorous studies to get an indication of the impact of an intervention on an outcome, in this case the impact of PD on student achievement. An "effect size" is the difference in a criterion measure (e.g., student achievement) between an experimental (or intervention) and a control group. The 16 studies that met the rigorous criteria for selection in the meta-analysis generated a total of 104 effect sizes in the CCSSO study (Blank & de las Alas, 2009). This meta-analysis found that studies using student measures that are closer to the heart of what the PD is intended to impact do report larger effect sizes, or more impact of PD on student achievement.

The CCSSO meta-analysis also provides strong evidence in support of active methods of teacher learning during PD, such as coaching, modeling, classroom observations and feedback, discussions with colleagues, developing assessments, and professional networks (Blank & de las Alas, 2009). Another key finding concerns the nature of teacher learning goals in the PD design: Each of the PD programs analyzed focused on helping teachers improve their knowledge of how students learn in the specific subject area, how to teach the subject with effective strategies. The PD programs were deliberately designed to transmit specific skills tailored to content areas that could be immediately used in the classroom by teachers.

More generally, studies find that teacher and student performance is enhanced when teachers are more knowledgeable about all students and can coordinate efforts to meet students' needs through collaborative efforts (Baker et al., 2010; Blank & de las Alas, 2009). Collaboration among teachers with different levels and areas of skill and different types of experience can capitalize on the strengths of some, compensate for the weaknesses of others, increase shared knowledge and skill, and thus increase their school's overall professional capacity.

Optimal effectiveness is achieved by individually tailored sequential PD plans that span a period of years. Each stage in the developmental sequence is planned conjointly by teachers and overseers based on classroom observations and other artifacts, such as portfolios, student assignments and assessments, lesson plans, collegial duties, and student performance on exams.

EVALUATING THE EFFECTIVENESS OF TEACHERS

The primary goal of PD is to enhance teachers' effectiveness in facilitating student achievement (Koretz, 2008; Newton et al., 2010; Glazer et al., 2011; Wei et al., 2010). Recently, more systematic approaches to evaluating teacher effectiveness have emerged in the professional literature and state policy-making in response to emphases on evaluation and accountability in the No Child Left Behind (NCLB) Act (2002) and Race to the Top (RTT) funding criteria (U. S. Department of Education, 2010), as well as to evidence on existing evaluation practices (Daley & Kim, 2010; Weisberg et al., 2009). Evidence has indicated that, until very recently, teacher evaluations nationally had been characterized by infrequency, lack of explicit expectations and rating criteria, limited useful feedback, and no continuity across years (Daley & Kim, 2010; Weisberg et al., 2009). A recent survey of 15,176 teachers in 12 districts found that nearly 75% of teachers had not received specific feedback on how to improve their instructional practice. Newly inducted teachers also reported they had not received feedback on any area of performance in need of improvement over the course of their first three years as teachers. Known as the "Widget Effect," this same widely-cited study found that districts rarely undertook formal dismissal procedures for poorly performing teachers; in half of the districts studied, not a single non-probationary teacher was dismissed on the grounds of poor performance within a five-year period (Weisberg et al., 2009).

In their 2nd phase application for a Race to the Top (RTT) grant, the Arkansas Department of Education wrote,

"...we took a hard look at our own performance to promote teacher quality and found ourselves lacking. Our technical assistance to infuse effective teacher evaluation had been lax far too long.... data from one of the studied districts showed that none of the probationary teachers were non-renewed for performance in five years, and 99.7% of tenured teachers received a satisfactory rating or equivalent. Similarly, data from another district showed that 100% of tenured teachers received a satisfactory rating or equivalent. The data on the Arkansas LEAs that participated in this research project and the resulting report are representative of most districts in Arkansas." (U. S. Department of Education, 2010, p. 96)

Similar statements are found in other states' RTT grant applications (U. S. Department of Education, 2010). Historically, teacher evaluations have been characterized by perfunctory bureaucratic exercises guided by few if any expectations or criteria or concern for continuity from one year to another. These evaluations were subjective professional judgments based on limited periodic classroom observations. Succinctly stated, evaluators rarely provided teachers with specific useful feedback for professional development and changes in teaching strategies

aimed at improved student performance (Dwyer, 2007). The intensity and frequency of evaluation vary according to teachers' tenure status; for example, tenured teachers in approximately half of the Midwestern districts studied by Brandt et al. (2007) were evaluated once every three years.

The more rigorous approaches to teacher evaluation entail efforts to devise systematic tools and procedures for measuring effective teaching, involving frequent classroom observations based on detailed protocols and assembling artifacts that represent student performance and teacher responsibilities (Daley & Kim, 2010).

METHODS OF TEACHER EVALUATION

By far the most common method of teacher evaluation involves observing a teacher's instruction, without considering student achievement scores on standardized tests (Brandt et al., 2007; Donaldson, 2009). Presently, classroom observations are generally structured around an observation instrument that measures teachers' performance on an assessment metric, often a dichotomous satisfactory/unsatisfactory rating scale. However, there is an increasing use of more fine-grained scales such as unsatisfactory, basic, proficient, and distinguished. "Walk-throughs" are increasingly popular, and some states provide training and instructional materials to guide the implementation of these classroom observations. Formative assessments aimed at coaching and supporting teachers, instead of rating them, are being adopted by many districts.

Many districts and states also have moved toward performance-based assessments as a means for teacher evaluation. This method involves observation, but assesses the teacher's instruction against an articulated set of performance standards translated into detailed descriptors within a rubric. Charlotte Danielson's (2011) *Framework for Teaching*, aligned with the Interstate New Teacher Assessment and Support Consortium standards, is one of the most commonly used system of evaluation. The Arkansas Department of Education has adopted this framework for their new teacher evaluation and support system (Act 1209, Cushman, 2011). Some educators argue that performance-based assessments deliberately include teaching behaviors that promote student learning. Others qualify this statement by observing that performance-based assessment method is focused more on teacher inputs than student outputs (Donaldson, 2009).

ARKANSAS TEACHER EVALUATION AND SUPPORT SYSTEM

The legislative intent of Act 1209 of the Arkansas 88th General Assembly, 2011, is to create a teacher evaluation system that ensures effective teaching, promotes professional learning, and provides feedback and support that encourage teachers to develop knowledge and skills that contribute to student achievement gains. The intent also is to link evaluation procedures with curriculum standards, professional development and support, and employment decisions. A major strength of this evaluation system is the effort to focus on both teacher inputs and student outcomes by using a balance of "artifacts" or indicators (e.g., lesson plans, PD participation, samples of student work, formative and summative assessments) of student and teacher performance. The artifacts used in a particular case in each district is a conjoint decision made by the evaluator(s) and the teacher being evaluated.

Another major strength of the new Arkansas teacher evaluation is the use of standardized (The Danielson Group, 2011) "frameworks," or teacher evaluation categories (e.g., setting instructional outcomes), and "rubrics" or descriptors of these categories (e.g., students link learning goals), with levels of performance (i.e., unsatisfactory, basic, proficient, and distinguished). In the summative (or rating) teacher evaluation, one-half of the artifacts must be external assessment measures (i.e., scored by someone other than the teacher evaluated). A

teacher is placed in intensive support status, for a time specified by the evaluator, if that person receives a rating of “unsatisfactory” in any evaluation category, or a rating of “unsatisfactory” or “basic” in a majority of descriptors in a category. If a teacher does not accomplish the goals and complete the tasks established for the intensive support status, a recommendation for termination is made based on the Teacher Fair Dismissal Act of 1983, § 6-17-1501 et seq.

The Arkansas teacher evaluation system also links PD to individual teachers’ learning needs identified through the evaluation process. In concert with the teacher, evaluators are required to develop PD plans in which at least half of the 60 hours required by law are directly related to the teacher’s area of teaching and identified teaching needs. For teachers in intensive support status, all PD hours, except those required by law, must be directly related to the individual teacher’s needs.

Two other features of the Arkansas evaluation system are particularly noteworthy. ADE has devoted considerable resources to training evaluators over a period of time, instead of assuming that a new comprehensive system can be learned in a single session. Secondly, there are online video vignettes to model effective teaching in each domain, which is especially useful for teachers who are struggling with instruction.

In summary, some of the salient advantages of the Arkansas Teacher Evaluation and Support System include the use of a well-established framework for teaching (The Danielson Group, 2011), with empirically-grounded rubrics and performance ratings. This system also links teacher evaluation to PD, intensive support, and employment decisions. Furthermore, recommended artifacts, half of which must be external measures, cover a wide range of teacher responsibilities in addition to classroom instruction, such as PD participation and collaborative research. This diversity of performance data comports well with the criticism directed at using “high stakes” testing as the sole measure of teacher effectiveness (Baker et al., 2010; Koretz, 2008). It also avoids the problems raised in the next section of this report on using portfolios as the sole or primary measure of teacher effectiveness.

PORTFOLIO AS A MEASURE OF TEACHER EFFECTIVENESS

Some school districts rely extensively or exclusively on the use of a portfolio to evaluate teachers, and many portfolios include a performance-based element, and other districts also require evidence of student learning (Tucker et al., 2003). The National Board for Professional Teaching Standards portfolio, for example, includes video of a candidate teaching, samples of student work, and evidence of impact beyond the classroom in working with colleagues, community members, or families. Additionally, National Board candidates must pass content- and grade-specific examinations. Candidates’ entries are scored by a minimum of 12 teachers who have been trained in how to evaluate these portfolios. Connecticut has used a content-specific portfolio system coupled with mentoring in its licensure decisions since 2000 in its Beginning Educator Support and Training Program.

Many districts that rely on portfolios allow teachers to “cherry pick” items included in portfolios. As a result, they tend to not be a representative sampling of teacher and student performance. In addition, many items included do not lend themselves to inter-rater reliability, such as essays and poster-board displays. Scoring or rating items in portfolios is based on professional judgment, and therefore, subject to idiosyncratic biases and preferences. Ratings and scoring of portfolio items are rarely subjected to inter-rater reliability checks, nor is the validity questioned. Research on use of portfolios as teacher evaluation tools is discussed on p. 13 of this report under the section on Comparison of Methods of Teacher Evaluation.

VALUE-ADDED MODELING MEASURE OF TEACHER EFFECTIVENESS

The most quantitative approach to teacher evaluation is “Value-Added Modeling” (VAM) (Koretz, 2008; Newton et al., 2010; Glazerman et al., 2011), which focuses directly on growth in student achievement scores. It has been argued that VAM is a more sophisticated analysis of fine gain scores than commonly used examinations of test scores at a single point in time, or comparisons across time of cohorts consisting of different students (Koretz, 2008; Newton et al., 2010; Glazerman et al., 2011).

Conceptually, VAM's promise of quantifying the “added value” that teachers and/or schools produce in terms of student learning offers intuitive appeal because it is more objective than professional judgments of classroom teaching and “cherry picked” items in a portfolio. Methodologically, VAM is attractive because it appears to offer a way to disentangle the “effects” of teachers and/or schools from those of other uncontrolled (or extraneous) factors, including students' demographic and socioeconomic characteristics, family education and language background, and community characteristics such as poverty and crime.

However, despite its conceptual and methodological appeal, the use of VAM to estimate teacher effectiveness or to rank teachers for high stakes purposes poses daunting and seemingly insurmountable challenges stemming from many factors, including but not limited to: 1) the non-random assignment of students to teachers and schools, 2) the limitations of particular tests both for measuring the full range of desired knowledge and skills and for measuring learning gains, and 3) the difficulties of disentangling the contributions of many influences on learning, such as multiple teachers, parents, tutors, specific curricula, availability of useful learning materials, class sizes, leadership, and a host of student and community characteristics (Baker et al. 2010; Corcoran, 2010; Glazerman et al., 2010; Newton et al., 2010).

Since neither students nor teachers are randomly assigned to classes, it is very likely that value-added estimates contain systematic bias owing to class assignments. This assignment bias is confounded by the well-established effects of student (e.g., disabilities, family dysfunction) and community (e.g., poverty, crime) characteristics on student performance (Baker et al., 2010; Corcoran, 2010; Wei et al., 2010). The complexities of these interactive effects have not been fully analyzed or well understood, and therefore, raise serious questions about the adequacy of VAM procedures to account for all of the interrelationships involved (Baker et al. 2010; Corcoran, 2010; Glazerman et al., 2010; Koretz, 2008; Newton et al., 2010).

To further complicate this problem of measuring teacher effectiveness, it is exceedingly difficult, if possible, to isolate the “effects” of individual teachers from the overall school influences, and the impact of other teachers and tutors on students (Baker et al., 2010; Darling-Hammond, L. et al. Darling-Hammond, L. et al., 2008, 2009; Glazerman et al., 2010; Wei et al., 2010). Individual student learning, even in a particular subject, is actually a reflection of the composite of instruction received from several teachers and tutors, including parental assistance (Baker et al., 2010; Blank & de las Alas, 2009; Corcoran, 2010; Koretz, 2008).

Based on an extensive review of VAM research, scholars convened by the Economic Policy Institute (Baker et al., 2010, p. 9) concluded:

“Although value-added methods can support stronger inferences about the influences of schools and programs on student growth than less sophisticated approaches, the research reports cited...have consistently cautioned that the contributions of VAM are not sufficient to support high-stakes inferences about individual teachers. Despite the hopes of many, even the most highly developed value-added models fall short of their goal of adequately adjusting for the backgrounds of students and the context of teachers' classrooms.”

Researchers at the prestigious Rand Corporation made similar comments (McCaffrey et al., 2003):

“The estimates from VAM modeling of achievement will often be too imprecise to support some of the desired inferences. The research base is currently insufficient to support the use of VAM for high-stakes decisions about individual teachers or schools.”

In addition to statistical procedure issues, studies indicate a lack of stability in teacher ratings of effectiveness across years, with correlations as low as .20 to .35 (McCaffrey et al., 2009). Some of this instability in ratings is due to variance in teachers' performance and extraneous factors already discussed, but it is also the result of measurement deficiencies (Baker et al., 2010). Value-added measurement of growth from one grade to the next should use vertically scaled tests, which most states do not use (Baker et al., 2010). In order to be vertically scaled, tests must evaluate content that is measured along a continuum from year to year. Following an NCLB mandate, most states now use tests that measure grade-level standards only and, at the high school level, end-of-course examinations, neither of which are designed to measure such a continuum. Without vertically scaled tests, VAM can estimate changes in the relative distribution, or ranking, of students from last year to this, but cannot do so across the full breadth of curriculum content in a particular course or grade level, because many topics are not covered in consecutive years.

Furthermore, it is important to recognize that the standardized tests are not perfect, unerring measurements of student achievement. Not only are they subject to errors of various kinds, they are narrow measures of what students know and can do, relying largely on multiple-choice items that do not evaluate students' communication skills, depth of knowledge and understanding, or critical thinking and performance abilities (Baker et al., 2010). The present widespread practice of giving students intense preparation for state tests—often to the neglect of knowledge and skills that are important aspects of the curriculum but beyond what tests cover—has in many cases invalidated the tests as accurate measures of the broader domain of knowledge that the tests are supposed to measure. This phenomenon is reflected in the continuing need for remedial courses in universities for high school graduates who scored well on standardized tests.

A report from the National Research Council and the National Academy of Education (Braun, Chudowsky, & Koenig, 2010, p. 1) concluded:

“Value-added methods involve complex statistical models applied to test data of varying quality. Accordingly, there are many technical challenges to ascertaining the degree to which the output of these models provides the desired estimates. Despite a substantial amount of research over the last decade and a half, overcoming these challenges has proven to be very difficult, and many questions remain unanswered...”

In summary, basing teacher evaluation primarily on student test scores does not accurately distinguish “more” from “less” effective teachers because even relatively sophisticated approaches cannot adequately address the full range of statistical problems that arise in estimating a teacher's effectiveness. Efforts to address one statistical problem often introduce new ones. These challenges arise because of the influence of student socioeconomic advantage or disadvantage on learning, measurement error and instability, the nonrandom sorting of teachers across schools and of students to teachers in classrooms within schools, and the difficulty of disentangling the contributions of multiple teachers over time to students' learning. As a result, reliance on student test scores for evaluating teachers is likely to misidentify many teachers as either poor or successful.

COMPARISON OF METHODS OF EVALUATING TEACHER EFFECTIVENESS

Well-respected scholars at the Brown Center on Educational Policy at Brookings (Glazerman et al., 2010) argue that we know a good deal about how other means of classification of teachers perform in comparison to value-added approaches. They maintain that research is quite clear that if student achievement is the outcome, value-added is superior to other existing methods of classifying teachers. "Classification that relies on other measurable characteristics of teachers (e.g., scores on licensing tests, routes into teaching, nature of certification, National Board certification, teaching experience, quality of undergraduate institution, relevance of undergraduate coursework, extent and nature of professional development), considered singly or in aggregate, is not in the same league in terms of predicting future performance as evaluation based on value-added." (Glazerman et al., 2010, p. 9)

Glazerman et al. (2010) discuss a study done in New York City, where they noted that if teachers were laid off based on seniority they would be distributed across the full range of teachers' effectiveness in raising student test scores, whereas teachers laid off based on low value-added scores would be at the bottom of the distribution. In other words, many more effective teachers would be retained if layoffs were based on value-added than if they were based on seniority. Principal ratings perform better than teacher seniority in identifying teachers with low effectiveness in raising student achievement, but not nearly as well as value-added scores.

Other research on various approaches to teacher evaluation suggests that they all provide some information on student learning. Jacob and Lefgren (2008) found that principals' ratings of teachers' ability to raise student achievement were correlated with the teachers' actual value-added student gains. The researchers found that principals successfully identified teachers who produced the highest and lowest value-added student gains, but they were not as good at identifying those in the middle. Milanowski (2004) found that performance-based ratings had modest to moderate correlations with value-added estimates in math (.43), reading (.32), and science (.27) in Cincinnati. Based on these results, he concluded that rigorous, performance-based assessment results do have criterion validity and could be used to make high-stakes judgments about teacher job security. Similarly, Gallagher (2004) observed a relationship between evaluation scores and student achievement in reading (.50). He further noted that the school's curriculum, professional development, and evaluation instrument were more aligned with each other and with the student tests in reading than they were in math.

Much of the research on portfolios has investigated the relationship between National Board certification and student achievement, and this evidence is mixed. The certification process does appear to select the most effective teachers from the applicant pool. Moreover, Goldhaber and Anthony (2007) found that students taught by National Board certified teachers in North Carolina elementary schools experienced significantly larger gains in math and reading than students of other teachers. However, a study of National Board certified teachers in Florida between 2000 and 2004 in grades 3-10 found that future National Board certified teachers were not more effective than non-Board certified teachers as measured by most value-added estimates (Harris & Sass, 2007). These researchers also found that the process of becoming National Board certified did not raise teachers' effectiveness either during or after certification. A major problem with studies of National Board certified teachers is that they are a self-selected group that is unlikely to be representative of the larger teacher population (Little, Goe, & Bell, 2009). A study on Connecticut's BEST program found that students of teachers who scored highest on the portfolio assessment gained three additional months of learning over one year in comparison to student taught by the lowest scoring teachers (Wilson et al., 2007).

Value-added modeling putatively offers the most direct evidence on the effect of teaching on student achievement. However, in addition to the limitations already discussed, evidence indicates that value-added estimates are likely to vary from one year to another. For instance, Goldhaber and Hansen (2008) found moderate correlations (.30 in reading, .52 in math) in the year-to-year teacher effects in their study of fifth-grade teachers in North Carolina. Specifically, these researchers found that of teachers who were in the bottom quintile on average in math or reading in their first 2-3 years of teaching in North Carolina, 29 percent were in the top two quintiles in reading and 24 percent were in the top two quintiles in math in years five and up, on average. Papay (2011) also found low-to-moderate correlations (.17 to .51) among value-added estimates based on three reading achievement tests administered in one district to the same teachers during a year.

DISCUSSION OF STRENGTHS AND WEAKNESSES OF EACH EVALUATION APPROACH

Based on an extensive search of the professional literature, and conversations with national experts and ADE officials, a discussion ensues concerning primary strengths and weaknesses of current teacher evaluation approaches. In the interest of parsimony, this discussion does not address other less commonly used approaches, such as teacher self-evaluations and parental and student evaluations.

Presently, it seems judicious to conclude that each of the teacher evaluation methods discussed, while suffering important limitations, offers useful information for evaluation, professional development, and strategies for improving student achievement. Therefore, the wisest approach to evaluating teachers may be to blend different approaches, as done in Arkansas, to get a more robust and multi-faceted understanding of a teachers' influence on student learning and achievement.

Each approach appears to offer advantages over other methods of teacher evaluation, as well as weaknesses that make it less desirable than alternative procedures. For example, the traditional approach involving classroom observations by the principal and/or other evaluators has the advantages of direct observation of teaching and interactions with and between students. Supplemented with lesson plans, pacing guides, curriculum frameworks and standards, knowledge of individual students, and school and community characteristics, evaluations based on classroom observations offer the promise, if not the reality, of a broader, more encompassing scope of influences on teaching. These evaluations allow for individualized differences in how material is taught, and nuances in skills that go undetected in portfolios and formalized measures of teacher knowledge and skills.

However, teacher evaluations based on classroom observations rely on professional judgments, which are susceptible to subjective inferences and personal preferences. Traditionally, evaluations based on observations (and other associated instructional materials) have not been guided by explicit systematic standards and expectations, and they have not been conceptualized as developmental plans of knowledge and skills acquisition over a period of time.

Historically, classroom observations have been infrequent, involving brief periods of time, and feedback tended to be generic and lacking in detail and specificity to be useful for instructional change. Traditional classroom evaluations typically are not based on calibrated and validated observation protocols or explicit definitions of effective teaching. As a result, there is no assurance of intra- or inter-rater reliability. In most cases, evaluators have not been trained and their observations have not been calibrated (Little et al., 2009).

Classroom observations are often supplemented with portfolios consisting of lesson plans, pacing guides, research, student work and exams, formative assessments of students, formative and summative teaching evaluations, peer evaluations, notes from parents, and special awards and recognitions. Portfolios cover a broad range of teacher responsibilities and allow the teacher to present diverse indicators of performance in addition to classroom instruction, including student achievement measures. The portfolio process often requires teachers to reflect on why artifacts were included and how they relate to particular standards. They may contain exemplary work as well as evidence that the teacher is able to reflect on lessons, identify problems, and make modifications for future lessons. A salient weakness of this approach is that teachers typically are permitted to “cherry pick” items to be included in the portfolio. Ratings of portfolios are based on professional judgment, and often are not guided by systematic standards or criteria. Generally, there is no uniformity in what items are to be placed in a portfolio. Use of portfolios for summative evaluations have not been validated (Little et al., 2009). It is difficult to verify consistency in scoring or rating, and it is unclear whether materials included in a portfolio are an accurate representation of a teacher’s practice. As with classroom observations, portfolios often are not evaluated with an explicit rating scale or system. Many times a global assessment is made of satisfactory or unsatisfactory, which doesn’t provide specific guidance to teachers regarding their effectiveness.

Currently, the most systematic, objective approach to teacher evaluation is VAM. It is the only systematic statistical analysis of linkage between teaching and student achievement gains. As a whole, VAM procedures are very sophisticated statistical procedures which can examine complex interrelationships between factors from several domains, such as classroom performance, faculty relationships, school culture, and familial and community characteristics (Braun, Chudowsky, & Koenig, 2010; Corcoran, 2010; Goldhaber & Hansen, 2008; McCaffrey et al., 2003; Newton et al., 2010).

In practice, however, results indicate that VAM is not able to fully isolate the effects of individual teaching from the impact of other teachers, school culture, and wide-ranging parental and community influences on student achievement gains. This analytic inadequacy is due in large part to unavailability and deficiency of measures, including indicators of teacher effectiveness, but also to the utter complexities of trying to sort out the individual effects of dynamic interacting influences, such as the cumulative impact of multiple teachers.

Finally, extensive discussion with Kaneal Alexander, Director of Teacher Evaluation, Tennessee Department of Education, and Dr. Tony Davis, Senior Director, Mid-continent Research for Education and Learning, leads to the conclusion that pursuing a value-added modeling approach to teacher evaluation is a daunting prospect at best. Linkage between state, district, and course standards has to be established and aligned. Teaching frameworks have to be aligned with evaluation measures. Measures of teacher effectiveness and student outcomes have to be developed and aligned - especially for teachers who do not teach courses covered on state benchmark exams - and validated. Herculean efforts are required to train state and district staff on new evaluation instruments, procedures, and policies.

In addition, longitudinal datasets and data processing systems that are interconnected and can communicate across systems must be developed. According to Ms. Alexander and Dr. Davis, some states at the forefront of linking teacher evaluation system to student performance are Colorado, Tennessee, and North Carolina. No state has achieved a fully operational system of teacher evaluation linked to student achievement, and both Dr. Davis and Ms. Alexander indicated it is a Herculean undertaking to mount a complete system.

CONCLUSIONS AND POLICY IMPLICATIONS

In many states, teacher effectiveness is determined based on results from a single measure, typically classroom observation, or less often value-added modeling (Little et al., 2009; Wei et al., 2010). However, using one or both of these measures cannot account for the many significant ways teachers contribute to the success and well-being of their students, classroom, and schools. Providing a comprehensive score and/or detailed commentary for teachers based on multiple measures is necessary to capture important information that is not included in most classroom protocols or value-added scores. A comprehensive evaluation of the full scope of teaching responsibilities is essential to designing detailed, individually-tailored professional learning plans for sequential development of knowledge and skills over a span of time.

Scoring schemes provide at least the appearance of empirically-based evaluation, but unless the fidelity of administration and validity of measurement is known, the validity and reliability of scoring classroom observation protocols and portfolios is uncertain and subject to personal biases and preferences. Scoring teachers also can have negative consequences, especially for those with low scores. At the same time, scores or ratings do provide an ordinal scale for teachers to have a sense of their relative performance level. Ratings indicate what knowledge and skills need to be learned or enhanced in professional learning experiences and instruction.

Written commentary on teacher evaluations has the potential to provide nuanced details and specific guidance about content and skills that need to be enhanced in professional learning plans. Written comments can elaborate on exactly what strengths and weaknesses were observed in instruction and other teacher responsibilities, as well as actions and interventions to develop skills and content knowledge. Written comments are subject to unspecified frameworks (e.g., philosophical) and preferences, and often inter-rater reliability is unknown. A major problem historically, continuing into the present, is the lack of a protocol and training for evaluators. As a result, teacher evaluations have been replete with generic and uninformative statements that provided little if any guidance for making changes and professional development.

In reaction to these vacuous evaluations, states like Arkansas have adopted evaluation frameworks, or standardized evaluation categories with specific performance descriptors or rubrics (Act 1209, 2011; Danielson, 2011). To provide a measure of ordinal position, rubrics are assessed with four performance levels in Arkansas (unsatisfactory, basic, proficient, and distinguished). A major problem noted by teachers, administrators and ADE officials is the “Basic” performance level written into Act 1209 (2011, p. 10) because of the pejorative connotation associated with that categorical designation on the state Benchmark exams.

An ordinal scale performance measure certainly offers greater discrimination in performance levels than the long-used dichotomous scale of “*satisfactory*” versus “*unsatisfactory*.” The old dichotomous scale pressured evaluators to rate teachers as *satisfactory* who had deficits that needed to be remedied. Having a performance level above *unsatisfactory*, but below *proficient*, allows the evaluator to rate the need for change in addition to written comments.

Another noteworthy upgrade in evaluating teacher effectiveness is the use of multiple measures or methods of observation. Instead of relying solely on limited classroom observations, teacher evaluations increasingly are based on artifacts or evidence in addition to observations of classroom instruction, such as student assessments, class exams, assignments, and state testing. Act 1209 (2011) requires 50% of these artifacts to be external assessment measures. External measures are measures developed by sources other than the teacher being evaluated. External measures provide direct indications of student performance, and are essential in planning professional learning.

Many states, including Arkansas, are moving to linking teacher evaluation and professional development planning. For example, Act 1209 (2011) in Arkansas states that the professional learning plans shall require at least half of the PD hours required by law to be directly related to teachers' content area, instructional strategies applicable to their content area, and their identified needs. For teachers in intensive support status, all PD, except units required by law, are to be directly related to their needs.

Some states have taken an additional step and are linking student achievement gains to teacher evaluation and professional learning (e.g., Southern Regional Education Board, 2011). The most common methodology used to examine this linkage is to analyze growth scores with value-added modeling (e.g., Braun et al., 2010; Corcoran, 2010; Daley & Kim, 2010; Newton et al., 2010). The flaws and limitations of the value-added approaches are well-documented (e.g., Corcoran, 2010; Newton et al., 2010), and they have been articulated in this report. Succinctly stated, faulty measures, non-random assignment of students and teachers, and the complexities of extraneous influences render isolating the effects of individual teachers virtually impossible. As a result, sophisticated statisticians (e.g., Baker et al., 2010; McCaffrey et al., 2003) have cautioned against using value-added methods for "high stakes" decisions such as dismissal.

At the same time, researchers also note that it is an overreaction to ignore or dismiss value-added modeling. As Glazerman et al. (2010, pp. 8 - 9) observe, "...research is quite clear: if student test achievement is the outcome, value-added is superior to other existing methods of classifying teachers." The efficacy of predicting student performance of other teacher evaluation approaches typically is not addressed in criticisms of value-added methods (Glazerman et al., 2010). What is also overlooked is that the problems of faulty measures, complexity of extraneous influences, and difficulty of isolating the impact of individual teachers on students outcomes applies to classroom observations, portfolios, assembling artifacts, and professional judgments. In addition, professional judgments of classroom performance, portfolios, and artifacts do not systematically examine extraneous factors. It is simply assumed that professional evaluators will mentally consider such factors as student and community characteristics in scoring and/or classifying teachers' performance based on observation, portfolios, and artifacts. While establishing inter-rater reliability would offer more assurance that evaluators are consistent in their evaluative judgments (scoring and classifications), this step in establishing teacher evaluation systems is rarely taken.

A salient flaw of professional judgments is their susceptibility to idiosyncratic philosophical frameworks and personal preferences. Training is done to try to establish uniformity in professional judgments by having well-defined frameworks and rubrics, and these efforts are likely quite successful. However, training typically does not include efforts to establish inter-rater reliability. Professional judgments also are more or less like a "black box" in terms of how individual evaluators weight the many extraneous factors that impact student achievement, in addition to teacher effectiveness.

An advantage of value-added methodology over other approaches is systematic examination of the relative "effects" of teaching and extraneous influences. Although clearly flawed, value-added provides systematic or empirical approximations of the interactive effects of several factors that are considered in a "black box" with other approaches to teacher evaluation. However, owing to faulty measures and statistical complexities that are still being investigated, value-added modeling should not be a deciding factor in personnel decisions.

In conclusion, the most judicious approach to teacher evaluation and professional development would seem to be blending the various approaches discussed in this report. Classroom

observations of teaching, especially when guided by explicit criteria and expectations, provide direct evidence of content knowledge and skills unparalleled by any other method of evaluation. It permits a wider scope of observation of teaching per se than artifacts or designed measures of teacher effectiveness. Artifacts, on the other hand, provide evidence of other teacher responsibilities and products of teaching. Artifacts also indicate progression in individual and collective learning. Portfolios can be a treasure trove of information about the scope and growth of student learning. Teacher portfolios can be valuable indicators of professional growth and present needs for additional PD.

Finally, value-added analyses can be used to get systematic estimates of the relative impact of teaching in a particular school after examining the effects of a variety of extraneous factors such as student and community characteristics. This level of systematic analysis is not achieved with other methods of evaluation.

REFERENCES

Act 1209 – To establish the teacher excellence and support system, and to align provisions of the Arkansas code concerning professional development and teacher fair dismissal. Arkansas 88th General Assembly, Regular Session, 2011

ACT, Inc. *The condition of college and career readiness 2011*. Retrieved November 30, 2011, from, <http://www.act.org/research/policymakers/cccr11/>

Arkansas Educator (2011). *AEA educators work with legislators at the 88th General Assembly*. Retrieved June 22, 2012, from, www.aeaonline.org/images/EducatorJuly13v.2.pdf

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Ravitch, D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute. Retrieved June 5, 2012, from, <http://www.epi.org/publication/bp278/>

Blank, R. K., & de las Alas, N. (2009). *Effects of teacher professional development on gains in student achievement: How meta analysis provides scientific evidence use to education leaders*. Washington, DC: Council of Chief State School Officers. Retrieved May 23, 2012, from, http://www.ccsso.org/Resources/Publications/Effects_of_Teacher_Professional_Development_Gains_in_Student_Achievement_How_Meta_Analysis_Provides_Evidence_Useful_to_Education_Leaders_.html

Brandt, C., Mathers, C., Oliva, M., Brown-Sims, M., & Hess, J. (2007). *Examining district guidance to schools on teacher evaluation policies in the mid-west region*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved June 13, 2012, from, http://www.ies.ed.gov/ncee/edlabs/regions/midwest/pdf/rel_2007030.pdf

Braun, H., Chudowsky, N., & Koenig, J. (Eds.) (2010). *Getting value out of value-added: Report of a workshop*. Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Accountability; National Research Council. Retrieved June 11, 2012, from, <http://www.nap.edu/catalog/12820.html>

Corcoran, S. P. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice*. Providence, Rhode Island: Annenberg Institute for School Reform at Brown University. Retrieved June 8, 2012, from, <http://www.annenberginstitute.org/pdf/valueAddedReport.pdf>.

Cushman, K. (2011). *The teacher excellence and support system: Act 1209 of 2011*. Power Point Presentation, AAEA Conference, August 2, 2011. Retrieved June 14, 2012, from, <http://arkansased.org/educators/index.html>

Daley, G., & Kim, L. (2010). *A teacher evaluation system that works*. Retrieved October 18, 2010, from, http://www.tapsystem.org/publications/wp_eval.pdf

Danielson, C. (2011). *The framework for teaching*. Princeton, NJ: The Danielson Group. Retrieved June 13, 2012, from, <http://www.danielsongroup.org/article.aspx?page=frameworkforteaching>

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan*, 76, 597-604.

Darling-Hammond, L. et al. (2008). *Powerful learning: What we know about teaching for understanding*. San Francisco: Jossey-Bass.

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Washington, DC: National Staff Development Council. Retrieved May 23, 2012, from, <http://www.learningforward.org/news/NSDCstudytechnicalreport2009.pdf>

Donaldson, M. L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality*. Washington, DC: Center for American Progress. Retrieved June 13, 2012, from, http://www.americanprogress.org/issues/2009/06/teacher_evaluation.html

Dufour, R., Dufour, R. Eaker, R., & Many, T. (2006). *Learning by doing: A handbook for professional learning communities at work*. Bloomington, IN: Solution Tree Press.

Dufour, R., & Marzano, R. J. (2011). *Leaders of learning: How district, school, and classroom leaders improve student achievement*. Bloomington, IN: Solution Tree Press

Dwyer, C. A. (Ed.) (2007). *America's challenge: Effective teachers for at-risk schools and students*. Washington: National Comprehensive Center for Teacher Quality. Retrieved June 13, 2012, from, <http://www.tqsource.org/publications/NCCTQBiennialReport.pdf>

Fulton, K., & Britton, T. (2011). *STEM teachers in professional learning communities: From good teachers to great teaching*. Retrieved May, 21, 2012, from, <http://nctaf.org/research/research-papers/>

Gallagher, A. H. (2004). Vaughn Elementary's Innovative Teacher Evaluation System: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education*, 79, 79-107.

Garet, M., Porter, A., Desimone, L., Birman, B., & Yoon, K.S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38, 915-945.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H. S., Doolittle, F., Zhu, P., & Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE 2008-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., & Whitehurst, G. J. (2011). *Passing muster: Evaluating teacher evaluation systems*. Washington, DC: Brown Center on Education Policy at Brookings. Retrieved May 29, 2012, from, <http://www.brookings.edu/research/reports/2011/04/26-evaluating-teachers>

Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., & Whitehurst, G. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: The Brookings Institute. Retrieved January 26, 2010, from, http://www.brookings.edu/reports/2010/11/17_evaluating_teachers.aspx

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed? National Board Certification as a signal of effective teaching." *Review of Economics and Statistics*, 89, 134-150.

Goldhaber, D., & Hansen, M. (2008). *Assessing the potential of using value-added estimates of teacher job performance for making tenure decisions*. Seattle, WA: Center on Reinventing Public Education. Retrieved June 15, 2012, from, <http://www.urban.org/url.cfm?ID=1001369&renderforprint=1>

Harris, D. N., & Sass, T. R. (2007). *The effects of NBPTS-certified teachers on student achievement*. Washington: Center for Analysis of Longitudinal Data in Education Research.

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, 26,101-136.

Koretz, D. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard university Press.

Little, J. W. (1990). The persistence of privacy: Autonomy and initiative in teachers' professional relations. *Teacher College Record*, 91, 509-536.

Little, J.W. (2003). Inside teacher community: Representations of classroom practice. *Teacher College Record*, 105, 913-945.

Little, O., Goe, L., & Bell, C. (2009). *A practice guide to evaluating teacher effectiveness*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved June 25, 2012, from, www.tqsource.org/publications/practicalGuide.pdf

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati." *Peabody Journal of Education*, 79, 33-53.

McCaffrey, D. F., Koretz, D., Lockwood, J. R., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND Corporation. Retrieved June 8, 2012, from http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf

McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The inter-temporal variability of teacher effect estimates. *Education Finance and Policy* 4, 572–606.

Retrieved June 8, 2012, from,

<http://www.mitpressjournals.org/doi/abs/10.1162/edfp.2009.4.4.572>

MET Project (2010). *Learning about teaching: Initial findings from the measures of effective teaching project*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved

May 29, 2012, from, http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf

Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). *Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts*. Educational Policy Analysis Archives, 18, 1-27. Retrieved June 5, 2012, from,

<http://epaa.asu.edu/ojs/article/view/810>

No Child Left Behind Act of 2001, Pub. L. 107-110, 20 U.S.C. § 6301 (2002), January 8, 2002.

Obama, B. (2009, March 10). *Taking on education. Remarks made at the U.S. Hispanic Chamber of Commerce*, Washington, DC. Retrieved May 23, 2012,

<http://www.whitehouse.gov/blog/09/03/10/Taking-on-Education/>

Papay, J. (2011). *Different tests, different answers: The stability of teacher value-added estimates across outcome measures*. *American Educational Research Journal*, 48, 163-

193. Retrieved June 15, 2012, from, <http://aer.sagepub.com/content/48/1/163.full.pdf+html>

Saunders, W. M., Goldenberg, C. N., & Gallimore, R. (2009). Classroom learning: A prospective, quasi-experimental study of Title I schools. *American Education Research Journal*, 46, 1006-1033. Retrieved May 29, 2012, from,

<http://aer.sagepub.com/cgi/content/abstract/46/4/1006>

Saxe, G., Gearhart, M., & Nasir, N. S. (2001). Enhancing students' understanding of Mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education*, 4, 55-79.

Snow-Renner, R., & Lauer, P. (2005). *Professional development analysis*. Denver, CO: Mid-Content Research for Education and Learning.

The Danielson Group (2011). *The framework for teaching evaluation instrument*.

Princeton, NJ: The Danielson Group. Retrieved June 14, 2012, from,

<http://www.danielsongroup.org/downloads.aspx?file=DFfTEI2011.pdf>

Southern Regional Education Board (2011). *Focus on teacher reform legislation in SREB states: Evaluation policies*. Retrieved June 26, 2012, from,

http://publications.sreb.org/2011/11S07_Focus_Teach_Eval.pdf

The New Teacher Project (2010). *Teacher evaluation 2.0*. Retrieved May 29, 2012, from, <http://tntp.org/ideas-and-innovations/view/teacher-evaluation-2.0>

Thompson, M. & Goe, L. (2009). *Models for effective and scalable teacher professional development*. Princeton, NJ: Education Testing Service. Retrieved May 24, 2012, from, www.ets.org/Media/Research/pdf/RR-09-07.pdf

Thoonene, E. J., Slegers, P. J., Oort, F. J., Peetsma, T. D. & Geijsel, F. P. (2011). How to improve teaching practices: The role of teacher motivation, organizational factors, and leadership practices. *Educational Administration Quarterly*, 47, 496-536.

Tucker, P. D., Stronge, J. H., Gareis, C. R., & Beers, C. S. (2003). The efficacy of portfolios for teacher evaluation and professional development: Do they make a difference. *Education Administration Quarterly*, 39, 572-602.

U. S. Department of Education (2008). *The Final Report of the National Mathematics Advisory Panel*. Retrieved October 20, 2010, from, <http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf>

U. S. Department of Education (2010). *Race to the Top Fund: States' applications for phase 2*. Retrieved June 7, 2012, from, <http://www2.ed.gov/programs/racetothetop/phase2-applications/index.html>

U. S. Department of Education (2011). *Schools and Staffing Survey 2011*. Washington, DC: Institute of Education Sciences. Retrieved May 21, 2012, from, <http://nces.ed.gov/surveys/sass/>

U. S. Department of Education (2012). *The Condition of education 2012*. Washington, DC: Institute of Education Sciences. Retrieved May 24, 2012, from, <http://nces.ed.gov/pubs2012/2012045.pdf>

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Dallas, TX. National Staff Development Council. Retrieved May 29, 2012, from, http://srnleads.org/resources/publications/nsdc/nsdc_2009-02_execsumm.pdf

Wei, R. C., Darling-Hammond, L., and Adamson, F. (2010). *Professional development in the United States: Trends and challenges*. Dallas, TX. National Staff Development Council. Retrieved May 29, 2012, from, <http://srnleads.org/resources/publications/nsdc.html>

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*. Retrieved October 18, 2010, from, <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>

William, D (2008). *Keeping Learning on Track*. Princeton, NJ: Educational Testing Service. Retrieved May 24, 2012, from <http://keepinglearningontrack.nwea.org/about.html>

Wilson, M., Hallman, P. J., Pecheone, R., & Moss, P. (2007). *Using test scores as evidence of external validity for indicators of teacher quality: Connecticut's Beginning*

Educator Support and Training Program. Berkley, CA: University of California, Berkley, Graduate School of Education.

Wylie, E. C. (2008). *Tight but loose: Scaling up teacher professional development in diverse contexts*. Princeton, NJ: Educational Testing Service. Retrieved May 24, 2012, from, http://www.ets.org/research/policy_research_reports/rr-08-29

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Retrieved May 30, 2012, from http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/REL_2007033.pdf

APPENDIX A

DATA ON ARKANSAS PROFESSIONAL DEVELOPMENT AND TEACHER EVALUATION

The following data came from surveys of all 239 districts and of a random sample of 74 schools included in the 2011-12 Adequacy Study conducted by the Bureau of Legislative Research (BLR). Table 1 shows what percentage of professional development (PD) is provided by seven different sources according to responses by superintendents on district survey. It reveals, for example, that the bulk of PD is provided by the district, with an average (mean) percentage of 45.69 and a midpoint (median) of 50. The average difference in district-provided PD (standard deviation) is large, indicating considerable variance between districts. Also indicated in Table 1 is the percentage of PD provision at different percentile intervals. Missing data are shown because many superintendents left blanks even though they were instructed to use zeros if a source did not provide PD.

TABLE 1. PERCENTAGE OF PROFESSIONAL DEVELOPMENT PROVIDED BY SOURCE

	AETN	ADE	Co-op	District	District/AETN	Contract	Other
Statistics							
Mean	2.98	7.96	28.88	45.69	0.61	4.41	2.48
Median	1.00	5.00	25.00	50.00	0	0	0
Std. Deviation	4.34	8.37	21.85	25.85	2.70	9.50	9.19
Minimum	0	0	0	0	0	0	0
Maximum	30.00	57.00	94.00	100.00	100.00	75.00	100.00
Percentiles							
20 Percentile	0	0	10.00	20.00	0	0	0
40 Percentile	1.00	5.00	20.00	41.80	0	0	0
60 Percentile	2.00	10.00	31.00	53.80	0	0	0
80 Percentile	5.00	15.00	45.00	70.00	0	10.00	0
Missing Data	43	1	0	1	70	63	94

Note: Missing data are number of districts, not percentages as reported in the rest of the table.

One of the noteworthy advances of the Arkansas Teacher Excellence and Support System (Act 1209 of 2011) is the linkage of PD to teacher evaluation. Under the section on Support Components, Act 1209 states, "...a teacher being evaluated and the evaluator, working together, shall develop a professional learning plan for the teacher that: (A) identifies professional learning outcomes to advance the teacher's professional skills; (B) clearly links professional development activities and the teacher's individual professional learning needs identified through the Teacher Excellence and Support System." (p. 11)

Act 1209 of 2011 also indicates, "The professional learning plan shall require that at least one-half (1/2) of the professional development hours required by law or rule for a teacher are directly related to one (1) or more of: (A) the teacher's content area; (B) instructional strategies applicable to the teacher's content area; (C) the teacher's identified needs."

Furthermore, "...until [a] teacher is removed from intensive support status, all professional development identified in the professional learning plan, except professional development that is required by law or by the public school where the teacher is employed, shall be directly related to the individual teacher's needs." (pp. 11-12)

Additional support is provided by requiring ongoing formative assessments. Act 1209 of 2011 reads, “Interim teacher appraisals shall be used to support teachers on an ongoing basis throughout the school year and: (A) provide a teacher with immediate feedback about the teacher’s teaching practices; (B) engage the teacher in a collaborative, supportive learning process; (C) help the teacher use formative assessments to inform the teacher of student progress and adapt teaching practices based on the formative assessments.” (p. 12)

Table 2 shows the number of formative evaluations of teachers in a given school year according principals of 74 randomly selected schools who responded to the BLR Adequacy Survey. Formative evaluations are designed to support teachers and provide feedback for growth, rather than to rate them as is done at the end of the year in summative evaluations. Two schools were excluded owing to their responses of 180 evaluations being far out of range. The average (mean) number of formative evaluations are nearly the same for inexperienced and experienced teachers; however, the midpoint (median) is higher for inexperienced (2) than for experienced (1) teachers. It may also be noted that 80% of the schools indicated that inexperienced and experienced teachers receive four or fewer formative evaluations in a year.

TABLE 2. NUMBER OF FORMATIVE EVALUATIONS IN A SCHOOL YEAR

	< 2 Years of Experience	3 or > Years of Experience
Mean	3.75	3.21
Median	2.00	1.00
Standard Deviation	6.52	6.48
Minimum	0	0
Maximum	36	36
20th Percentile	1.00	1.00
40th Percentile	2.00	1.00
60th Percentile	2.00	2.00
80th Percentile	4.00	4.00

Note: Two schools were excluded owing to their responses of 180 days being far out of range.

Fifty-six, or 75.7%, of the 74 randomly selected schools reported that they factor formative evaluations in their summative (or final) teacher evaluation (or rating). Nearly all selected schools reported an annual summative evaluation, with the exception of one that left the question blank, for both inexperienced and experienced teachers.

Table 3 shows that 40, or 54.1%, of the 74 randomly selected schools use a protocol with a rating scale (commonly, unsatisfactory, basic, proficient, and distinguished) for their summative evaluations of teachers. Seventy-one principals reported that they always meet with teachers concerning summative evaluations, whereas two principals indicated this meeting occurs when

TABLE 3. TYPE OF SUMMATIVE EVALUATION USED IN ARKANSAS

Type of Evaluation	Number of Schools	Percent
Open Format	2	2.7
Checklist	17	23.0
Protocol with open comments	6	8.1
Protocol with rating scale	40	54.1
Other	9	12.1

requested. One principal responded to an “other” option, and specified that teachers “are given the results.”

Principals indicated that they spend an average (mean) of 109 minutes in a school year observing each teacher in the classroom (Table 4). Obviously, these are educated estimates, and therefore they should not be interpreted as numerical data at an interval or ratio level of measurement.

TABLE 4. TOTAL MINUTES OF CLASSROOM OBSERVATION IN A YEAR FOR SUMMATIVE EVALUATIONS

Minutes of Classroom Observation	
Mean	109.20
Median	80.00
Standard Deviation	118.80
Minimum	0
Maximum	585.00
20th Percentile	38.00
40th Percentile	60.00
60th Percentile	90.00
80th Percentile	156.00

Note: One school was excluded owing to out-of-range response of 3600.

Table 5 indicates how many teachers were separated from the school for performance issues over the past three years. For example, 51 principals reported that no teachers had been separated from the school over that time period. The relatively small number of separations is congruent with what is reported in the professional literature, which discusses the reluctance to dismiss ineffective teachers (e.g., Weisberg et al., 2009).

TABLE 5. NUMBER OF TEACHER SEPARATIONS FROM SCHOOL DUE TO PERFORMANCE

Number of Teachers	Number of Schools	Percentage
0	51	68.9
1	13	17.8
2	6	8.1
3	3	4.1
4	1	1.4

Note: Number of certified teachers separated from the school for performance issues in the past 3 years.

PD AND EVALUATION ISSUES ARISING FROM SURVEYS, SITE VISITS AND ADE OFFICIALS

There are certain specific issues that have been repeatedly raised in the BLR surveys, site visits, and by ADE officials that need to be brought to the attention of legislators. For example, on every BLR survey, and in many onsite visits, teachers and administrators have observed that there are too many mandated professional development (PD) topics and hours. Most of these mandated units of PD have limited apparent impact on classroom instruction, and research evidence shows that teacher effectiveness and student achievement are enhanced by PD devoted to content knowledge and skill development. Instead of requiring many units every year, units thought to be necessary could be staggered over some interval of years (e.g., 5 years). It is also possible to deliver many of these topics, such as child maltreatment and parental involvement, in forums other than PD.

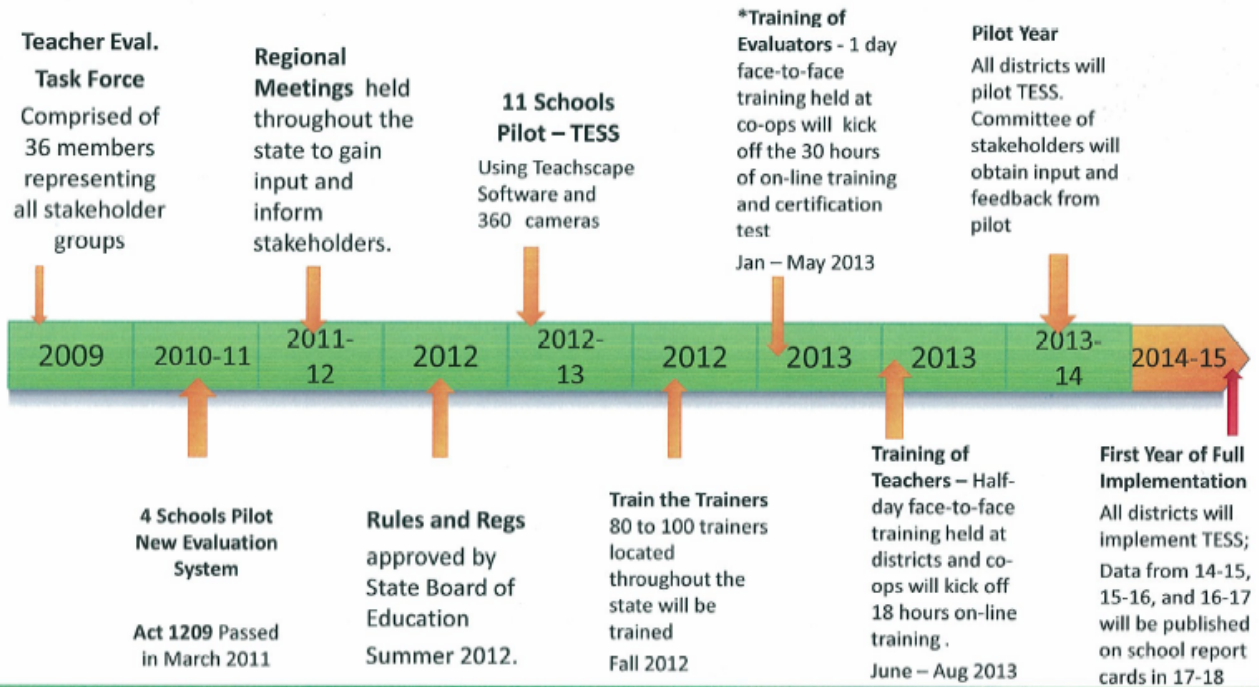
A couple of related concerns that are consistently discussed whenever a question is asked about what changes should be made in PD in Arkansas are redundancy and irrelevance. Some PD units present redundant materials from one year to another, and some units are not relevant for certain teachers. There is ample empirical evidence to show that having too many mandated units, redundancy, and irrelevance are real problems in constructing PD programs that enhance teacher effectiveness and student achievement – they do not contribute to improving teaching and they diminish the time devoted to improving content and skills of teachers (e.g., Baker et al., 2010; Darling-Hammond et al., 2008, 2009; Wei et al., 2009).

Based on evidence of effective teacher professional development as a concept rather than a program, many researchers and policy analysts recommend shifting from designated units of PD to individualized professional learning plans that target specific needs (content and skills) of each teacher. These plans could be developed conjointly between the principal or academic coach and the teacher and they might be periodically monitored by ADE. According to recent professional literature, PD should be based on teacher evaluations.

Based on onsite interviews with teachers, academic coaches, administrators, and ADE officials responsible for implementing Act 1209 (2011), it would appear that the teacher evaluation system being incrementally implemented holds real promise of elevating teacher effectiveness in raising student achievement. There seem to be Herculean efforts being made to devise sound measures of teaching and to train evaluators to achieve fidelity in evaluation measurement. The choice of multiple measures, including requiring 50% of the artifacts to be external measures, and the integration of teacher evaluation and professional learning are especially noteworthy. Given all the various methodological considerations discussed in this report, it appears that choices made by the coalition of educators that created Act 1209 are aligned with the state of knowledge about evaluating teachers (Arkansas Educator, 2011). The timeline for implementing the various components of Act 1209 is in Appendix B.

APPENDIX B

ARKANSAS TEACHER EXCELLENCE SUPPORT SYSTEM (TESS)



*The face-to-face training will take place in January and February 2013 – Evaluators will need to complete the on-line training and certification test by August 2013