Evaluating the Continued Implementation of the Teacher Excellence and Support System (TESS) and the Implementation of the Leader Excellence and Development System (LEADS) in Jonesboro, Arkansas

Gregory E. Cox and Thomas A. Suchman

Vanderbilt University

Table of Contents

Executive Summary	3
Section One: Introduction	7
Section Two: Geographical Context	16
Section Three: Methods	19
Section Four: Question 1	32
Question 1A	44
Question 1B	50
Section Five: Question 2	54
Question 2A	55
Section Six: Question 3	59
Section Seven: Themed Findings	74
Section Eight: Discussion	83
Section Nine: Recommendations and Conclusion	91
References	99
Appendixes	105

The Arkansas Teacher Excellence and Support System (TESS) was enacted in 2011. TESS is based on Danielson's Framework (Danielson & McGreal, 2000) and is grounded in a wave of policy changes made to improve teacher evaluations systems throughout the nation. The system was piloted during the 2013-2014 school year, and was fully implemented during the 2014-2015 school year. The Leadership Excellence and Development System (LEADS) was enacted in 2013 to improve principal evaluations in a similar way that TESS intended to improve teacher evaluations. LEADS was fully implemented during the 2014-2015 school year.

In 2014, an evaluation of early implementation of TESS was conducted. This study found that while there were some successes with early implementation, there were also challenges related to variation in teacher perceptions and school site implementation as well as multiple tradeoffs that affected the quality of implementation.

This study is an update to the prior study designed to investigate continued implementation of TESS and provide similar information on the implementation of LEADS in Jonesboro, Nettleton, and Westside Consolidated school districts. In order to approach this task, we devised three new, multi-part, project questions:

1. To what extent has implementation of TESS reflected best practices and achieved policymakers' intended goals?
    a. How have teacher and principal perceptions of TESS implementation changed over time?
    b. How have principals and superintendents worked together to improve TESS implementation and promote desired outcomes?
2. To what extent has the implementation of LEADS reflected best practices and achieved intended goals?
    a. What role has LEADS played in influencing how principals implement TESS?
3. What evidence is there that TESS has impacted teacher practice and principal leadership?

Similar to the first study, we used a mixed methods approach. We conducted interviews and surveys with both teachers and administrators. Modified original

surveys and interview protocols were used to see how both teachers' and administrators' perceptions of TESS had changed since implementation. Data from the first study were obtained so that a statistical analysis could be made of the change in teacher and administrator perceptions. Prior research focused on policy implementation, teacher and principal evaluation, and the social and organizational context of schools.

Findings

Many of the findings of the previous study are still applicable, yet there have been changes in the intensity and scope of the findings.

Growing Levels of Comfort and Frustration

In 2016, teachers feel more comfortable with the TESS process and generally believe that the rubric reflects effective teaching. Teachers are also confident that they can demonstrate the skills covered by the rubric. However, teachers are increasingly frustrated that TESS is taking too much time away from educating students and is having a negative effect on their ability to conduct their work as they think best.

Continued Variation Between Schools and Districts

Teacher and administrator perceptions of TESS and LEADS vary by schools and districts. Qualitative data suggests that variations in implementation approaches, school personnel, and school culture have an effect on teacher and principal perceptions of the processes.

Unrealized Aspirations and Unintended Outcomes

There is disagreement about the effect of TESS on teacher practice. Some outcomes seem to confound the intended purpose of TESS, and some teachers believe TESS hampers the emergence of school communities. Moreover, the process to evaluate educator practice and professionalism has a negative effect on some teachers' senses of autonomy and professional well-being. There are also conflicting understandings of how TESS and LEADS should be used for summative and formative purposes.

A Series of Continued Trade-offs

While teachers and administrators struggle to find the time to complete the requirements associated with TESS and LEADS, teacher and administrator

perceptions of TESS are improving. The possible improvement in teacher and principal quality must be balanced with what many educators view as invasive observations and a disproportionate amount of work. Also, while principals suggest that while TESS should be used for non-contract renewal, it is rare for the process to actually work toward that purpose. This conflict between resources and opportunity cost is a major finding in the study.

The implementation of LEADS has stalled behind the time commitments associated with TESS. Some principals are also frustrated that the training for LEADS has not been as rigorous as it was for TESS.

Recommendations

Our recommendations fall into three main categories: general suggestions about program implementation, refining the use of resources, and separating the formative and summative aspects of teacher and administrator evaluations.

Program Implementation Should Consider Training Methods and the Needs of Workers

When additional tasks are added to employees' workloads, there needs to be a lessening in other duties in proportion to the new ones added. High quality training is more involved than simply providing information (Fixsen, Naoom, Blase, Friedman & Wallace, 2005); as a result, implementing new training requires a significant amount of time. Teachers and principals are overwhelmed. Moreover, implementation of individuals' TESS requirements needs to be phased in gradually as new teachers and administrators learn their jobs in order to give educators time to adjust. Danielson (2011) agrees that evaluation systems should be differentiated for different levels of professional development. Feiman-Nemser (2001) and Curtis and Wurtzel (2010) write about the importance of reducing new teachers' workloads. Finally, during implementation, it is essential to attend to the individual needs of the workers; some teachers still do not know their current track placement.

Seek a More Efficient Use of Resources

The school districts might consider reducing or eliminating the requirement of submitting artifacts. However, teachers should be encouraged to add artifacts for the benefit of their own practice. We also recommend that the teachers' growth efforts

be more fully differentiated according to the level of teacher development.

Consider Separating Formative and Summative Functions of TESS and LEADS

Administrators (and some teachers) doubt that TESS effectively removes low quality teachers. However, the possibility of adverse teacher employment decisions undermines effectiveness as a developmental tool by reinforcing a "fixed" rather than a "growth" mindset by tying consequences to evaluation results. TESS and LEADS should be redesigned as purely formative exercises so that they can be effective for this purpose, and another system should be established for effective summative accountability. Such changes could include more observations by peers rather than administrators, teacher input into the choice of observer, de-emphasizing the distinction between the two top ratings levels, and removing punitive consequences for the ratings.

Section One: Introduction

NCLB and Teacher Evaluations

TESS and LEADS are indicative of the broader national policy shift toward teacher and administrator quality. It is important to understand how the continued implementation of methods to improve teacher and administrator evaluation affects individuals, schools, and districts. President George W. Bush signed The *No Child Left Behind Act* (NCLB) into law on January 8, 2002. NCLB sought to improve student outcomes by putting in place a series of sanctions for those outcomes. This increased the incentives faced by schools and districts to improve student performance. As teacher quality is linked to student performance (Darling-Hammond, 1995; National Commission on Teaching and America's Future, 1996; Rockoff, 2004), efforts to measure and promote teacher quality gained greater emphasis (National Center for Teacher Quality, 2012). As a result, a strategy that accompanied the advent of NCLB was to implement new approaches to teacher evaluations.

In 2010, a revision in the NCLB mandate called for all states and districts to revise teacher and administrator evaluation systems (U.S. Department of Education, 2010). Traditionally, teacher qualifications such as education and career history had been the measure of teacher effectiveness during teacher reviews (Greenberg, Rhodes, Ye, & Stancavage, 2004). Even when teachers were evaluated more thoroughly, teacher evaluation systems were often composed of binary measures; teacher evaluations typically consisted of checklists where teachers were noted for "doing their jobs" or "not meeting expectations" (Sartain, et al 2011).

Critics and reformers could fault this approach for numerous reasons. For those concerned with the formative purpose of improving the quality of the existing teaching force, evaluation systems rarely provided feedback that was sufficient for teachers to improve their practice. For those more oriented to the summative purpose of removing poor teachers from the classroom, this system also failed for that purpose; only the most egregiously incompetent teachers or teachers whose practices were otherwise detrimental to student progress were rated as unsatisfactory. The frequency of unsatisfactory ratings was low enough to cast doubt on the assumption that all teachers who rated "satisfactory" actually merited such a rating.

While NCLB does not specifically mention using evaluation scores to make teaching personnel decisions, The National Council for Teacher Quality writes, "Using evaluation results to make personnel decisions is where the rubber meets the road, and more states than ever before are attaching consequences to teacher effectiveness" (2012, p. 3). As a result, a significant debate has risen about the purpose of teacher evaluation: whether or not the purpose should be for helping teachers improve, or for identifying weak teachers for dismissal. Whichever might be the case, before changes in evaluation systems were made, they rarely afforded schools the ability to fire teachers whose work was ineffective, or to recognize teachers who successfully led classrooms of American students.

Teacher Evaluation in Arkansas

In Arkansas, "Because there were no descriptors or rubrics, expectations were not clear. This lack of clarity provided little targeted feedback for teachers in improving their professional practice and improving student learning. The prior evaluation system relied on a vague checklist of classroom practices" (Office of Educator Effectiveness, n.d., p.2). As a result, teachers in Arkansas, as well as other states, tended to merely try to meet the mediocre standards of the evaluation tool, rather than improve their craft as teachers (Office of Educator Effectiveness, n.d.). Moreover, the typically crude measures of teacher performance rarely gave schools a clear enough picture of what was actually occurring in the classrooms of individual teachers on a daily basis. In order to attempt to improve the quality of their education in compliance with the NCLB mandate, many states needed to adopt new ways of measuring and promoting teacher quality.

The state of Arkansas set up a taskforce of diverse stakeholders in 2009 "with the purpose of researching, evaluating and recommending a framework for summative evaluation that would include valid assessment of educator practice and professionalism, as well as evidence of educator impact on student growth and performance" (Arkansas Department of Education, 2015). The taskforce worked with Charlotte Danielson, author of *A Framework for Teaching*, to "research, evaluate, and recommend" a new teacher evaluation system that would help teachers focus on the quality of their teaching (Office of Educator Effectiveness, n.d., p.1).

Drawing on this work, the Arkansas legislature created the Teacher Excellence and Support System (TESS) in an effort to improve evaluation practice. In 2011, the Arkansas General Assembly introduced and passed Ark. Code Ann.§6-17-2802 to standardize comprehensive evaluation and support for licensed educators and non-licensed teachers employed in public charter schools under a waiver of teacher licensure requirements granted by the State Board of Education in the schools' charters. TESS was enacted in 2013, and districts had an option of piloting the system that first year. The system was fully implemented statewide during the 2014-2015 school year.

The stated goals of the system are as follows:

• Provide school districts with a transparent and consistent teacher evaluation system that ensures effective teaching and promotes professional learning;
• Provide feedback and a support system that will encourage teachers to improve their knowledge and instructional skills in order to improve student learning;
• Provide a basis for making teacher employment decisions;
• Provide an integrated system that links evaluation procedures with curricular standards, professional development activities, targeted support, and human capital decisions;
• Encourage highly effective teachers to undertake challenging assignments;
• Support teachers' roles in improving students' educational achievements;
• Inform policymakers regarding the benefits of a consistent evaluation and support system in regard to improving student achievement across the state;
• Increase the awareness of parents and guardians of students concerning the effectiveness of teachers. (Arkansas Annotated Code 6-17-2802).

Along with the goals of the state listed above, TESS is also designed to empower every teacher to be able to actively seek to improve his or her practice, and facilitate more in-depth conversations between teachers and their principals about their methods used in the classroom. In addition, TESS is designed to promote collaboration between educators to improve teaching practice. While the documentation and evaluation process of TESS is time intensive, the process is intended to foster reflection and dialogue concerning best teacher practices.

Meanwhile, because school leadership effectiveness is related to student

learning and teacher effectiveness (Minor, Porter, Murphy, Goldring, Cravens, & Elloitt, 2014; Whalstrom & Louis, 2008), in 2009, the Arkansas General Assembly passed legislation that also sought to create a better way to evaluate administrator effectiveness. Act 222 created a taskforce consisting of a diverse group of stakeholders including, "superintendents, principals, teachers, educational cooperatives, institutions of higher education, school boards, and other education associations" (Office of Educator Effectiveness, 2015, p. 62). The result was the Arkansas Leader Excellence and Development System (LEADS).

Teacher Excellence and Support System (TESS)

TESS has adopted Danielson's framework of four domains: 1) Planning and Preparation, 2) Classroom Environment, 3) Instruction, and 4) Professional Responsibilities. Each domain houses multiple components that embody Danielson's Framework for Teaching. In all, there are twenty-two components that teachers need to document throughout the TESS evaluation process. TESS's domains attempt to connect teacher quality to classroom practice, which is a connection that had been missing from many earlier forms of teacher evaluations (National Council on Teacher Quality, 2011).

While TESS is primarily designed to improve teaching (and by extension, student performance), low scores on the TESS evaluation system can result in a teacher's placement on an "Intensive Support Track" which can eventually lead to the teacher's dismissal (Arkansas State Department of Education, 2015).

TESS Teacher Tracks

The TESS system is split into three tracks. While all teachers are observed yearly, announced formal observations take place on a one- or four-year cycle. No matter the particular track on which teachers are placed, all teachers need to record their proficiency in each of the four domains on BloomBoard, an electronic tool used to store teacher artifacts and give teachers access to professional development content. The artifacts on Bloom Board ideally inform yearly decisions about Professional Growth Plans (PGP).

Track 1 is a three-year program for probationary or novice teachers. The cycle is virtually the same for each of the three years. Track 1 is designed to teach teachers the TESS process and how to develop PGPs

and document artifacts on BloomBoard. If the need should arise, school districts have the right to extend Track 1 for an extra year with a majority vote from the individual school district's directors. All new teachers, and all teachers new to school districts, are placed on Track 1. During each of the three years on Track 1, each teacher has a summative evaluation on all of the components of Danielson's framework. Formative evaluations during each of the three years focus on target growth areas. After the first formative evaluation of each year, the teacher develops his/her PGP, which is reviewed at the end of each year.

Track 2 is a four-year cycle for teachers who have moved beyond probationary status. Track 2 is divided into two sub-sections: 2A Summative Appraisal and 2B Interim Appraisal. An element of 2A is a summative evaluation occurring every four years over each component of Danielson's framework. During Track 2A, teachers will continue to receive formative observations focusing on the teacher's PGP. PGPs may be modified after each formative observation. Teachers in Track 2B do not receive a summative evaluation for three years. Even though summative appraisals do not take place, teachers will still receive informal formative evaluations during each of these three years. In all interim appraisal years, observations may be targeted on specific areas of the teacher's PGP. During each year, teachers and administrators collaborate on the teacher's PGP, and each teacher receives a performance rating based on PGP goals.

In 2015, changes were made to the TESS model, and an extra year was added to Track 2. While the changes are significant, the timetable of the implementation of the TESS system did not change.

Track 3 is designated "Intensive Support Status". Teachers can be placed on Track 3 at any time from either Track 1 or Track 2. Track 3 is for teachers who have received an "unsatisfactory" rating on any of the entire domains, or who have received a majority of "basic" or "unsatisfactory" ratings in the majority of components of a domain. On Track 3, teachers receive summative evaluation on all components. Teachers on Track 3 also experience multiple formal and informal evaluations as well as multiple conferences with evaluators. Teachers remain in Track 3 for two semesters. If needed, two additional semesters may be added. If teachers improve, they may be placed back in Track 1 or Track 2A. If teachers do not improve,

TESS claims is designed not to conflict with the Arkansas Teacher Fair Dismissal Act (ATFDA), and teachers may be recommended for termination or non-renewal.

Principal Evaluation in Arkansas

In 2009, the Arkansas General Assembly passed different legislation that sought to create a better way to evaluate administrator effectiveness. The result was the Arkansas Leader Excellence and Development System (LEADS). Using a similar structure as TESS, LEADS has become Arkansas' official method for monitoring public school administrators. The purpose of the Arkansas Principal Evaluation System is to:

- Provide a cohesive process that includes clear expectations to guide principal preparation, induction, and continued professional development;
- Guide and sustain excellent leadership performance that ensures the improvement of teaching and learning;
- Assist higher education programs in developing the content and requirements of degree programs that prepare prospective principals;
- Provide a process that includes instruments to be used by reflective practitioners to promote their professional growth. (Office of Educator Effectiveness, 2014).

Leader Excellence and Development System (LEADS)

Based on the recommendations of the task force, the framework for LEADS passed in 2011 and was reaffirmed in 2013. LEADS is used to evaluate principals, assistant principals, superintendents (including associate, deputy, and assistant superintendents), and other educational administrators. The LEADS program was piloted between the 2011-2012 and 2012-2013 school years, and was implemented across the state during the 2013-2014 school year.

The LEADS rubric is divided into six standards. These standards are based on the Interstate School Leaders Licensure Consortium (ISLLC) standards. Administrators are scored with one of four ratings for each category: Exemplary, Proficient, Progressing, and Not Meeting Standards.

LEADS Administrator Evaluation Tracks

The LEADS model is structured much like the TESS model. Like the TESS model, the LEADS model has three basic categories (similar to the TESS Tracks): 1) Probationary/Novice, 2) Inquiry, and 3) Intensive.

The Probationary/Novice Category is for new administrators or for administrators who have relocated to Arkansas from other states. This category lasts three years. During each year an administrator is in the Probationary/Novice category, the administrator will receive a summative assessment as well as formative assessments. The administrator's professional growth plan will also be monitored and revised. Much like the first track of the TESS system, the first category of the LEADS system is designed to help new principals and assistant principals become more familiar with and learn the LEADS model.

Again, like TESS, the second category, Inquiry, is a four-year cycle broken up into two sub-categories: a Summative Evaluation Year and three Interim Appraisal years. During the Summative Evaluation Year, principals and assistant principals undergo a summative assessment conference. Administrators in this sub-category also monitor and revise their PGPs and have formative assessment conferences.

Administrators in the Inquiry category who are NOT in their summative year, work on (and collect evidence for) their PGP, engage in conversations about their PGP and evidence, and have an end of the year conference in which they receive a new PGP and performance rating.

The third category is the Intensive Category. Administrators can enter the Intensive Category from any year of the Probationary/Novice Category, or the Summative Evaluation Year of the Inquiry category. While in the Intensive Category, administrators complete intensive work on their PGP, and participate in multiple formative assessment conferences and observations. During each year administrators are in the Intensive Category, they will also receive a summative evaluation. If at the end of two semesters, the administrator is not able to leave the Intensive Category, two additional semesters may be added. If an improvement is made, administrators may return to any year of the Probationary/Novice Category or the Summative Evaluation Year of the Inquiry Category.

Requirements Associated with TESS and LEADS

Both TESS and LEADS take a significant amount of time for teachers and administrators to document their practice and observations. Formal observations are preceded by a pre-conference and followed by a post-conference. There is also a tremendous amount of teacher and administrator effort that is concomitant to the observation practice including multiple meetings and logging evidence into a program called BloomBoard.

The state of Arkansas also has determined that the TESS and LEADS systems must be used by school districts to consider termination for teachers who make "unsatisfactory" ratings for three consecutive semesters and administrators who make "unsatisfactory" ratings for two consecutive semesters (Office of Educator Effectiveness, n.d.).

Purpose of the Project

The purpose of this project is to follow up on the work of Ashby, Frank and McClain (2014) evaluating the implementation of the TESS teacher evaluation system in Arkansas, and to extend that study by examining early implementation of LEADS. To supplement state-level monitoring of the implementation of TESS, four districts near Jonesboro commissioned an independent study of early stage implementation of TESS by Vanderbilt University doctoral students. The results of this study helped inform early refinements of TESS implementation.

Having had the opportunity to address early stage implementation issues, the Jonesboro, Westside, and Nettleton school districts have asked us to look at the continued efficacy of the implementation. This study follows up on that work and also looks at LEADS. This project asks specifically:

1. To what extent has implementation of TESS reflected best practices and achieved policymakers' intended goals?
    a. How have teacher and principal perceptions of TESS implementation changed over time?
    b. How have principals and superintendents worked together to improve TESS implementation and promote desired outcomes?

2. To what extent has implementation of LEADS reflected best practices and achieved intended goals?
    a. What role has LEADS played in influencing how principals implement TESS?
3. What evidence is there that TESS has impacted teacher practice and principal leadership?

In order to address these questions, we designed a mixed-methods project. By using mixed methods of quantitative and qualitative data collection, we believe we developed a clearer picture of the implementation of the TESS and LEADS systems: "The benefits are derived from drawing on the strengths of qualitative methods to answer questions about *how* and *why* a phenomenon occurs and those of quantitative methods to examine how often a phenomenon occurs and establish generalizable, empirical associations between variables and outcomes" (Smith, Cannata, Haynes 2014). Taking into account the varied issues illuminated by the previous study, the literature review, and the prior learning and experiences of project participants, we expanded the original set of survey questions for teachers and administrators. The more extensive surveys were administered in late November of 2015. We also made three trips (October, November, and January) to interview teachers, principals, and superintendents. This report will present evidence about the ongoing implementation of TESS and LEADS, and the impact those systems have had so far, analyze that evidence in light of relevant scholarship, and on the basis of that analysis, make recommendations to the state and to local districts about how the systems can be made more effective.

Section Two: Geographical Context

The study involves three school districts in and around Jonesboro, Arkansas (Jonesboro Public Schools, Nettleton School District, and Westside Consolidated School District), which serve as a microcosm of the state. Each of the three school districts vary in size, which gives us a broader view of how TESS and LEADS implementation may affect differently sized school districts across the state.

Jonesboro, Arkansas, was established in 1859 in the northeast corner of the state. A city with a population of 72,210, Jonesboro's nearest town of over 200,000 people is Memphis, Tennessee, which is located 63.8 miles to the southeast (factfinder.census.gov accessed, February 10, 2016). While Jonesboro is not a large city, it houses branches and outlets of significant numbers of national and chain businesses. As the home of Arkansas State University and four school districts, there are many fast-food restaurants where students can afford to eat, and images of Howl, the mascot of the Arkansas State Red Wolves, is omnipresent (www.city-data.com accessed January 16, 2016, www.jonesborochamber.com accessed, January 15, 2016).

Economically, Jonesboro seems like a fairly typical small town. In 2013, the median household income in Jonesboro was $40,046. This median is slightly lower than the Arkansas median income of $41,264. The median house or condominium value in Jonesboro is $135,600, which is higher than the Arkansas median value of $108,700 (factfinder.census.gov accessed, February, 10, 2016). The cost of living index rating is 82.7, which is lower than the US average of 100. The unemployment rate in Jonesboro during June 2014 was 5.6, which was higher than the Arkansas rate of 5.0 (factfinder.census.gov accessed, February, 10, 2016).

Jonesboro is also a growing city. According to the US Census Bureau and Community Analyst, between the years 2000 and 2012, Jonesboro has grown 26.4 percent. Jonesboro is projected to have a population over 75,000 by the year 2017 (www.city-data.com, accessed January 15, 2016). There is also a new shopping mall, golf courses, and a downtown artistic district filled with shops and restaurants.

Jonesboro's racial make-up is primarily white. Caucasian citizens make up 74.7 percent of Jonesboro's population. African Americans are the

second most prominent racial demographic, representing 18.4 percent of the population. Hispanics and Asians are less prominent racial presences at 5.2 percent and 1.5 percent, respectively (factfinder.census.gov accessed, February, 10, 2016). As we visited the schools in Jonesboro and toured the city, we noticed that some areas appeared to have higher specific racial concentrations, and the schools generally seemed to reflect the racial demographics of the surrounding neighborhoods (www.city-data.com, accessed January 15, 2016).

The educational attainment of the city is somewhat different from the rest of the state. Jonesboro has higher than state average levels of adults who have bachelor's degrees, master's degrees, and doctorate degrees, likely due to the proximity of Arkansas State University (www.city-data.com, accessed January 15, 2016).

The major industries in Jonesboro are educational services, health care, social services, manufacturing, and retail trade. Accommodation and food services, construction, finance and insurance, and department and town administration and support (including waste management services) are also significant employers in the town (factfinder.census.gov accessed, February, 10, 2016). It is important to note that as of 2013, 11.5 percent of the town's population was employed in education services. This again demonstrates the significant presence of the University and four school districts (www.city-data.com, accessed January 15, 2016).

Jonesboro is the home of four school districts: Jonesboro School District, Westside School District, Nettleton School District, and Valley View Public Schools. Valley View did not participate in this study. The Jonesboro School District schools are typically located in the central and north part of the city. The Westside campuses are to the south west of most of the Jonesboro schools, on the outskirts of the city. The Nettleton schools are primarily based on the southeast side of the city. There is also a private high school in the city, the Ridgefield Christian School.

The Jonesboro School district is the largest of the school districts involved in this study. It teaches over 6000 students and consists of one pre-Kindergarten school, six magnet elementary schools, two middle schools, and one high school. The school district employs twenty-five principals and 405 teachers. The student racial makeup is fifty percent Caucasian, forty-one percent African American and eight percent

Hispanic. Sixty-eight percent of the students are on free or reduced lunch.

Somewhat smaller than Jonesboro, the Nettleton School District houses one pre-kindergarten school, two elementary schools, two intermediate schools, one junior high, one middle school, and one high school. The district teaches 3266 students and employs 255 teachers and fifteen principals. The district is thirty-seven percent Caucasian, forty-seven percent African American, and seventeen percent other minorities. Sixty-eight percent of Nettleton's students receive free or reduced lunch.

Westside Consolidated Public Schools is the smallest of the three districts involved in the study. The elementary, middle, and high school are all located on the same campus. There are 1776 students in the district. The district is ninety-five percent Caucasian, and five percent minority. Fifty-four percent of the students receive free or reduced lunch. Westside is also classified as a "fringe rural" district by *US News and World Report*, whereas Jonesboro Public Schools and Nettleton Public Schools are both classified as "small town" school districts (www.usnews.com, accessed January 15, 2016).

Section Three: Methods

Study Type

A starting point for this project was selecting the type of study to conduct. This process of selection was driven by a combination of the object of study and the nature of the questions for which we sought answers. One set of considerations had to do with a need for broad sets of data. We needed to gather data that applied to all three school districts as a whole, not fragmentary or unsystematic information. Insofar as we were successful at assembling comprehensive data sets about these three districts, we hoped to be able to shed light on the implementation of TESS and LEADS throughout Arkansas, even on the implementation of similar systems nationally. These concerns led to the conclusion that we needed to collect quantitative data from as broad a range of respondents as possible. This set of considerations favored a quantitative approach.

However, we also wanted to make sure that the information we gathered would provide perceptive insight into the inner functionings of the evaluation systems and the organizational contexts in which they are applied. In order to understand more fully, we needed to investigate not merely what was the case, but also why that was so. In a case such as ours, where we are looking at behavior and processes within human organizations, we needed not only to pose open-ended questions, but also to be able to probe and follow up on responses. This requirement favored a qualitative approach.

Rather than seeing such approaches as incompatibly in epistemic opposition, we selected a mixed methods approach that seeks to combine the strengths of both. The quantitative approach would yield numerical answers that would in turn permit the use of descriptive and inferential statistical techniques. Descriptive techniques would yield measures of dispersion and of central tendency in the data from particular questions, while inferential techniques would shed light on how various factors fit together and influence one another. Qualitative approaches might give additional information about *what* is the case, but the main value of using qualitative methods is that they help us understand *why* something is (or is not) the case. Combining the two approaches in our mixed methods approach leverages the strongest aspects of both research traditions to enable a broader and deeper account of the project questions than

would be possible using one or the other approaches alone (Sieber, 1973). Johnson and Onwuegbuzie (2004) write, " . . . research methods should follow research questions in a way that offers the best chance to obtain useful answers. Many research questions and combinations of questions are best and most fully answered through mixed research solutions."

Conceptual Framework

Our process of developing the conceptual framework for our study began with using the original conceptual framework of the prior study as a model. The prior group's initial conceptual framework had categories of rural context, teacher evaluation, and policy implementation. We kept the latter two of the three categories, and added a third of our own. We left out rural context category, since our initial fact finding visit to the site led us to conclude that this was a somewhat inapt way to describe the community, and we added a category of principal evaluation since our study added in LEADS (the principal evaluation system) as a topic.

Within the first category of policy implementation, we have two subcategories: capacity and will, and local context. "Capacity" has to do with the capabilities of the teachers and of the administrators. These capabilities must be developed with high quality training including coaching, feedback, reflection, and collaboration (Darling-Hammond, 2012; Fixsen et al., 2005; Goe, Biggers, & Croft, 2012) 2012), and this in turn requires adequate time. Evaluators in particular must have the capacity to give high quality feedback (Halvorsen et al., 2004). "Will" is adequate motivation to do what is required. This requires "buy-in" (Coburn, 2003; Desimone 2002), and for administrators in particular it requires a willingness to take on the political conflict that candidly critical feedback generates (Halvorsen et al., 2004; Kimball & Milanowski, 2009). It is in the local context in which this must be generated.

Circumstances and conditions vary across the locations where policies are implemented, and these can affect the outcome. Local conditions can shape how implementation is understood and the ability to carry it out effectively (Crowson and Goldring, 2009; Datnow and Park, 2009; Honig, 2006), and the motivation to carry it out to deep effect and in sustainable ways (Coburn, 2003; Datnow and Park, 2009; Desimone, 2002).

For teacher and principal evaluation, we had subcategories of measurement and formative vs. summative purpose of evaluation. There are a number of measurement issues. One is a historical lack of agreed credible standards of practice (Danielson, 2011; Darling-Hammond, 2012; Goe et al., 2012; Grossman, 2001; Lortie, 1969). Another is the consistency of evaluators across time and between different evaluators (Danielson, 2011; Goe et al., 2012). The subcategory of formative vs. summative addresses for what purpose the evaluations should serve. Formative purposes would use the evaluations as part of a system of fostering growth, while summative purposes would use the evaluations as part of a system of accountability. Since both are important functions, the debate is largely over whether or not the same system can serve both purposes (Clipa, 2011; Danielson, 2011; Duke, 1990; Goe et al. 2012).

After our data collection, we refined our framework to better fit the data we actually collected, as the previous group had revised their conceptual framework in light of their findings. In our case, we combined teacher and principal evaluation and added a category addressing the social organization of schools, which emerged as important in the initial fieldwork.

This third category, social organization of schools, has sub-categories of trust, collaboration, power and authority, and organization of time. Trust of the evaluator's competence and impartiality has a crucial role to play (Clipa, 2011; Johnson, 2012; Kimball & Milanowski, 2009; Sartain et al., 2011). Trust more generally within the organization is a precondition to collaboration, the collective and cooperative effort that is crucial in promoting improved teaching and learning (Johnson, 2015; Lee, Dedrick, and Smith, 1991; Murphy, 2013; Scribner, Hager, & Warne (2002). How power and authority work within a school crucially affect professionalism (Lortie, 1969), the closely related issue of teacher autonomy (Lee et al., 1991), and the broader extent to which the school functions as a community (Beck & Foster, 1999; Lee et al. 1991; Sergiovanni, 1994). Finally, Firestone et al. (2013) found that these systems were quite time-intensive, and considerations detailed above suggest some of the reasons why, from the time needed to build capacity to the need to manage policy specificity (Desimone, 2002). These three categories played a crucial role in the analysis and interpretation of our data, described below.

Refined Conceptual Framework		
Policy Implementation	Teacher and Principal Evaluation	Social Organization of Schools
• Capacity and Will • Local Context	• Measurement • Formative vs. Summative Purposes	• Trust • Collaboration • Power/Authority • Organization of time

Quantitative Data

Nearly all of our quantitative data is contained in our survey results. Our survey instruments were adapted from those used in the earlier study (Ashby et al., 2014) in order to facilitate an examination of how perceptions of implementation changed over time. These original instruments had categories addressing perception of the systems and the organizational capacity to implement the evaluation systems. The teacher survey asked questions corresponding to the elements of the first study's revised conceptual framework and map onto elements of our conceptual framework: communication, training, experience and expertise, attitudes and beliefs, time and resources, compatibility with competing policies and programs, professional culture, and alignment with human capital, as well as typical demographic questions (White, Cowhy, Stevens, & Sporte, 2012). The administrator survey asked about the same issues. The wording and format of the questions followed the pattern of several state surveys of teacher evaluation implementation.

Several elements were added to these original instruments that reflect our revised conceptual framework. The first set of adjustments included the addition of questions about the LEADS system to the administrator survey pursuant to our second project question. These questions were largely patterned after corresponding questions about the TESS system. We also added questions about whether the systems were perceived as formative or summative, about perceptions of evaluator expertise and impartiality, and (for LEADS) about the ability of the evaluator to gather relevant evidence in conducting the evaluation. In all cases, these questions reflected recurring

concerns about educator evaluation from the scholarly literature.

The teacher and administrator surveys were entered into Qualtrics, a widely used and reputable electronic survey platform. Email links to the surveys were distributed to the superintendents of the Jonesboro, Nettleton, and Westside districts in late November 2015. The superintendents then sent the survey links to all building principals. The building principals coordinated the distribution of the email links to other administrators (in the case of the administrator survey) and to teachers (in the case of the teacher survey). The building principals used a variety of approaches in arranging for the completion of the surveys. Forty-five percent of administrators (25/56, including superintendents, principals, and assistant principals) and fifty-six percent of teachers (377/674) completed the surveys. When analyzing the school level of the respondents, we found that teachers of younger students were overrepresented in our set of respondents compared with the population of teachers in the districts, while teachers of older students were correspondingly underrepresented. By contrast, the years of experience in our samples were within the known ranges for the populations (see tables 1 and 2 below).

Table 1: Percentage of Total Respondents by School Level and Percentage of Teachers by School Level

Table 1: Percentage of Total Respondents by School Level and Percentage of Teachers by School Level

Teacher Sample and Population	Elementary/Intermediate	Middle/Junior High	High	Other (PK, K, ALC, Tech)	Median Years of Experience
Survey Sample	58% (216/371)	19% (69/371)	16% (59/371)	7% (26/371)	10-20 (range)
Three District Population	42% (284/674)	24% (162/674)	22% (148/674)	12% (81/674)	15.3

Table 2: Administrator Sample and Population	Median Years of Experience
Survey Sample	10-20 (range)
Three District Population	17.1

The other type of quantitative data we collected was the evaluation data for each district from the 2014-2015 school years. The central offices of each district provided these data to us, with 234 results from Nettleton, 85 from Jonesboro, and 110 from Westside.

Interview Data

Similar to the survey instruments, our interview protocols were adapted from those developed by the prior study team and refined according to our conceptual framework. The interview protocols contained broad categories of questions regarding perception and capacity and were patterned after published interview protocols in the state surveys of teacher evaluation system implementation. In the case of the teacher interview protocols, the questions addressed training, expectations, skills, instrument quality, the role of the evaluator, effects of the process, scope for improvement, supports, challenges, and collaboration. The principal interview protocol asked about the same issues, as well as changes in teacher evaluation practice. The superintendent interview

Table 3: Teacher Interviews Conducted by District and School Level				
	Jonesboro	Nettleton	Westside	Total
Elementary	5	0	3	8
Intermediate/ Middle	3	4	4	11
High	4	5	3	12
Other	1	0	0	1
Total	13	10	10	33

protocol asked about all of the above as well as state support for the policy and collaboration between districts. As with the survey instruments, for the protocols we added additional questions about key considerations suggested by our conceptual framework, and in the case of the principal and superintendent interviews, added questions about principal evaluation as well.

In order to avoid the risk that potential teacher interview subjects would perceive the request as coercive, we contacted potential interview subjects directly, using lists of employees with job titles and email addresses from each district to identify potential participants rather than having administrators make the requests. We used a purposive sampling strategy to try to interview teachers at all instructional levels (elementary, middle, and high) in each district. We interviewed a total of 33 teachers, including teachers at every level for all districts (with the exception of Nettleton elementary teachers).

All superintendents and principals were contacted with interview requests. We conducted interviews in person with the administrators who responded to the interview requests. After initial data analysis, we decided that the number of interviews conducted did not provide a sufficiently broad perspective. Therefore, we sent out additional interview requests, followed up with phone calls, and conducted an additional combination of in-person and phone interviews with more principals.

Two superintendents and eight principals were interviewed for this study. Principals from each of the three school districts were interviewed. The superintendent from the third district was unavailable for an interview. We feel that these leaders offered an interesting look at how both superintendents of larger school districts, and smaller, more rural, districts led their school systems through the implementation of TESS. In our conversations, superintendents tended to talk about their relationships with principals (rather than with teachers) and principals tended to talk about their relationships with teachers (rather than superintendents).

As interviews were conducted, we worked with building administrators to schedule times to interview those teachers who had agreed to participate. As an additional measure to ensure that participation was truly voluntary and non-coercive, building administrators were told only the identities of those who agreed to participate, not of those who had been contacted and either declined or did not

respond. That way, there would be no negative repercussions for those who did not participate.

We also worked with building administrators to arrange for the interviews to be conducted in appropriate locations and circumstances. Interviews were conducted at the school sites where interview subjects worked. The location of the interviews was in the classrooms or offices of the interview subjects, or in common rooms or work areas. In all cases the interviews were conducted in private, so that in addition to occurring in a comfortable and familiar location for interview subjects, the subjects could be confident that their responses would not be overheard.

The interviewers asked for and obtained permission from the interview subjects to make audio recordings of the interviews in order to ensure that we had a verbatim account of the interviews. In addition, interviewers took written notes of key points and illuminating quotes. At the beginning of the interviews, interview subjects were given assurances of confidentiality (the recordings and notes were only for the use of the project team). Subjects were told that the one circumstance where others might have access to their words was in the event that we used a quote from their interview. Still, subjects were told their names and any other non-generic identifying information would be redacted from the quotes.

The interviews were semi-structured, with scripted questions alternating between open and close-ended questions, starting off with "icebreaker" type questions to develop rapport. For specific questions, follow up probes were pre-scripted, and in other cases follow up questions explored topics broached by unanticipated but relevant responses.

Qualitative Analysis

Our analysis of the qualitative data collected started with reviewing notes and listening to the recordings of interviews in order to refresh our memories and gain a greater familiarity with the content. This process was repeated in an effort to discern themes and patterns in the responses given in the interviews. As these themes and patterns were identified, we matched them up with our conceptual framework by developing a coding scheme for the more formal stage of our analytical process. Meanwhile, we compiled our notes and the recordings into written accounts of the interviews. Following this, each of us

wrote analytical memos summarizing the patterns in the responses of the interviews.

Once completed, we used a matrix embodying the categories and subcategories of our coding scheme to process the information from our interviews. For this stage of the process, we identified and articulated themes and patterns and compared these with our revised conceptual framework. For the theme of *policy implementation*, we found sub-themes of *capacity* and *will* as well as *local context*. For the theme of *teacher and principal evaluation*, we found sub-themes of *measurement* and *formative and summative* uses of evaluation data. For the theme of the *social organization of schools*, we found sub-themes of *trust, collaboration*, *power and authorit*y, and the *organization of time*. We developed a matrix with the major themes on the vertical axis and the subcategories of these themes on the horizontal axis. Each of us then began the process of reviewing the interviews we conducted and coded this data by placing the information contained therein into cells of this matrix. As we proceeded, we examined how the other partner was categorizing the information, and these comparisons formed the basis of an ongoing series of discussions about how to interpret concepts and where certain types of information was best placed, often deciding that some types of information ought to be in several different locations. Where thus far the information had been organized by interview subject, we then had the responses of the various interview subjects organized around major ideas in our conceptual framework and the themes and patterns that we identified over the course of the initial analysis of our interviews. This process made it clearer to us what points were made most frequently and with the greatest emphasis, as well as where unexpected findings were true of numerous individuals, rather than simply being isolated anomalies. Finally, having our findings organized around these central ideas allowed us to place our findings within the broader context of the academic literature using those ideas that appeared both in our data and in the literature as binding links.

Quantitative Analysis

Our quantitative analysis began with the effort to "clean" the data set compiled from the responses to our surveys. We inspected the sets of numbers and used frequency analysis to discover cases in which values were missing. For ordinal

variables, where the relative magnitude of the number has intrinsic meaning, we replaced missing values with the mean value of the set of valid responses so that our ability to perform the full range of relevant statistical analyses would not be compromised, and our results would not be skewed one way or another. Since the data were entered electronically, we did not look for reverse coding due to the possibility of transcription errors.

Once our data were cleaned, we calculated descriptive statistics. Most of our ordinal variables were Likert-type scale items, and the rest contained personal information like educational attainment, years in education, and hours per week spent on TESS. For these ordinal variables, since there were no distant outliers, we calculated means and standard deviations. This showed us where in each range the typical response fell, and how widely distributed the sets of responses were. We also noted minimum and maximum values. This gave us a summary account of how the set of survey respondents as a whole answered each question, and a snapshot of how the group as a whole leaned in their answers.

After generating descriptive statistics for responses to individual questions, we sought to create scales that would inform us about underlying constructs. Since many of the issues that arose in our interview data were different from the questions we initially sought to answer, we chose to do an exploratory factor analysis. The method we chose for this was principal component analysis. For this analysis we began by using SPSS software to see which items formed components with eigenvalues greater than one.

We then scrutinized the set of items in each of the components to see which ones seemed to measure coherent underlying constructs. After eliminating those that did not, we re-ran the principal component analysis, this time limiting the number of components generated to the number that could both be sensibly interpreted and which had eigenvalues greater than one. We eliminated generated components that had fewer than three items and those to which a coherent interpretation could not be given. For those items that appeared in more than one component, we allocated that question to the component for which its impact was greatest as measured by the absolute value of the question's impact within that component. We then took the items grouped together as components in the rotated component matrix and verified that each had a Cronbach's alpha above 0.7 (See table 4

below). Having confirmed this, we then computed new variables representing the constructs from the items grouped together in the components. Next, we ran linear regressions using these as independent variables to see how well they predicted whether respondents perceived an overall improvement in teaching and whether the time and other resources expended for TESS would have been better used for other purposes. We went through this process twice, once with the teacher survey data for questions relating to TESS and once with the administrator data for questions relating to TESS. (We attempted it a third time to develop scales for LEADS oriented questions, but none were generated).

Table 4: Scale Reliability Analysis

Teacher Scale*	Cronbach's Alpha	Administrator Scale*	Cronbach's Alpha
TESS Training	0.896	TESS Capacity	0.912
TESS Skill	0.930	Value of TESS	0.886
TESS Beneficial	0.932	Opportunity Cost of TESS	0.828
TESS Collaboration and Development	0.797	Desire for more training and support	0.75

*See appendices 6 and 7 for scale components Threshold for reliability: $\alpha>0.7$

For the teacher evaluation data, we sorted the spreadsheet for each district by evaluation category and counted how many teachers were in each evaluation category. We then calculated a total across the three districts for each category, and then calculated the percent of ratings in each category for each district and for the three districts overall.

Limitations

Our project faced a number of limitations. Some of these limitations were apparent from the outset of the study, while others emerged as the study proceeded. Of the limitations known in advance, the most significant is our inability to adopt or implement an experimental or quasi-experimental design. We would have

preferred to have been able to do this since the districts asked for us to look at the impact of the system, and experimental designs permit the generation of evidence of causality. Since we were neither able to conduct pre-testing nor able to assign subjects to control and experimental groups randomly, the external validity of our findings about the causal impact of the system are limited. Further limiting the external validity of our findings is the fact that we do not anticipate being able to find appropriate comparison groups. We have been unable to isolate the effects of the evaluation systems, and are therefore unable to make strong correlational, much less causal, claims.

Part of our inability to adopt an experimental or quasi-experimental design is due to another limitation, finite time and other resources. Had we been able to conduct our data collection periodically over the course of many months or years, our data would enable us to make inferences about how key circumstances and perceptions changed over time, and examine how these changes correlated with various aspects of TESS implementation. As it stands, we were only able to compare our findings with those of the previous study, without being able to say exactly *when* in the intervening period changes occurred or what elements of implementation were happening around that time that could account for the change. Leaving aside the time limitations, even within the scope of the handful of site visits we were able to conduct as part of our fieldwork, we would have preferred to have more time for additional interviews. We would have preferred to have teacher interviews from each level in each district, and in our administrator interviews we continued to hear new insights in our final interviews, suggesting that there were additional new insights to be had that we did not gather.

Another set of limitations emerged during the course of our study. One such limitation arose from our inability to obtain more than one year's worth of teacher evaluation data. We had hoped to compare ratings in a specific domain for the same teachers from one year to the next, in order to see if measured teacher quality changed, and if so, what patterns could be detected. Since we were only able to obtain one year's worth of evaluation data, we were unable to determine whether there were changes. Similar considerations limited our ability to draw any conclusions about impacts on student achievement. We were unable to obtain testing data of sufficient

granularity, or to perform an analysis of the impact on measured student achievement over time.

A third emergent limitation was the difficulty encountered in recruiting interview subjects, administrators in particular. In order to ensure that participation in our interviews was truly voluntary, we recruited subjects directly rather than relying on supervisors to arrange interviews; however, one result of making the choice as voluntary as we could was a lower rate of participation than we might otherwise have obtained. Another limitation was perhaps the most significant. Based on prior discussions about our methods with district administrators, we had anticipated response rates for our surveys to be near one hundred percent (well in excess of the standard threshold of seventy percent), but actual rates were much lower at fifty-six percent for teachers and forty-five percent for administrators. This lower than expected response rate limits our ability to extrapolate to our populations of interest and hence further limits the external validity of our work. In both cases, the low participation rate in our data collection could mean that the views gathered are from those who were unusually motivated to share their views (and this group might have different views than those of the population as a whole). On the other hand, notwithstanding our attempts to make participation truly voluntary, some might have felt pressure to participate and therefore to have been on some level coerced, even if unintentionally.

A final limitation concerns our ability to investigate all of the questions posed by the districts. The districts had requested information about a number of goals that would have required substantial additional resources to investigate, such as the impact on student achievement and views of parents. Given available resources, we limited the study's exploration of policy goals to those that could be answered through teacher and administrator surveys and interviews.

Section Four: Question 1: To What Extent has the Implementation of TESS Reflected Best Practices and Achieved Policymakers' Intended Goals?

Implementing a new system of teacher evaluation has been a complicated endeavor for the three Jonesboro school districts because the new system affects all stakeholders. While teachers may feel that they are the most affected by the TESS system, principals indicated that they have to shoulder tens of additional work hours weekly in order to keep up their role in completing the evaluations. Moreover, the state fully intends for the TESS evaluation system to have a positive effect on teaching, which should theoretically affect student performance. The policymakers who adopted the TESS system targeted specific outcomes for the new system. The state wanted to meet best practices and intended goals. This section will focus on how TESS has met the intended goals set forth by the State of Arkansas. An examination of best practices will come later in this paper in the "Discussion" section.

Intended Goals

While the scope of and time frame for this study did not allow us to uncover information about all eight of the state's goals, the following section will examine what the qualitative and quantitative data suggests about how well TESS has met the first five of the policymakers' intended goals in the Jonesboro, Westside, and Nettleton Consolidated school districts.

Providing school districts with a transparent and consistent teacher evaluation system that ensures effective teaching and promotes professional learning

The purposes of the TESS system have not always been realized in implementation. As discussed in the introduction, the TESS system utilizes fixed rubrics. Therefore, every principal at every school using TESS should be using the same instrument of teacher performance measurement. Elementary school teachers are evaluated with the same rubric as middle school and high school teachers. The actual metrics and mechanics of the measurement system are consistent within schools and districts, and across different districts. Furthermore, the TESS system is designed to be transparent. The TESS system requires principals and teachers to document their work and the findings that accompany teacher observations. As a result, all parties should have access to the notes from the observation and the artifacts

that the teacher has entered to document his or her work and growth. Principals have also had to complete rigorous training that included simulated observations. In order to complete the TESS training, principals had to score the simulated observations in a manner in which the principals' scores were consistent with the scores of the trainers. The training was rigorous enough that some principals had to repeat the test several times until they passed.

However, the way the system is designed is not always the way it has worked in each of the three school districts. In a variety of different ways, the effort to make the system disinterested for the sake of consistency runs aground in the limitations of human objectivity, both perceived and actual.

The superintendents interviewed for this project said that their districts made ample use of the online training resource made available by the state (IDEAS), conducted a book study of Charlotte Danielson's work about teacher evaluation and the framework embodied in TESS, and devoted time beyond that mandated by the state to training. The teachers and principals felt this training was effective. Fifty-eight percent of the teachers felt the TESS training they had received was "good" or "very good". Thirty-four percent of teachers felt that the training was "fair". Only nine percent of teachers felt the TESS training was "poor" or "very poor". One hundred percent of administrators "agreed" or "strongly agreed" that they had received adequate training to perform TESS duties.

However, TESS training is ongoing for teachers. While the bulk of training occurred early in the implementation process, teachers continue to receive refresher training each year. Since implementation, the refresher training now consumes much less time than the initial training. Teachers and survey data show that (in general) the amount of TESS training has decreased over the past two years. One teacher said, "The first year there was a lot [of training]; since then we haven't had any." Another teacher in a different school district said, "[there was] a lot during implementation training. I don't remember anything after that." As a result, evidence suggests that the initial TESS training for teachers was strong, but teachers have not received the subsequent refresher training as positively.

Despite the decline in the amount of training over the past two years, teachers overwhelmingly indicated in both the survey and the interviews that they understood the

TESS rubric and believed that the rubric's four domains (planning and preparation, classroom environment, instruction, and professional responsibilities) reflected effective teacher practice. Moreover, the survey shows that sixty-nine percent of the respondents either agreed or strongly agreed that they were adequately informed about the TESS evaluation system (only thirteen percent disagreed or strongly disagreed). The survey and interviews also suggest that teachers thought that the expectations for the TESS evaluation were communicated clearly and consistently.

Table 5: Teacher Perception: TESS Training and Understanding

	Strongly agree or agree	Uncertain	Disagree or strongly disagree
I am prepared to carry out the following aspects of TESS: - Collect and document artifacts for each of the four domains	72.43%	18.92%	8.65%
I am prepared to carry out the following aspects of TESS: - Complete paperwork for pre-and post-conferences	78.80%	14.95%	6.25%
I am prepared to carry out the following aspects of TESS: - Develop lesson plans that incorporate principles from the "Planning and Preparation" domain	82.26%	12.63%	5.11%
I am prepared to carry out the following aspects of TESS:- Implement instructional practices that affect principles from the "Instruction" domain	81.40%	14.02%	4.58%
I am prepared to carry out the following aspects of TESS: - Create a classroom environment that reflects principles from the "Classroom Environment" domain	90.44%	7.38%	2.19%
I am prepared to carry out the following aspects of TESS: - Choose and fulfill the duties under the "Professional Responsibilities" domain	87.50%	9.78%	2.72%

In sum, the teachers from the three school districts felt aware of the TESS system and generally believed that they had been properly trained to participate in the evaluations (See Appendix 3 - Teachers survey). Seventy-two percent of the teachers either agree or strongly agree that they are prepared to collect and document artifacts for each of the four TESS domains. Seventy-nine percent of teachers feel prepared to complete the paperwork for pre- and post-conferences. Teachers overwhelming agreed or strongly agreed that they were able to develop lesson plans that incorporate the principles from the TESS "Planning and Preparation" domain (82%), implement instructional practices from the "Instruction" domain (81%), create a

classroom environment that reflects the "Classroom Environment" domain (90%), and choose and fulfill the duties under the "Professional Responsibilities" domain (88%).

The administrators' and principals' survey responses also indicated that they felt equally prepared to administer the TESS evaluations. The administrators unanimously felt that they had received adequate training to complete the evaluations and that the training was of adequate quality. The surveyed principals (N=25) all agreed or strongly agreed that their training was adequate. Moreover, thirty-two percent of the principals felt that their training was "very good"; fifty-two percent of the principals thought the training was "good". Though sixteen percent of the principals rated the training as "fair", not a single principal rated the training as "poor" or "very poor".

The majority of teachers and administrators also felt that the rubric did an accurate job of cataloguing the skills teachers needed to have in order to be effective with general pedagogy. One principal said, "I think the rubric is pretty spot on. It's specific enough that the steps are incremental." A middle school teacher in another district also shared her support for the rubric: "I believe the rubric [captures effective teaching practice]. It fully explains what we're supposed to look like." Another middle school teacher said, "Absolutely, the rubric covers everything in detail. I love the rubric."

Table 6: Administrator Perception of TESS Training Quality

	Very Good	Good	Fair
Administrators from 3 Districts	32%	52%	16%

That being said, some teachers felt that one single rubric does not always fit for all teaching situations. One elementary teacher said, "One rubric for all grade levels is insane. Fifth graders are not the same as second graders. Some ideas aren't developmentally appropriate." This opinion was shared by a middle school teacher: "There's a lot of generalized information and categories that need more specific rubrics and artifacts that are more appropriate for particular grade levels." One superintendent agreed, offering as an example that this superintendent thought it was unfair to SPED teachers to use the same rubric as teachers in conventional settings.

Despite feeling well prepared to receive and give the evaluations, some

teachers and administrators disagree about whether or not the evaluations resulting from the TESS system are consistent. Though sixty-three percent of the surveyed teachers agreed or strongly agreed that there was a great deal of trust between administrators and teachers in each school and most teachers and administrators thought that the rubrics demonstrated effective teaching, some of the interviewed teachers and administrators expressed concerns involving bias, evaluator subjectivity, and asked questions about the subjectivity of the rubric itself. One interviewed teacher said, "The rubric is very important. It takes the subjectivity of the ratings out." However, another teacher added, "The rubric is not as objective as it wants to be. Different districts and different individuals will implement it differently. I know this from discussions with teachers in other schools and districts. Also, there will always be favorites." Even some principals were concerned about the issue of bias: "Bias is always hard. We do a good job of being as unbiased as possible. Sometimes, we'll re-assign observations if we worry about bias."

Some teachers and principals were also concerned about inter-rater reliability, even if bias is not a concern. One principal stated, "The difference between a 3 and 4 is subjective." This idea was echoed by a high school teacher, "The rubric is good, but there's not enough of a distinction between a 3 and 4." Another high school teacher said, "The rubric has redundant elements and fuzzy distinctions between levels." Another teacher worried that the lack of frequent observations might not lead to a complete picture of teacher practice: "With so few observations, a fluke can lead to a lower rating than would be representative. Process can't capture some things. Ratings can also be distorted by the level of the teacher's comfort with the evaluator."

Other teachers are worried that the observations and rubrics do not always match the realities of a daily lesson. Many of the teachers lamented the difficulty of designing a lesson that would be able to demonstrate the skills in each of the domains that the teachers needed for a desired score. One teacher said, "I will say this: If an evaluator came in just one time, they wouldn't see everything on the list." Other teachers worry about putting on a "dog and pony show" in order to score well on observations: "The observed lesson doesn't represent actual daily practice. Students and colleagues should have a voice concerning teacher

performance." Another teacher worried that the documentation of the observations did not always represent what actually happened during the lesson: "[Principals] will sometimes put down scores for something they didn't observe." This discrepancy in ratings, or in the perception of the quality of the rating, demonstrates a trade-off between the accuracy of the rating and the time it takes the principals to complete the evaluation.

Provide feedback and a support system that will encourage teachers to improve their knowledge and instructional skills in order to improve student learning

Several elements of TESS are intended to serve the function of feedback and support. This begins with the teacher and observer both reflecting on the lesson, and using those reflections as part of the basis for a post-observation conference. Examining these elements begins to highlight a key tension in our findings. A strong consensus existed among administrators that these were useful, but teacher opinion was divided. The teacher survey results showed that while teachers agreed that they were able to complete paperwork for the conference (which included the reflection), the median response to the question of whether they could improve as a result of feedback was "uncertain". The mean response was between "agree" and "uncertain", indicating that some teachers did think they could improve from feedback. This same gap between the perceptions of administrators and those of teachers showed up in our interview data as well. While administrators generally thought that the process of reflection and conferencing was among the most valuable parts of the process for teacher growth, teacher opinion was more divided. Some teachers agreed that the reflection and conferencing were among the most valuable parts of the process; others did not find them to be useful. A number of teachers said that they had become more reflective in their teaching, with one singling out her planning as having become better differentiated as a result. Another said he had become more mindful of his role in student outcomes. A different teacher said she had become more mindful in everything she did and attributed some of this to the TESS process, but this teacher was also quick to point out that the key factor was her internal desire to improve.

The results of the post-conference are to be incorporated into each teacher's Professional Growth Plan (PGP), and

professional development activities for each teacher are supposed to be aligned with the needs identified in the evaluation and the goals articulated in the PGP. Teachers also disagreed about whether or not professional development activities were linked to their evaluations. Here again, we see a gap between administrator and teacher perceptions. The median administrator response to questions regarding their ability to identify professional development and support for individual teachers based on their ratings, for their schools based on effectiveness data, and their access to resources to recommend for growth was "agree", although the mean response in two of these three cases fell between "agree" and "uncertain", indicating some uncertainty and/or disagreement on these points (in the third case the mean was equal to the median). For teachers, the median response to whether their evaluation score informed their professional development and whether they had access to adequate support in their areas of refinement was "uncertain". Even though the mean responses for these questions fell between "agree" and "uncertain" and therefore indicate that some teachers perceive an alignment and adequacy of professional development, the considerable ambiguity remaining about these points is striking. (See table 7).

Table 7: Administrator and Teacher Perceptions About the Alignment of PD to Evaluation Data and The Availability of Appropriate Resources

	Administrators ("At your school")	Teachers ("Your own")	Difference	Sig. (two tailed)
PD Aligned to Evaluation Results*	2	2.82	0.82	<0.0001
Adequate Access to Developmental Resources*	2.42	2.64	0.23	0.2091

*= Paraphrases of differently worded corresponding questions
1="Strongly Agree", 5="Strongly Disagree"

One teacher said, "PD sessions are not always related to my specific PGP goals." Although the interviews and surveys suggest that teachers are engaging in PD activities related to TESS, it is not clear that

the activities are always related to the individual teacher's developmental needs. Other teachers are not convinced that the PD activities are particularly useful, even if they are related to professional goals. In an interview, another teacher said, "We're expected to work towards our PGPs (Professional Growth Plans), which are useful for younger teachers but are a stretch for veterans." Also, some schools have identified PGPs for their whole staff to pursue. While this may unify the staff in working through the TESS process, individual teachers' areas of refinement are not necessarily being addressed.

It is also unclear what scope teachers have to align their professional development to their needs by selecting that PD themselves. While more teachers (forty-three percent) agreed that TESS results helped them choose the professional development activities in which they participated, a third of the teachers were uncertain about the effect of TESS on their PD choices, and twenty-four percent disagreed.

Besides providing support for growth, TESS is designed to provide additional incentives for growth in the form of the ratings levels. Our evidence indicates that the effect of the system itself was to incentivize increasing the score by means other than improving the quality of their teaching. During interviews, when teachers talked about improving their skills, comments were usually made in the context of improving scores on the TESS evaluation rather than improving the underlying skills. Moreover, while some teachers talked about an innate desire to become better teachers, those comments were never connected to the TESS model.

A final means by which TESS could have facilitated teacher growth is through peer learning. The majority of the interviewed teachers discussed collaboration as either a) something in which they were already engaged, or b) a means to improve performance on the TESS evaluations. One teacher said, "Teachers work together informally to figure out the system (in part due to a lack of timely answers from the state). The need to [improve] has brought teachers closer… but no increase since… TESS [was implemented]". Another high school teacher opined, "Teachers work together to prepare for evaluation, and are always talking about student achievement, but no more so than before TESS." Another teacher succinctly said, "Collaboration for sake of documentation has increased, but not for [teacher] quality." TESS does not

appear to have increased teacher growth through peer learning.

Provide a basis for making teacher employment decisions

Despite the fact that a teacher's failure to move out of track 3 should start the process for dismissal or the non-renewal of contract, there are mixed feelings from superintendents, principals, and teachers about how meaningfully TESS affects staffing decisions. In the administrator survey, administrators tended to agree that TESS results should be used for hiring and promotion. This is important because it demonstrates the summative mindset of the principals conducting the evaluations, but not necessarily for the purpose of firing teachers. For example, one principal also said that he did not see the TESS system as a means for replacing poorly performing teachers. One principal said, "For the most part, if there's an issue, we will know about it before the TESS system. It's really hard to get fired in Arkansas. I don't see TESS as a tool to initiate termination."

Quantitative data suggest that the issue will be moot for quite some time. Our analysis of teacher ratings for 2014-2015 indicate that not a single teacher in any of the districts earned the "unsatisfactory" rating needed for identification for Track 3, and only eight percent were rated "basic" (see table 8). An influential report calling attention to the low incidence of low ratings in binary ratings systems reported that Jonesboro Schools had rated only 0.3% of teachers with the lowest ratings between school years 2003-2004 and 2007-2008 (New Teacher Project, 2009). TESS has not changed this by simply adding more ratings levels. Since a major justification for the new evaluation systems was to increase the rate of identification of unsatisfactory teachers, this unchanged rate represents a serious failure. One could instead make the comparison to teachers rated either "unsatisfactory" or "basic" under TESS, and here we do see eight percent of teachers rated "basic" (leaving 92% rated in the upper half of the ratings scale). However, a closer look at these numbers shows that 5/6ths of the basic ratings come from just one district, Westside. Jonesboro rated 5.8% of teachers as "basic", and Nettleton rated 0.4% of its teachers as "basic". Jonesboro does rate more of its teachers in the lower categories than it did before, while Nettleton's results do not meaningfully differ from the historical ones from Jonesboro. Westside rated 27.3% of its teachers "basic", and this is consistent with

the Westside superintendent's emphasis with his evaluators that he wants them to be candidly critical with their evaluations. However, Westside teachers also have more negative perceptions of the system as a whole, and this could be related to those teachers having higher rates of critical feedback.

Table 8: TESS Evaluation Levels by District

	Nettleton	Jonesboro	Westside	Overall
Unsatisfactory	0	0	0	0
Basic	1	5	30	36
Proficient	167	78	80	325
Distinguished	66	2	0	68
	234	85	110	429

Perhaps resulting from this disconnect, teachers also had mixed opinions about how TESS results would be used. Some teachers wanted bad teachers to be fired, and some teachers worried about their TESS scores causing them to lose their jobs. Other teachers were not particularly worried about TESS being used as a tool for firing weak teachers. One teacher said, "[TESS] doesn't even get rid of teachers who are bad." Another teacher indicated that she expected some principals would use TESS as a tool to get rid of people they did not like. However, she qualified her remark to say that teachers would have to perform quite badly in order to earn an unsatisfactory rating. Another teacher merely said, "The threat of job loss is implied, but not stated."

With respect to the issue of possible positive consequences, the survey indicated that the majority of principals thought that teacher evaluations should be used for decisions involving hiring and promotion. The superintendents agreed. Additionally, both superintendents interviewed see the processes as successfully promoting better conversations among principals, teachers, and students about teaching and learning, and increasing focus on improving student achievement. Both also use TESS as part of an effort to identify and cultivate future leaders for the district.

However, during the principal interviews, principals were mixed about TESS actually being used as a tool to identify strong teachers for leadership opportunities. One principal specifically

said that she asked teachers who scored "5s" to lead PD sessions in their areas of strength. Another principal said that TESS scores were a factor in a teacher's possible promotion. Other principals unequivocally said that TESS scores were not part of their decision making process (despite the fact that surveys indicated that principals largely viewed TESS scores as summative). None of the interviewed teachers mentioned that they viewed TESS as an opportunity to promote their teaching capabilities for future employment opportunities or additional leadership roles.

As a result, there seems to be a disconnect between whether or not TESS is viewed as a tool for summative or formative evaluations. Of the principals who look at TESS as summative, there is disagreement about how TESS results should be used. Some principals use TESS results for hiring and promotion, while others do not. Other principals have mixed views about using TESS as a tool to remove poorly performing teachers from their posts. In interviews, teachers acknowledged that TESS could be used to fire teachers, but no teacher ever mentioned that it could be used as a tool for hiring or promotion.

Provide an integrated system that links evaluations procedures with curricular standards, professional development activities, targeted support and human capital decisions

Although the previous section discussed our findings about the use of the TESS system to make human capital decisions, there is more ambiguity about how teachers perceive TESS's link to curricular standards, professional development activities, and targeted support. While forty-three percent of the teachers surveyed agreed or strongly agreed that TESS fits well with other school/district initiatives, fifty-two percent of the teachers surveyed were uncertain or disagreed that the TESS system met their district's initiatives. Five percent of teachers strongly disagreed.

One teacher wrote on the survey, "The evaluation system helped me to develop a program that addressed Common Core expectations, and allowed me to measure how well I was doing this." Another teacher tied TESS to Common Core by writing, "The biggest relation to Common Core, that I see, is the correlation to student led activities." Another teacher wrote, "I think TESS fits fairly well with CCS." However,

not all teachers agree on the coherence between TESS and Common Core. The write-in comments are fairly evenly split between negative and positive about the relationship.

According to the survey, principals are less certain about whether or not the TESS system is consistent with other district improvement initiatives. Only thirty-eight percent of the principals agreed that the TESS system supported district improvement initiatives. The remaining sixty-two percent of principals either were uncertain or disagreed.

Encourage highly effective teachers to undertake challenging assignments

Again, despite the fact that the principals' survey indicated that most principals saw TESS as a means of identifying teachers for additional responsibilities or promotions, only a few interviewed principals used the TESS system to identify teachers for more challenging assignments or career paths. One teacher who participated in an interview was a lead teacher who helped train other teachers for TESS. However, he applied and interviewed for the position, rather than being identified and encouraged to apply because of his TESS scores. It is unknown whether or not his TESS scores were part of the principal's decision to hire him to the leadership position. Again, no other teacher mentioned TESS as a pathway to promotion or additional leadership roles.

Question 1a: How Have Teacher and Principal Perceptions of TESS Implementation Changed Over Time?

Teachers

We used T-Tests to compare attitudes about the implementation of the

TESS system between the previous study (Ashby et al., 2014) and the current project, and uncovered five major findings:

1. There have been meaningful increases in the understanding of expectations and ability to describe the processes and procedures of the TESS system.
2. Teachers are closer to agreeing that feedback informs professional development and that they have access to support for growth in identified areas.
3. Teachers disagree that TESS helps foster better and more professional conversations. However, the disagreement is less strong than it was two years ago.
4. There has been a decrease in the belief of the potential for feedback to improve teaching.
5. There has been a slight improvement in the teachers' views that the system interferes with their ability to perform other duties; however, teachers still do not believe that TESS is a worthwhile use of their time.

These results show that while teachers understand the rubric, and are comfortable with the rubric and the training they have received on TESS, they largely have an ambivalent to negative opinion about the system in general. This disconnect between understanding and comfort, yet not supporting TESS, raises larger questions about professionalism and the use of resources in the districts and Arkansas as a whole. The fact that teachers are required to spend time and energy on a process they believe they understand and yet do not find worthwhile constitutes an infringement of professional autonomy, and by establishing an accountability system operated by those outside of the profession, it undermines the internal accountability to fellow practitioners characteristic of a profession.

As we compared the responses to questions about teacher understanding of the TESS system from the 2014 study to the current study, the change in attitudes and perceptions in most questions demonstrated positive growth. This growth indicates that through training, practice, and increased familiarity, teachers have become more comfortable with the TESS rubric and the tasks teachers need to complete as they are being evaluated. The information covered in the questions that yielded statistically significant increases from the 2014 study to the current one included the understanding

of the expectations of each of the TESS domains and sub-domains, as well as the ability to accurately describe the processes and procedures associated with TESS evaluations.

Teachers remain uncertain that TESS feedback informs professional development choices, and that they have adequate support to improve in areas of refinement identified in their evaluations. However, teachers' responses have moved from being uncertain to closer to "agree". In 2014, teachers' average score for the survey item, "Feedback from my teacher evaluation informs the professional activities in which I participate" was a 3.10. In 2016, the average was a 2.81, which shows a move closer to agreeing with the statement. Similarly, in 2014, teachers' perception of the support they had to improve areas of refinement identified by TESS evaluations was a 2.77. In 2016, the scores moved to a 2.64, which, again, shows a move toward agreement.

Teachers are also uncertain that TESS improves the quality and frequency of professional conversations with colleagues. However, the disagreement in 2016 is less than the disagreement in 2014. In fact, the average score on the question has moved from 3.36 to 3.19 (which is much closer to the "uncertain" score of 3.0, but still leans towards "disagree").

In an apparent contradiction, even as teachers' perceptions improved (while remaining ambivalent to negative) about whether their professional development was informed by feedback, whether they had access to adequate support in identified areas, and whether the quality and frequency of professional conversations had increased, teacher perceptions deteriorated about whether feedback could help improve their teaching. This exception is problematic, but also needs to be somewhat discounted because of the wording of the question between the two studies. In the original study, the belief statement was worded, "Overall, I think the new teacher evaluation system *will have* a positive impact on my own teaching practice" (emphasis added). The 2016 survey worded the statement as, "I believe the feedback given to me though the TESS process *has helped* improve my teaching" (emphasis added). The difference in the wording between the two questions is significant for a couple of reasons: 1) the first question was asked before teachers were actually participating in the TESS system, and 2) the 2016 statement looks at the teacher's

perception of TESS feedback in retrospect rather than as an anticipation of a perception. However much impact this shift in perspective might have had, teachers' attitudes and perceptions moved from an average score of 2.69 (between "agree" and "unsure") to 2.91, which is much closer to "unsure". If perceptions have, in fact, deteriorated about the potential for feedback to help improve teaching, it might indicate that direct feedback from the evaluator is not helpful, but collaboration and professional development might be.

There has been a statistically significant reduction in the perception that TESS interferes with the ability of teachers to fulfill their other responsibilities. Nonetheless, the perception remains that it does interfere (There was a similar reduction in the still critical perception about whether the system consumes time and other resources better used elsewhere, but this change did not quite meet standard thresholds of significance). To place this in context, it is helpful to observe two additional instances in which we did not see statistically significant changes. There were slight increases in the perceptions about whether the system would have an overall positive impact on teaching, and whether it would improve student achievement, but these were not statistically significant. More importantly, the mean response fell between uncertainty and the belief that there would not be a positive impact. Put another way, sixty-six percent of teachers are unsure, disagree, or strongly disagree that TESS has had a positive effect on their teaching practice. Similarly, 71.7% have that range of views about the prospective impact of TESS on student achievement. Broadly speaking, perceptions of TESS have improved the most with respect to teacher's understanding of TESS and of their skills to do what is required. Perceptions about the benefits of TESS remain largely negative even in areas where there were small improvements.

Table 9: Teacher Perceptions: TESS and Professional Development

	2014 Mean	2016 Mean	Mean Difference (2014-2016)	Sig. (2-tailed)
I understand what is expected of me in the domains and sub-	2.18	2.01	0.17	0.007

domains of the rubric				
I can accurately describe to others the processes and procedures by which I will be evaluated	2.69	2.43	0.26	<0.000
Feedback from my teacher evaluation informs the professional development activities in which I participate	3.1	2.82	0.28	<0.000
I have access to adequate support to improve areas of refinement identified in my teacher evaluations.	2.78	2.64	0.14	0.043
The quality and frequency of professional conversations with colleagues has increased under the new system	3.36	3.19	0.17	0.025
I believe the feedback given to me through the TESS process has helped improve my teaching	2.69	2.91	-0.22	0.005
The new teacher evaluation system consumes time and other resources that could be better spent elsewhere	1.87	2.02	-0.15	0.0627
I believe that the obligations of TESS interfere with my ability to carry out other teaching responsibilities	2.14	2.33	-0.19	0.0251
Overall, I think the new teacher evaluation system has had a positive impact on my own teaching practice	3.2	3.1	0.1	0.238
Overall, I think the new teacher evaluation system has had a positive impact on student achievement in my school	3.28	3.19	0.09	0.245
1= 'Strongly agree', 5= 'Strongly disagree'				

Administrators

In general, principals and superintendents have a much more positive sense of the TESS system when compared to the views of teachers. However, like teachers, administrators also feel that the time associated with conducting TESS evaluations takes away from aspects of their jobs. The major findings we have discovered in comparing the 2014 and 2016 administrator survey results about the implementation of TESS are:

1. The administrators see themselves as more able to accurately rate teachers, assess the suitability of artifacts, conduct teacher conferences, and complete TESS-related paperwork.
2. Administrators are more confident in their ability to coach teachers in the four domains of TESS and provide substantive feedback.
3. Administrators feel less need for more training in accurately assessing the suitability of artifacts and having critical conversations.

Unlike the teacher surveys, administrator responses were much more likely to be "agree" or "disagree" with the item on the survey, rather than having most answers average into the "uncertain" range. Administrator surveys also reflected more positive attitudes than the teacher surveys.

In fact, looking between the two surveys, principals' (with few exceptions) beliefs in their abilities and in the TESS system improved. Unlike the teacher survey, where teachers generally moved from "uncertain" averages to slightly less "uncertain" averages, the principals generally moved between "agreeing" to "strongly agreeing".

Administrators also grew in their beliefs (which were already positive) that they were able to provide substantive feedback and coach teachers in the four domains. It is striking that administrators had significant improvements in their perceptions of the value of their feedback even as teachers had significant declines in their perceptions of the value of that same feedback.

Table 10: Administrator Perceptions of the Quality of Their Feedback

	2014 Mean	2016 Mean	Mean Difference (2014-2016)	Sig. (2-tailed)
Provide Substantive Feedback	2.14	1.67	0.47	.002
Coach Teachers in the Four Domains	2.1	1.71	0.39	0.007

Also, between the 2014 and 2016 surveys, administrators became less likely to feel the need for more in-depth TESS-related training. In the following four categories: 1) accurately rating teachers using the TESS rubric, 2) conducting teacher conferences, 3) having critical conversations with teachers regarding performance, and 4) identifying professional development and support for specific individuals based on their evaluation results, the principals' desire for training shifted from some wanting "refresher training" to more feeling that "no further training [is] needed."

Question 1b: How Have the Principals and Superintendents Worked Together to Improve TESS Implementation and Promote Desired Outcomes?

Both of the interviewed superintendents raised concerns about changes in state policy arising from a change in administrations. One superintendent described the situation at the state department of education as “chaos”, with little to no guidance or support for

implementation, seemingly arbitrary changes of deadlines at times, and unpredictable bugs in state level online systems. This superintendent said that, as a result, it is unclear what the state's expectations are, but that his/her district holds high expectations of itself. The other superintendent cited the lack of clarity about the amount of documentation required as a particular concern, and pointed out inequities that this independent interpretation creates between districts because some do more and some less. This superintendent also lamented the amount of paperwork needed to dismiss an ineffective teacher, and expects that one of the first dismissals to arise from TESS will be challenged in court under the state's fair dismissal law, notwithstanding the claim by TESS proponents that the dismissal is consistent with this law.

As a result, the superintendents have had to lead their districts through the implementation of TESS without consistent help from the state. Both of the interviewed superintendents had positive views of the training their districts had conducted for TESS, noting that in each case the approach of their districts either went beyond that required by the state, differed from the approach recommended by the state, or both. In both cases, the superintendents showed high levels of motivation to adopt the system, with one having been a pilot district and the other having implemented the system in advance of the school year for which implementation was mandatory. A middle school principal from one of these districts said, "We were ahead of the game. We were proactive." Furthermore, the superintendents had to change their approach to TESS as the state changed the TESS model and legislation.

The superintendent and principals of the Jonesboro school district changed the state training models by redesigning the training so that it was done at building sites and was more interactive and learner-centered. Along with the change in training, they broke the training down into smaller units and had month-by-month benchmarks. They had a component that drew not only on the aforementioned book study and IDEAS website, but also on individualized support and guidance on all aspects of the implementation process provided by a consultant from the Danielson group. This consultant, brought in by the district as a supplement to the training as envisioned by the state, conducts ongoing monthly professional development and individual conferences with teachers, both as

guidance for the teachers and as modeling for evaluators. One principal said, "Honestly, we brought in Shirley Hall with the Danielson group. It was one of the best PD sessions I've ever had. Principals learned first, and then we brought it out to the buildings. We actually watched classes and then immediately scored them." The superintendent also allowed principals to revamp daily schedules and create new PD directives in order to facilitate the implementation of TESS. One high school principal said, "Late starts, PLCs, and RTIs all go toward TESS." Another Jonesboro principal said, "We [got] rid of BloomBoard glitches and made changes in the software."

Multiple interviews relayed that the superintendent of the Westside district supplemented the training provided by the state by paying for additional resources and having the principals conduct training at the building level with their staffs. He said superintendents became more focused on principal and teacher quality and made more of an effort to know all staff personally. This training was passed from the superintendent to the principals to lead and oversee the training for their teachers.

However, the superintendent acknowledged, "Teachers vary in how much they take advantage of this." The district did a three-hour face to face with teachers, and principal training with co-op for conferencing, coaching, and evaluation. Education service cooperatives are designed to assist member school districts and communities with the development of more effective uses of shared resources and to provide shared services promoting student achievement. The principals and assistant principals work with the co-op to train as evaluators, as well as in coaching, pre-conferencing, and evaluation, and these administrators have ongoing development in the form of continued access to co-op training and feedback from the superintendent. Currently, the Westside school district wants to give teachers more of an opportunity to observe other teachers. The superintendent also observed that since the state now requires teacher education programs to complete three hours of training on TESS, new teachers should enter the profession more familiar with the system. Looking back, he would have changed how his district provided TESS training by coordinating the pace of the book study better, and by giving teachers experience in scoring videos of other teachers. These changes in training would give them more examples and non-examples

of particular qualities of effective teaching as described by the rubric, and also give them a better appreciation of how difficult it is to generate accurate scores. The superintendent also communicated a sense that the degree of training provided to staff varied widely from district to district: he said that while some had done the recommended book study and others had done additional training, he also thought that some had done only the mandatory three hours.

Both districts perceived implementing TESS early as a distinct advantage. A superintendent in the Jonesboro district said that she wanted to give her principals and teachers the time to "do it right". She mentioned the book study specifically, but as an example of the benefits of not being rushed. She also described this careful approach as being a departure from the state recommendations. Jonesboro reported an increased use of technology to coordinate and facilitate implementation, especially the use of Google Classroom. They also relied heavily on implementation feedback and guidance from the Danielson Group consultant. The Jonesboro district based individual and building level professional development on the evaluation results, and invested heavily in this training. These decisions are consistent with Fixsen, et al."S2005 findings that policy implementation requiring staff training should involve the presentation of the information, but also the demonstration of techniques with practice and feedback.

The Westside superintendent said he places great emphasis on forming and developing personal relationships with his principals, teachers, and other staff so that he can monitor what is really happening in classrooms, but grants that this is only feasible due to the small size of his district. He judged the efficacy of implementation in terms of how well it allowed staff to meet the expectations laid down by the state, and viewed the work with TESS as part of a broader effort to prepare staff well in advance in order to enable adjustments to coming changes in national, state, and local policies and expectations.

Section Five: Question 2: To What Extent has Implementation of LEADS Reflected Best Practices and Achieved Intended Goals?

Introduced in 2014, the LEADS evaluation system is still finding its footing in the daily practices of principals and assistant principals. Both superintendents and principals mentioned in their interviews that although LEADS implementation mirrored that of TESS, it had received much less emphasis from the state. Beyond this lesser emphasis, according to the principals

interviewed and surveyed for this study, LEADS is generally seen to be of lesser importance due to the principals' obligations with TESS and other duties. Also, when principals and superintendents were interviewed about LEADS, there is less familiarity with the system and less enthusiasm than that with which they discuss TESS. As above, this section will concentrate on the achievement of goals, and best practices will be addressed in the discussion section.

Intended Goals

The scope of this project did not cover how the use of LEADS at the local level informed practice at Arkansas colleges and universities. However, we were able to gather school leader perceptions about the other three goals.

Provide a cohesive process that includes clear expectations to guide building- or district-level leader preparation, induction, and continued professional development in Arkansas school districts, open-enrollment public charter schools, and the Arkansas Correctional School

LEADS provides a detailed process that also provides generally clear expectations for the continued professional development of school leaders. However, the implementation of LEADS has not been particularly smooth or effective due largely to it having been given less emphasis and attention than TESS. Principals almost universally claimed that the additional responsibilities added by the implementation of LEADS were considered of secondary importance to the principals' other responsibilities. One principal said, "[LEADS is] a lot more hours and expectations on myself. I'm always juggling. I have to do too much to feel good at any one thing." The frustration with finding time to fully integrate LEADS into the principals' daily responsibilities was echoed by another principal: "I don't have time to focus on LEADS. I need to take care of the teachers in my building." One principal said that LEADS implementation has improved at her individual school, "But only because it has been made a priority." The principal added that, "Some administrators might just see [LEADS] as a burden."

While there has been professional development surrounding LEADS, principals and educators feel that the training has been uneven and less specific than the training associated with TESS. One administrator said, "The state should give

better clarity with LEADS about how much documentation is required… [LEADS] is still trying to decide how to use evaluations to inform principal PD." Another principal feels that BloomBoard for LEADS is hard to use, and as a result, documentation of administrator growth is more difficult to catalog.

Our interviews suggest the inconsistency in the use of the process stems from a lack of attention to appropriate organization and policy context associated with the implementation of LEADS. With the timeframe of TESS implementation overlapping with that of LEADS, one or the other almost inevitably had to be given priority, and the state's greater emphasis on TESS tipped the scale away from LEADS. Had there been excess capacity on the local level, perhaps both might have been able to have been implemented with fidelity. However, at least in Jonesboro, a principal's day was already rather full. Several principals reported that the implementation of LEADS seems as if it is an afterthought; they felt LEADS was quickly organized attempt to create a tool that can evaluate administrators in a similar manner to the way that TESS evaluates teachers. LEADS training has not been as thorough as TESS training, and LEADS has not been as blended into the lives of administrators as TESS has been into the lives of teachers and administrators.

Guide and sustain excellent leadership performance that ensures the improvement of teaching and learning

LEADS has the potential to guide and sustain leadership performance in a way that promotes improved teaching and learning, but current limits on organizational capacity make this goal unlikely to be realized. One superintendent said, "LEADS has been implemented as fully as possible given staffing constraints. We would like to have about four more trained evaluators to give feedback and dig deeper into what is happening." A principal echoed this sentiment when he said, "For LEADS, the superintendent is not in the building enough to form evidence based judgments." The principal feels that he sees enough of his assistant principal to make an effective evaluation, but does not think he is particularly effective in helping the assistant principal realize his goals.

One principal also expressed frustration with the fact that the superintendents did not have to go through the same level of training for LEADS that the principals did for TESS. Still bristling

over having to re-take the TESS certification, the principal suggested that there was a distortion in priorities when his evaluator was not as qualified to make evaluations about him as he was for his teachers. As a result, the situation suggests the question: if superintendents are less qualified than principals to make certain that school practices are ensuring the improvement of teaching and learning, how are districts to know that principals are truly effective?

According to survey data, there is broad disagreement about how difficult it is for administrators' evaluators to gather sufficient and accurate data about performances. Only thirty-three percent of principals thought it was easy for evaluators to gather sufficient and accurate data. While only twenty-one percent thought gathering accurate and sufficient data was difficult, the fact that forty-six percent of principals were neutral in their opinion suggests that there is wide uncertainty about the sufficiency and accuracy of the data collected in evaluations.

There is also a wide difference of opinion in whether or not evaluators give appropriate weight to each data point collected during evaluations. Twenty-six percent of administrators believe that their evaluators weigh data appropriately, and seventeen percent of administrators feel that their evaluators make errors in weighing data. Fifty-seven percent of administrators are uncertain about how appropriately evaluators weigh data in their evaluations. That means that seventy-four percent of administrators either think that their evaluations are erroneous or are not certain that their evaluations are conducted with fidelity.

The fact that such a high percentage of administrators either feel, or are uncertain, that their evaluators are unable to gather sufficient and accurate data and are uncertain (or feel that errors are made) in weighing their evaluations correctly, suggests that administrators are concerned about the quality of their evaluations. Also, the principals surveyed are inconsistent in the amount of time they spend on and the rigor with which they conduct LEADS.

Provide a process that includes instruments to be used by reflective practitioners to promote their professional growth

While LEADS certainly has a prescribed rubric, it does not appear that principals are always able to use the LEADS rubric to promote professional growth. The principals who participated in the survey had mixed feelings about the effectiveness of

LEADS. In the survey, fifty-two percent of the surveyed principals agreed that the LEADS evaluation system had a positive impact on administrative practice. Thirty-six percent of the surveyed principals had neutral opinions about the system, and twelve percent of the principals disagreed that the LEADS system positively affected administrative practice. Evidence from interviews suggests that much of the reason why principals struggle with LEADS is because their other professional duties spread them too thin. The principals often claimed in the survey and interviews that teachers and students were their priorities, not administrative evaluations.

Other principals are more optimistic about using LEADS; however, some freely admitted that they have not yet looked at the instrument. One principal said, "LEADS will help my development when I have a chance to read through it. Reflection really helps." The principal also said that he was excited about "giving better feedback to teachers and [learning] more about supports that exist." The principal plans to use observation data to set goals.

LEADS does provide a clear set of expectations and a reasonably well articulated process for working towards the stated goals of the system. Nonetheless, it does not appear that the Jonesboro districts had sufficient capacity to implement LEADS as intended, especially during the concurrent implementation of TESS. Beyond a lack of time, the lower priority given to LEADS also suggests that it was not presented in a way that convinced practitioners that it offered meaningful help in worthwhile areas.

Question 2a: What Role has LEADS Played in Influencing how Principals Implement TESS?

One way in which LEADS has affected TESS implementation is creating empathy. Several principals mentioned that being evaluated in this way made them more empathetic to what teachers go through with TESS. In a material sense though, surveys and interviews suggest that LEADS has virtually zero impact on how superintendents and principals implement TESS. Instead, TESS impacts how principals implement LEADS. This seems reasonable because TESS was piloted or

introduced early by two of the districts, and those principals were already familiar with TESS before they started using the LEADS system. In fact, one interviewed principal called LEADS, "TESS Lite". Also, principals indicated that since the models are so similar in structure, there was less training and interest surrounding the implementation of LEADS. One principal said, "I have not had as much PD over LEADS." Another principal was loath to demonstrate his excitement for LEADS as he said, "LEADS is fine. It's just kinda TESS." Other principals were more specific in their evaluation of the importance of TESS and LEADS: "TESS drives practice. LEADS is in the background. We have to take care of the teachers first. I haven't looked at the LEADS rubric all year."

Again, not all impressions of LEADS are negative. Some principals are actively trying to find ways to use LEADS in conjunction with TESS to improve school leadership. One principal said, "I share TESS responsibilities with my assistant principals. It helps with their LEADS." Other surveyed principals viewed LEADS as a way to understand how teachers view TESS. One principal responded, "[LEADS gives me] a better understanding of what is expected of teachers." Another principal used his experience with LEADS to shape his timetable for TESS: "I try to allow teachers time to reflect before having conferences." Another principal wrote more about how his experience with TESS "made [him] think more about what evidence [he] need[ed] to gather."

Section Six: Question 3: What Evidence is there that TESS has Impacted Teacher Practice and Principal Leadership?

If in any situation a policy has had an impact on the work behaviors of those it seeks to guide, it has had an impact on practice. In this narrow sense, the fact that TESS has impacted the decision making and allocation of time and energy of educators shows that it has impacted their practice. But this narrow construal of the question fails to capture what is of most interest to policymakers, practitioners, and academics: whether TESS has had the *desired* impact on teacher practice and principal leadership. An accurate portrayal

of the impact on practice and leadership must include both the intended and unintended impacts. Only in the light of such a fuller picture may any judgments about the system be made.

Impact on Teacher Practice

There is a difference between teacher and administrator perceptions in the overall assessment of the impact of TESS on instructional practice in the schools and districts. The median response of administrators to the question of whether TESS had had a positive impact on instruction in their schools was "agree", with the mean response falling between "agree" and "uncertain". For teachers, the median response to whether TESS had had a positive impact on their individual practice was "uncertain", with the mean falling between "uncertain" and "disagree". Hence, a preponderance of teachers either were unsure of whether it had helped them improve or flatly disagreed with the notion that it had (see table 10). Some of the differences in viewpoints on this question could be rooted in the fact that administrators were looking at the instruction of all teachers in their schools whereas teachers rated only the change in their own individual practice; however, the difference in perception is highly significant by conventional criteria (see table 10). If anything, the contrast between administrator and teacher perceptions was even clearer in the interview data. Administrators mostly thought the system had had a positive impact on teaching practice. Some teachers thought that it had helped them, but for the most part teachers did not see the system as an effective tool for improvement. The most common teacher response to the question of how TESS had impacted their day-to-day work inside and outside the classroom was, "It hasn't."

Table 11: Responses to Questions "TESS Has Had a Positive Impact on . . . Instructional Practice . . .?"

Administrators ("At your school")	Teachers ("Your own")	Difference	Sig. (two tailed)
2.33	3.1	0.77	0.0005
1 = "Strongly Agree", 5 = "Strongly Disagree"			

A concern about TESS's ability to positively impact teacher instruction is whether and to what extent the teacher behaviors enacted during observations are

authentic or are performances made to manipulate the outcome in the teacher's favor. This topic is such a common concern that it was spontaneously referred to by many interview subjects as a "dog and pony show". One superintendent interviewed took the view that TESS was an improvement over the previous system in that she saw it as less likely to promote the "dog and pony show". Principals were more mixed in their perception of whether TESS led to inauthentic teacher performances. Some did not mention it as a concern at all, while one, who did, explained his relatively sanguine attitude about the possibility by noting that in many cases teachers can show areas of relative weakness even when they are trying to teach better than they typically do, and that this still shows areas in which they need improvement. On the other hand, one principal cited as a major frustration her perception that teachers would artificially elevate their performance and therefore obscure their need for help in certain areas.

Teachers were directly asked about whether they would design lessons to get the best possible rating, and this question drew out a wide range of responses. Some claimed that they would not do anything differently for a lesson they knew would be observed, but several of these contrasted this with their perception that teachers typically did put on a "dog and pony show". Others who owned up to putting on such a show themselves noted that they did so in a variety of ways they thought would make them look good: by choosing certain lesson topics, specific pedagogical approaches, or lessons that were more refined and polished rather than works in progress on which they might otherwise wish to get feedback. All the teachers described these choices as designed to make them look good, not to improve student outcomes. One teacher memorably noted that he felt that he had no choice but to do this, since his evaluation score affected not only his employment and career prospects but also threatened to devalue the time and money he had spent in acquiring his qualifications.

Teacher views of whether the presence of a "dog and pony show" in TESS had implications for its ability to positively affect their instruction were complex. For one teacher, this question of teachers putting on an act for the observation represented a source of measurement error, showing not what teachers actually do but what they can do. Another teacher with a more benign interpretation of this observed that since teachers put many more hours into planning for the observed lesson, it is not reflective of

what teachers do on a daily basis. This teacher noted that his teaching benefited in the short term from going through the process of preparing this idealized lesson by reminding him of practices he should always be doing, but that this improvement faded quickly as the demands of day to day teaching made it impractical to invest the time needed to make the improvements last. Still other teachers had principled objections to this effect in addition to the practical ones. One resented feeling like the system forced her to place her own interests above those of her students: " . . . I'm scoring low in [TESS component]; I want to be proficient, so I'm focusing more on that and less on their needs."

This points to some of the least tangible but most often mentioned effects on teacher practice of TESS: on the mindset, morale, and motivation of teachers. The impact of the system on mindset has been broadly negative. Perhaps the best way to characterize the impact of TESS on the mindset of teachers is to adapt terminology from criminal law. Teachers feel that they are assumed incompetent until proven competent, and they feel that the burden of proof that they are competent falls on them. Teachers took umbrage at the sense that the process reduced their autonomy and devalued and disregarded their professional knowledge and expertise. As a general matter, they feel it is inappropriate for them to be expected to prove their worth in this way. Many noted that this makes them feel less respected as professionals and that it has a negative effect on morale. One teacher said, "TESS has really hurt morale. I'm a professional who has done this for many years. Expecting me to prove I'm good kills passion for teaching and doesn't treat us as professionals". Another captured the sentiment well when she said "Don't make me put on this circus act because it demeans me". Several expressed the concern that younger teachers were having their passion for teaching extinguished and predicted that the system would drive them away from teaching, with the time demands of TESS layered on top of those of learning how to teach. A number of teachers noted that making veteran teachers new to a district do three years of summative evaluations failed properly to take into account their experience or accomplishments. Several teachers mentioned that some high quality veteran teachers either had retired or were considering it rather than persist in having to work with TESS.

Job satisfaction and morale were also damaged by a loss of autonomy. One

teacher said that she had worked in [another professional sector] for [several decades] and had never felt as micromanaged as she does as a teacher, while another simply pointed out that TESS constrains their freedom. One teacher noted that she was constantly second guessing herself because of the pressure to document her practice. Other factors added to the demoralization. One prevailing tactic evaluators use to manage expectations about evaluation scores is to repeat the mantra "You live at a 'three', and a 'four' is someplace you visit", but one teacher said she found it de-motivating to be told this. This same teacher said that being "placed under a microscope" like this added to a more general sense that teachers were not supported by parents or the community, and therefore made her also feel that teachers are not supported by policymakers.

Effect on Teacher Collaboration

Since the TESS framework is based on Danielson's work, a component of the framework is centered on collaboration. While opinions vary among the three districts about the effect of TESS on collaboration, collaboration is certainly a focus of the districts' and state's goals in implementing TESS. Both the surveys and interviews asked teachers questions about the extent of their interactions with peers and administrators. The median teacher response in our survey to the question of whether there was a *great deal* of collaboration at their school was "agree" (with a mean response between "agree" and "uncertain"), but the median response to the question of whether the system *promoted greater* collaboration was "uncertain" (with a mean between "uncertain" and "disagree"). Our interview data are broadly consistent with these findings. Teachers tend to report that high levels of collaboration between teachers exist in their buildings, but many report that these levels of collaboration existed prior to TESS and persist in spite of rather than *because* of it. One teacher said that rates of collaboration have "gone up...but not because of TESS." In some school buildings, collaboration was already part of the school culture. In other buildings, while there is more collaboration, teacher interviews do not suggest that the quality of the collaboration is particularly strong.

Teachers at some schools did report higher levels of collaboration, but typically attributed it to building leadership and culture rather than to TESS as such. In contrast to the issue of peer collaboration,

several teachers in one district/building reported that they thought the system had led to greater awareness of and investment in their work by the principal, and expressed appreciation of the better relationships with their principal which they attributed to TESS. (It is, however, noteworthy that the principal of a building where many of these comments were made expressed an almost visceral hostility to TESS, seeing it as inhibiting his ability to provide timely feedback to teachers.)

There were also multiple unintended effects that ran counter to increasing collaboration. Teachers from three different schools indicated that some teachers refused to share ideas and best practices in an attempt to protect their scores on their TESS evaluations. Other teachers spoke of teachers forming cliques based on perceived teacher status and quality. There were also reports of increases in gossip about other teachers and perceptions of favoritism and bias on the part of evaluators. These results exemplify the exact opposite of how the TESS process is intended to encourage teachers to collaborate. While reports such as these did not surface in interviews at all school buildings, the fact that they surfaced in multiple schools demonstrates a potential negative consequence associated with collaboration tied to TESS. The one area in which collaboration clearly increased was in meeting the requirements of TESS. Teachers did work together to make sense of their new obligations under TESS and to figure out how best to meet these requirements. However, most teachers reported that this was driven solely by a desire to comply with the mandate rather than a desire to pursue authentic improvement.

The most disconcerting data that the comparison of the surveys uncovered was the remaining negative impression that teachers have of TESS's consumption of time and negative effect on their daily practice. In general, teachers agree that TESS interferes with their ability to carry out teaching responsibilities and believe that TESS consumes time and resources that could be better spent elsewhere. One question asked whether TESS interferes with the ability of teachers to carry out other teaching responsibilities. The median response was "agree" but the mean fell between "agree" and "uncertain". Similarly, the question about whether the system consumed time and other resources better spent elsewhere, the median response was also "agree", and even though for this question the median also fell between

"agree" and "uncertain", there was less uncertainty in the interference with carrying out individual duties (See table 12 below). This shows that many teachers thought that the system interfered with their ability to meet other important responsibilities, and even more felt that the system overall was a misallocation of time and other resources.

Table 12: Teacher Responses to Opportunity Cost Questions

	Mean	Median
I believe that the obligations of TESS interfere with my ability to carry out other teaching responsibilities.	2.33	2
The new teacher evaluation system consumes time and other resources that could be better spent elsewhere.	2.02	2
1="Strongly Agree", 5="Strongly Disagree"		

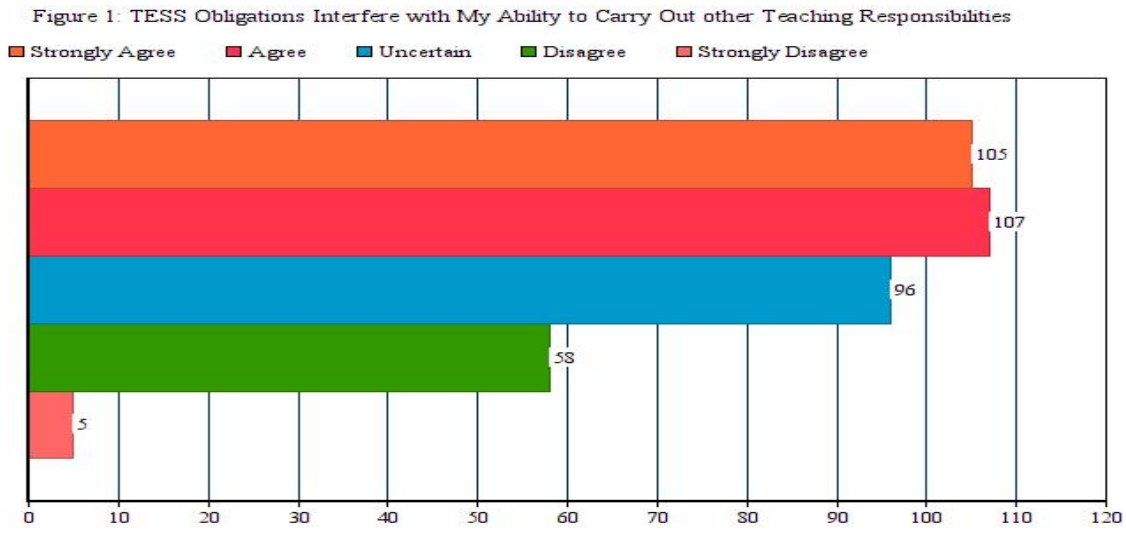
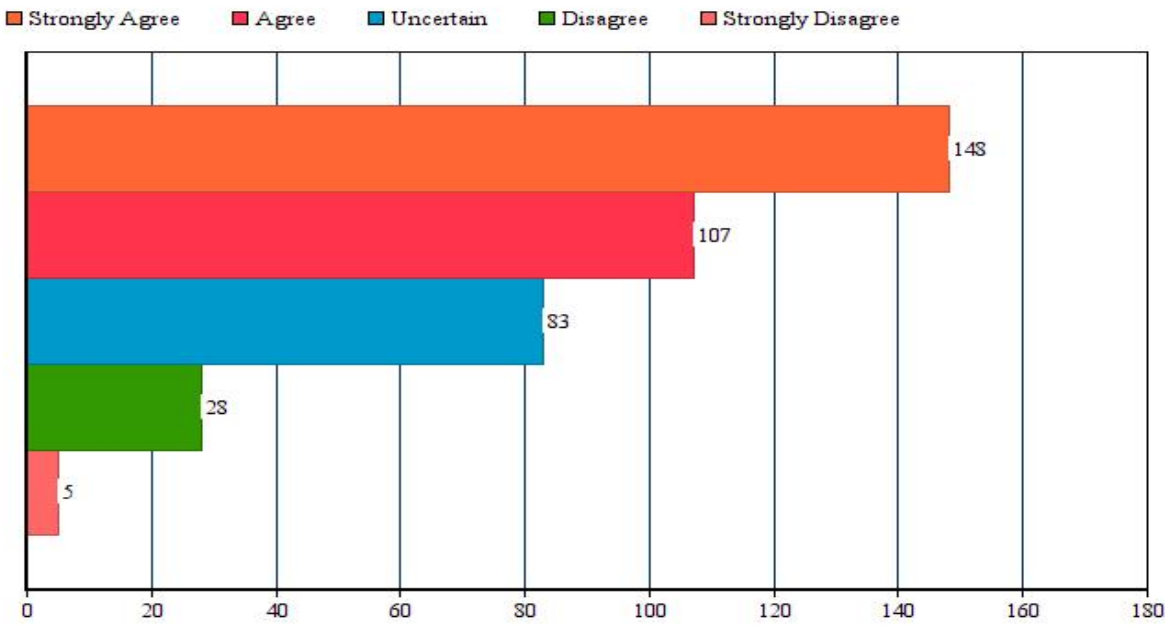


Figure 2: TESS Consumes Time and Other Resources Better Used Elsewhere



This same perception of TESS being an excessive time burden also came through clearly on the interview data. Every teacher emphasized this repeatedly. Many described it as "overwhelming", placing what many saw as an unrealistic burden on both teachers and administrators. One teacher expressed a view suggested by many, that policymakers imposed this additional time demand without fully appreciating the existing time demands of teaching: "Consider what you are giving a teacher to do, considering what they already do. Consider what you're making them do. And if you are going to implement something, come watch us first. Come into a regular classroom, and you decide if it is really going to be effective". Another expressed the view that the time demand was such that both teachers and principals rushed through it in a way that precluded it having desired benefits. One noted that the time teachers spent complying with the requirements of TESS took away from the ability of veteran teachers to mentor younger teachers, and most explicitly stated that the time they spend on TESS took away from their ability to devote as much time as they would like to responsibilities they perceived as having a greater impact on quality instruction, and therefore desired student outcomes. One teacher opined that

whatever was gained in teacher quality was not enough to justify the time it took to get that gain, while another noted that improvements from the process did not stick because of the time pressure of daily responsibilities. By the same token, many teachers perceived the time demands of the system as detracting from the ability of administrators to do part of their jobs that teachers saw as more important. Teachers thought that administrators had all the skills needed to meet the expectations of the system, but not the time in which to do it. Teachers also saw the time demands as compromising the ability of administrators to increase their skill in implementing the system. The collection and submission of artifacts stood out to teachers as a particularly poor use of resources, with one describing this as a "huge waste of my time" and another admitting to meeting this obligation by downloading and submitting documents and artifacts from the internet of which she made only *pro forma* use. (This reflected a common frustration about not receiving feedback about the artifacts, leading some to question whether anyone even looked at them).

Accounting for Teacher Perceptions of the Benefits and Costs of TESS

Our qualitative findings offer a straightforward and compelling account of why TESS is perceived the way it is by the teachers we interviewed. However, we also wanted to examine factors that contribute to the broader perception of the benefits and costs of TESS on the part of teachers. For this analysis we turned to our survey data, since we had data from many more individuals in this data set. We ran a linear regression using the question about overall positive benefit on each individual's practice as the dependent variable and the teachers' scales described above as the independent variables and obtained the following results:

Table 13: Model Summary (DV: Positive Effect of TESS on Teaching)

Adjusted R Squared
0.633

Table 14: Coefficients (DV: Positive Effect of TESS on Teaching)

Model	Unstandardized Coefficients		Standardized Coefficients	Sig.
	B	Std. Error	Beta	
(Constant)	0.182	0.144		0.206
TESS Training	0.233	0.070	0.158	0.001
TESS Skill	-0.197	0.075	-0.122	0.009
TESS Beneficial	0.865	0.056	0.686	0.000
TESS Collaboration and Development.	0.190	0.065	0.124	0.004

This model does a reasonably good job explaining the variation in teacher responses to the question of whether TESS had had a positive impact on their practice, explaining sixty-three percent of the variation in responses (Adjusted r squared 0.63). All of the scales used as independent variables in this model have a highly significant impact under conventional criteria. The strongest effect here is seeing TESS as helping in other ways. Seeing TESS as helping collaboration and professional development, and having received sufficient training, have a moderate effect in making teachers more likely to see TESS as having benefitted their teaching. Positive teacher beliefs about their own teaching skills made teachers less likely to see it as beneficial (in all cases, after controlling for the other variables).

We also wanted to see how well these constructs explained perceptions about whether the system consumed time and other resources that could be better spent elsewhere. Here is what we found:

Table 15: Model Summary (DV: Consumes Resources Better Spent Elsewhere)

R	R Squared	Adjusted R Squared
0.545	0.297	0.289

Table 16: Coefficients (DV: Consumes Resources Better Spent Elsewhere)

Model	Unstandardized Coefficients		Standardized Coefficients	Sig.
	B	Std. Error	Beta	
(Constant)	3.575	0.188		0.000
TESS Training	-0.246	0.091	-0.177	0.007
TESS Skill	0.398	0.098	0.263	0.000
TESS Beneficial	-0.655	0.073	-0.554	0.000
TESS Collaboration and Development	-0.030	0.085	-0.021	0.726

This model has considerably less explanatory power, explaining less than thirty percent of the variation in the responses to the question of whether the system was a positive use of resources. Only three of the four independent variables have significant impact within the model. Here again, the strongest influence on views of whether TESS is a negative use of resources is a positive view of its general effects (negative relationship). Views about the quality and adequacy of the training also vary negatively with the view about TESS's value as a use of resources. Seeing oneself as having the skill to implement TESS expectations predicted one's view of the system as a poor use of resources. In other words, viewing the system as beneficial and the training as well-done made one less likely to see it as a negative use of resources, whereas seeing oneself as already having the skill needed to meet expectations made it more likely that one saw TESS as a waste of time and other resources.

Principal Leadership

As noted above, teachers saw greater principal knowledge of and connection to what was happening in the classroom as a positive impact of TESS on principal leadership. Another benefit from the teachers' perspectives was that some felt that the process had helped build closer relationships with the principal. On the other hand, there were a number of teachers who saw already contentious relationships with administrators grow worse with the introduction of TESS.

Teachers thought the biggest positive impact of TESS on principal leadership was the post-conferencing stage of the process. When asked in interviews what aspect of the process was most helpful for their growth, the post-conference stage was by far the most common response from teachers. However, some teachers said they did not find post-conferences helpful. Some teachers cited aspects of this process as manifestations of principal leadership that helped the teachers become more effective. Many teachers mentioned the process of reflection as being helpful, with a number of these seeing value in the process of comparing their own self-ratings with the principal's ratings of the same lesson. One younger teacher singled out the questioning skills displayed by her principal in this process as being helpful to her. There were a number of teachers who thought that the feedback they received was helpful, although many others thought that principal feedback was of no value for a variety of reasons (e.g. rusty pedagogical skills of the principal, deficient content knowledge, or a lack of specificity). More broadly, even though some teachers said that they found the principal feedback to be helpful, teacher confidence in the ability of TESS feedback to help them improve actually declined since initial implementation.

Administrators had different views of the impact of TESS on their leadership. As noted above, the median response of administrators to the question of whether TESS helped them have better conversations with teachers was "agree", with the mean between "strongly agree" and "agree". In our interviews, numerous principals said they thought that TESS helped them give better feedback to teachers due to the greater specificity facilitated by the components of the rubric. The superintendents also thought that principal feedback for teachers had improved because of TESS.

As with teachers, the most salient impact of TESS on principal leadership was the amount of time it took for implementation. This comes out most comprehensively in the quantitative data. Administrators were asked a range of questions about the extent to which the time it took to implement TESS had affected a number of their other responsibilities: handling student discipline issues, conducting casual classroom walkthroughs unrelated to TESS, interacting with students, attending student-related conferences, completing other paperwork,

and spending time reflecting. For the conferencing and paperwork questions, the mean response fell between "somewhat impacted" and "slightly impacted", with a median response of "somewhat impacted". For the remaining questions, the mean response fell between "greatly impacted" and "somewhat impacted", and the question about the casual walkthroughs had a median response of "greatly impacted" (This was the only Likert-scale type item in either survey with a median response at one of the extremes). In evaluating the tradeoffs involved with performing TESS tasks, the median responses to the questions of whether or not TESS interferes with administrators' ability to perform other duties and whether it consumes time and other resources better spent elsewhere was "uncertain", with the median falling between "uncertain" and "agree". Only twenty-four percent disagreed that TESS interfered with their ability to perform other duties, and only sixteen percent disagreed with the idea that it consumes time and other resources better spent elsewhere. One way of interpreting these numbers is to say that fewer than one in six administrators thought TESS was a wise use of time and other resources. (Compare with the 7.5 percent of teachers who saw it as a good use of time and other resources.)

This concern about the time demands associated with TESS comes out in the interview data as well. One superintendent thought that principals possessed the skills to implement TESS effectively, but lacked the time to do so. One expressed concern about how sustainable it was for principals to spend their evenings doing work they were unable to finish during the workday due to the time demands of TESS. Principals highlighted the documentation, other paperwork, and working with teachers to collect and submit artifacts as particularly time consuming. All made reference to how formidable the time demands were, and one, who was eager to emphasize that he wanted to spend the needed time to make it work since teacher quality was such a priority for him, recommended that principals be relieved of other responsibilities to make the time demands manageable. A more skeptical (and more experienced) principal said that the time demands did not pay off since she would have done the parts of TESS that lead to better teaching anyway. Several teachers noted, even without prompting, how substantial the time demands on principals

were, with a number expressing empathy for the administrators.

Accounting for Administrator Perceptions of the Benefits and Costs of TESS

As we explored teacher views of TESS, we wanted to get a broader sense of what explained administrator views of the benefits and costs of the system. In a similar manner, we ran regressions on two key dependent variables using the scales we developed as independent variables. In this case, the regression on the variable about whether TESS is a good use of time and other resources produced an adjusted R squared of 0.000, so we will only report here what we found using the question about improved teaching practice in the school as the dependent variable. Here is what we found:

Table 17: Model Summary (DV: Overall TESS has positive impact on instruction)

R	R Squared	Adjusted R Squared
0.781	0.610	0.510

Table 18: Coefficients (DV: Overall TESS has positive impact on instruction)

Model	Unstandardized Coefficients		Standardized Coefficients	Sig.
	B	Std. Error	Beta	
(Constant)	0.229	1.111		0.839
TESS Capacity	0.264	0.256	0.155	0.314
Value of TESS	0.934	0.189	0.784	0.000
Opportunity Cost of TESS	0.041	0.209	0.031	0.846
Desire for more training and support	-0.322	0.237	-0.202	0.190

Overall this model is strong, predicting more than half of the variation in the dependent variable. However, since

only one dependent variable is significant, practically all of the predictive power of the model comes from this one variable. This tells us that for administrators, more general positive views of the effects and value of TESS predict the view that TESS has had a positive impact on instruction in the school. Since all of the power seemed to come from this one variable, we did two further tests. The first found a strong correlation between this variable and the dependent variable (0.729), and the second showed that in running the regression again with only this variable as the independent variable, the adjusted R squared is slightly higher and the standardized correlation coefficient is the same: 0.729. Positive views of TESS in general strongly predict positive views of the impact of TESS on instruction in their schools for administrators.

LEADS

LEADS has been a lower priority for the state, and has consumed much less time and attention than TESS. Accordingly, LEADS has already had some effect, but not as much of an impact on practice as TESS. Some of this influence is reflected in the quantitative data. Administrators tend to think the measures used in LEADS are well selected, with the median response to the question of how well selected they are being "well" (with the mean response falling between "well" and "fairly"). Administrators believe that LEADS has led to better feedback from their evaluators, with the median response to the question of whether it has led to better feedback being "yes". As a whole, administrators have some confidence that the system will help them improve their practice, with a median response to the question of whether it would being "agree", but a mean falling between "agree" and "unsure".

The interview data also suggested generally positive impacts. Several relatively newer principals thought that the framework provided a good guide to practice, with one saying that the clarification of expectations was especially helpful to him as a new administrator, and another asserting that the framework provided more clarity and guidance about how to do well than any supervisor ever could. On the other hand, a more experienced principal claimed that she would have done all of the things LEADS detailed anyway. In some instances, principals were uncertain about how to document their performance in the LEADS

categories. One thought that documentation was not possible in some cases, and another noted that she took the initiative to agree with her supervisor in advance on how to document her work. Several principals noted that they found the goal setting and reflection aspects of LEADS to be helpful. Several also mentioned that they thought that being evaluated by LEADS would better help them understand teachers' experiences of being evaluated by TESS. Superintendents thought that LEADS had improved their conversations with principals, helped with goal setting, and gave greater structure and clarity to the principal development process. They also thought that LEADS had helped with TESS implementation since it improved how principals were being held accountable for teacher development. On the other hand, all interviewed administrators thought that principals lacked sufficient time to implement LEADS effectively, and many implied (and one stated outright) that the time demands of TESS were so pressing that LEADS was an afterthought.

Section Seven: Themed Findings

Growing Levels of Comfort and Frustration

In the time since the implementation of TESS, an interesting tension has developed. While principals and teachers have become more familiar and comfortable with the systems, teachers have a largely negative perception of the process. While it is understandable that professional evaluations cause stress and take time, these costs can be justified if they yield sufficient benefit. This does not appear to have been the case with TESS. The extent to which teachers dislike TESS and feel that it is a waste of time is highly problematic. At the same time, teachers have become more certain that the TESS rubric itself reflects quality teaching practices.

Principals, in general, have a much more positive perception of TESS than teachers. This difference seems to be related

to differences in perception of how the tool is used and the difference in ways that administrators and teachers perceive work design. Principals view TESS as an improvement in the way that teachers are evaluated, which is central to their work as school administrators. While principals are frustrated with the amount of time associated with TESS, TESS forces principals to spend more time in classrooms and interact with teachers. As a result, the discomfort associated with extra work is palliated with an increased sense of professionalism, more time spent working with teachers, better quality interaction with teachers, and a better system of teacher evaluation. The implementation of TESS was fueled by administrator participation, and was guided by the idea that the old systems being used for teacher evaluation were unreliable and ineffective. In short, TESS was presented as a way to improve principals' work, and as a result, principals have a more positive perception of TESS than do teachers. .

On the other hand, TESS has added a lot of extra work and anxiety to teachers' jobs, yet offers little advantage. Teachers are increasingly frustrated that TESS is taking too much time away from students and is having a negative effect on their ability to conduct their work as they think best. Teachers also have frustrations about how they are scored. The fact that teachers do not think it is possible to receive regular scores of "4" hurts morale. One teacher says, "I hate, hate, to hear: 'just visit in distinguished'". Another teacher says, "In TESS, you start with a defeated attitude; you're only going to visit 'distinguished'". One principal acknowledged this frustration. She thought that the type of people who typically become teachers are generally the type of people who want to achieve and have had a lifetime of positive experience with schooling; they want to score well on evaluations. When teachers are relegated to a score of "3", they feel that their efforts are not acknowledged or rewarded. One teacher says, "'Threes' are hurtful...I want a pat on the back." More to the point, teachers generally struggle to find any reward in the TESS evaluation process. Some teachers mentioned that TESS has helped them become more reflective of their practice, but few expressed any tangible rewards or explicit appreciation of the process.

It seems that teachers' negative opinions of TESS stems from their perception that TESS is almost purely a method of evaluation rather than as a tool to

improve practice and student achievement. Since the training associated with TESS largely focused on the actual process, and not how the process would help teachers or how to actually design lessons to meet the needs of the rubric, a key issue in the implementation of TESS stands out as problematic (Ashby et al., 2014). Moreover, according to interviews and surveys, the amount of TESS training has decreased since its implementation.

While it makes sense that TESS training was front-loaded upon implementation, many teachers opined that they had received very little training about TESS over the past two years. This is important for two reasons. While all principals deal with TESS on a daily basis, and have grown to see how the process helps them, teachers on the four-year cycle have had fewer opportunities to see how the process can help their practice. Moreover, professional development and PGP work seem to be largely focused on improving performance on TESS, rather than classroom performance. While the two ideas (TESS performance and classroom performance) are intended to be inextricably linked, TESS is serving as a thin filter that is actually separating teachers from the work they find to be most important.

Continued Variation Between Schools and Districts

Teacher and administrator perceptions of TESS and LEADS vary by schools and districts. Using ANOVA to look at teacher perceptions associated with TESS across the three school districts, it was rare to find general trends on which the three districts agree. Generally, the Jonesboro school district has a more positive opinion of TESS than Nettleton or Westside. However, none of the districts' teachers have a positive perception of TESS in general.

Across the three districts, teachers are skeptical about the impact on student achievement, with the overall mean response to the question about whether it had had a positive impact on student achievement falling between "Unsure" and "Disagree" (3.19). Jonesboro teachers are slightly less skeptical (2.91), while Nettleton teachers (3.40) and Westside teachers (3.67) take an even dimmer view of the impact on student achievement. The disagreement between Jonesboro and both Nettleton and Westside is statistically significant ($p<0.000$), but the difference between Nettleton and Westside on this point is not significant ($p=0.203$).

The typical Jonesboro teacher remains unsure about the impact on student achievement, whereas the typical teacher in the other districts is fairly confident that it has not had a positive impact.

In general, the teachers are unsure about whether feedback from TESS evaluations has helped them improve their practice (2.92). Jonesboro teachers are slightly more likely to agree that feedback from evaluations have helped them improve their practice (2.74). Nettleton teachers are quite close to the overall view about the effect on their practice (2.93), while Westside teachers are more likely to disagree that feedback has benefitted them (3.48). In this case, the statistically significant differences are between Westside and Jonesboro ($p<0.000$), and Westside and Nettleton ($p=0.003$). Westside teachers stand out as having more negative attitudes than the teachers in the other districts about the value of feedback in improving their practice.

Survey results are also mixed showing whether or not teachers perceive the implementation of TESS to fit in with other initiatives that are also in the process of implementation, such as Common Core and other school or district wide policy implementations (2.72 overall mean). Jonesboro teachers almost lean toward agreeing (2.55). Nettleton is closer to uncertainty (2.83) and Westside teachers are closer to disagreeing than to agreeing (3.11). For this question, Jonesboro is significantly more positive than Nettleton (p-value of the difference 0.03) or Westside (p-value of <0.000). Jonesboro teachers are more likely than those in the other districts to see TESS as consistent with other initiatives, but even they are closer to uncertainty than to agreement.

Jonesboro teachers also have the rosiest perception of whether TESS interferes with their ability to carry out their professional obligations. Jonesboro teachers' (contrary to the overall results) point to "unsure"; however, their scores are closer to agree than to disagree (2.65). On the other hand, Nettleton (1.97) and Westside (1.95) teachers agree that TESS interferes with their ability to carry out professional obligations. Again, Jonesboro teachers display significantly more positive views on this question than do teachers at Nettleton (p-value of the difference <0.000) or Westside teachers (p-value of the difference<0.000), but they still lean towards the view that it interferes with their other professional obligations.

One relatively bright spot illuminated by the survey data suggest that teachers are collaborating throughout the three districts. Jonesboro (2.02) and Nettleton (2.06) "agree" that there is a great deal of collaboration throughout the three districts. Westside survey data indicate that teachers in the district almost agree (2.54).

Unfortunately, we cannot tie any increase in collaboration directly to the implementation of TESS. Jonesboro teachers are unsure, but lean toward agreeing that TESS helps teachers collaborate more (2.84). Nettleton teachers (3.21) and Westside teachers (3.46) are uncertain, but their survey results lean to disagreement. While the differences between Nettleton and Westside are not significant, those between Jonesboro and Nettleton (p-value of the difference 0.01) and between Jonesboro and Westside (p-value of the difference <0.000) are. The prevailing attitude about whether TESS has improved collaboration is uncertainty, but where Jonesboro teachers lean a bit towards agreement, Nettleton and Westside teachers lean toward disagreement.

We see a similar pattern in responses to the question of whether the quality and frequency of professional conversations has increased due to TESS. Jonesboro teachers were uncertain with a slight tilt toward disagreement (3.001), Westside teachers leaned more heavily toward disagreement (3.33), and Nettleton teachers were closest to outright disagreement (3.46). The only statistically significant difference in this aspect is between Jonesboro and Nettleton (p-value of the difference <0.000), so we can say that Jonesboro teachers are more likely to agree about TESS improving the quality and frequency of professional conversations than are the Nettleton teachers. However even the Jonesboro teachers are uncertain.

It is interesting to consider these findings from the perspective of how the three districts differed in their implementation of TESS. Jonesboro piloted the program, while Westside implemented the program a year early, but was not part of the state pilot group. Nettleton was on the standard Arkansas schedule for implementation. As a result, Jonesboro teachers had access to additional materials and training that teachers of the other districts lacked. While Westside used a book study to help the teachers acclimate to TESS, it is not apparent that the teachers were able to tie the book study to their perceptions of how their jobs would change. It also suggests a crucial role for

district and building leadership in these outcomes. The superintendents leading Jonesboro's implementation were exceptionally committed to the implementation and devoted innovative approaches and extra resources to implementation. Westside's superintendent had a similar level of commitment to implementing TESS. However, two of the three Westside principals are new to their positions. While they are described more as "instructional leaders" than the "building managers" they replaced, it might be that the instability in building level leadership over the course of this process compromised some outcomes. Finally, as we saw above, Westside differed from the other districts in giving more low ratings, and this greater rate of providing critical feedback to teachers might also account for the more negative perceptions of the system by Westside teachers.

Unrealized Aspirations and Unintended Outcomes

There are a number of ways in which the goals of TESS (and to a lesser extent LEADS) have not been realized. For TESS, the first goal was "Provide school districts a transparent and consistent teacher evaluation system that ensures effective teaching and promotes professional learning". The components of the framework are reasonably transparent, but the levels are not (as one principal said, "Don't tell me 'all' or 'most'; give me a number"). Despite much time and effort spent on certifying administrators as observers and scorers, there is widespread perception of imperfect inter-rater reliability. Insofar as this is true, it constitutes a lack of consistency. It is unclear the degree to which TESS promotes professional learning, but it does not do so to the extent that it 'ensures' effective teaching.

Goal two was "Provide feedback and a support system that will encourage teachers to improve their knowledge and instructional skills in order to improve student learning". The TESS process does provide feedback, and this feedback does help some teachers. However, more teachers do not find the feedback helpful. Moreover, the quality of the feedback is compromised by the measurement problems of the presentation of idealized lessons by teachers and the lack of content expertise and role conflict for principals. Similarly, TESS does provide a support system in the form of constructive feedback from principals and better aligned

professional development, but the time demands of the system and other resource limitations compromise the degree to which these supports are effective at helping teachers improve their knowledge and instructional skills to improve student learning.

Goal three of TESS was "Provide a basis for making teacher employment decisions". TESS can, in theory, be used this way, but we found no evidence that it had been. One principal described assigning new teacher leadership duties to a teacher he identified through TESS, but otherwise there was no indication of changes in job duties or promotions because of TESS results. On the other hand, there was widespread skepticism about whether TESS would be used for dismissal, and whether it would be effective for that purpose if it were so used. Broadly speaking, TESS did achieve the fourth goal of creating an integrated system, but we do not have evidence of it achieving the rest of the goals.

For LEADS, the goals were more modest, and many of them can be (and have been) met by the adoption of a quality rubric. Though, as with TESS, the time demands associated with LEADS (on top of the time demands on principals associated with TESS) are such that it seems unlikely that the goals will be completely realized.

There have also been some unintended outcomes (these are a result of TESS, as LEADS has not been implemented with sufficient intensity for concrete outcomes to emerge). The main unintended outcome is the weight of the time burden. The expectation was that the time burden would be manageable, but this does not appear to be the case. Another unintended outcome is the effect of combining the summative and formative functions of the system. The summative aspect was intended to be de-emphasized, but its presence seems to dominate teacher perception of the process and through that mechanism seriously compromises its effectiveness as a developmental tool. A third unintended outcome is goal displacement. The system is designed to promote effective teaching, but, largely due to the summative element, in many cases it motivates teachers to put on a performance to get a higher rating rather than to seek genuine improvement. On a related note, the system is intended to promote greater collaboration, but it appears to increase collaboration only for the sake of complying with the requirements of TESS rather than for actual improvement. In some cases,

teachers are either unwilling to share practices that lead to high ratings or unwilling to accept the help of highly rated teachers who offer it. Most significantly, TESS seems to have had a negative impact on teacher autonomy, professionalism, motivation, and job satisfaction.

A Series of Continued Trade-offs

The most tangible trade-off with TESS is the opportunity cost of the time TESS requires. Most teachers describe spending a great deal of time on TESS, partially on planning the observed lesson. While this planning directly impacts instruction, it has some hidden trade-offs. Because of the summative function of TESS, teachers have an incentive to plan an idealized, rather than, a representative lesson, which makes planning more time consuming than planning a typical lesson. However, the more teachers devote extra time to planning a lesson, the less representative the lesson is, thus degrading the accuracy of the rating. The bulk of the time cost of TESS is spent on pre- and post-conferencing, reflecting, completing paperwork, and collecting artifacts. Conferencing and reflecting impact instruction, and the time involved was not identified as cumbersome. Conversely, the paperwork and documentation were described as quite time consuming.

The prudence of this time commitment can be evaluated in a few ways. One is to weigh these costs against the benefits. Administrators and teachers disagreed about the benefits, with administrators generally seeing the system as having had a positive impact and teachers generally being more ambivalent or negative. Perceptions of the benefits have not improved; our qualitative data indicate that skepticism has grown about whether the system will ever factor into employment decisions, and there has been no statistically significant increase in perceptions of the benefits of the system. A more telling way to measure the value of the time spent is in terms of the opportunity cost. In interviews, many teachers and some principals said they believed that TESS obligations interfered with their ability to positively impact the school community (with principals noting that one effect is that they are not able to spend much time or attention on LEADS). Our survey data bring this out even more strongly. In terms of the impact on individual practice, forty-eight percent of principals see it as interfering with other job

responsibilities, and 56.7 percent of teachers agree (with 28.3 percent strongly agreeing). In terms of the allocation of resources within the school more generally, thirty-six percent of principals see it as consuming resources better spent elsewhere (against only sixteen percent who disagree), and 68.7 percent of teachers see it as a misallocation of resources (with 39.8 percent strongly agreeing).

Another key trade-off is the exchange between external accountability and professionalism. TESS and LEADS are designed to measure and promote teacher and principal quality, respectively. They both include notional summative elements in response to public desire for increased accountability in education. However, inasmuch as the TESS and LEADS systems increase accountability to external constituents, they undermine the professionalism of teachers and principals by shifting the emphasis away from internal accountability.

A third trade-off is between consistency and sensitivity to local context. One of the purposes of TESS is to provide consistent ratings. Such consistency is valuable in that it promotes confidence in the reliability and validity of the findings, and in the ability to make comparisons across schools, districts, and the state. TESS and LEADS have high specificity, and the greater the specificity, the greater the consistency. However, this consistency conflicts with sensitivity to local context, in turn limiting local buy-in and ownership of the process.

The rubrics are a strong example of this phenomenon. Teachers are evaluated by the same rubric in TESS, in part for the sake of consistency. However, using the same rubric for primary and secondary teachers, as well as for teachers of all subjects, results in many mismatches between the instrument and the pedagogical realities of a variety of teaching situations. In extreme cases, such as special education and resource room settings, a number of the metrics used simply do not apply. LEADS is somewhat better than TESS on this point, with different rubrics for different administrative roles.

The broader processes also involve trade-offs between consistency and sensitivity to local context. The systems are fairly prescriptive, for the sake of specificity (and in fact, one superintendent thought that it needed to be more prescriptive, lamenting a lack of guidance about how many artifacts were required for LEADS. Nonetheless, this superintendent also thought that some of the

different levels on the rubric were overly prescriptive). However, this prescriptiveness constrains flexibility to adapt to local situations. Both teachers and principals lack the discretion to balance the demands of TESS and LEADS with other demands on their time or attention, or to adjust the way they are conducted to suit particular situations. The result is that teachers and (especially) principals attend to the quantity of their work with the systems at the expense of quality, and in some cases will engage in the required activities while acknowledging that it does not suit the situation.

Section Eight: Discussion

Several different areas of scholarship help us develop a greater appreciation of how our findings compare with and contribute to the existing understanding of teacher and principal evaluation. Following our conceptual framework, we examine how our findings relate to scholarly understanding of policy implementation, teacher and principal evaluation, and the social organization of schools.

Policy Implementation

The literature we reviewed included two comprehensive accounts of what contributes to successful policy implementation (Murphy, Hallinger, & Heck, 2013; Spillane, Reiser, & Reimer, 2002; Stronge & Tucker, 1999; Taylor & Tyler, 2011). One set of scholars (Fixsen, et al., 2005), takes a structural/process oriented approach, whereas another (Desimone,

2000) looks for the presence or absence of decisive abstract qualities. Fixsen et al. (2005) emphasize the importance of going beyond the provision of information and initial training to include ongoing coaching and feedback for people assuming new roles. Our findings support the notion that more is needed to affect meaningful change in practice than information and initial training. Evaluators have had more training and practice since initial implementation, and over that time, teachers have only grown more pessimistic about the prospect of improving their practice due to the feedback they get in TESS. More positively, the district that had been a pilot and whose assistant superintendent had described their approach to training as building-based and interactive (and a departure from what the state had recommended) tended to have more positive attitudes than the other districts.

Another best practice advocated by this set of scholars is ensuring that the separate components of implementation overlap with one another in an integrated fashion (Fixsen et al. 2005). TESS has reshaped many aspects of the teachers' and principals' professional lives. TESS is used for both pre-service (at least introductory training for all new teachers and administrators) and in-service training. Teachers need to record their artifacts from lessons on BloomBoard regularly. TESS is designed for teachers to meet yearly with their principals in order to identify their PGP. By overlapping efforts to give teachers and administrators opportunities to learn about the TESS system and how to use it, the districts use a best practice technique to allow the policy to take hold. Beyond this, Datnow and Park (2009) write, "the sense making and co-construction perspectives build upon the importance of context in the mutual adaption view by elaborating on the interconnections between actors and explaining just exactly how context has shaped policy implementation" (p. 350). As teachers work through the TESS model with their administrators, what it does and does not mean for each individual and for the organizations as wholes will come to be defined and solidified.

Fixsen et al. (2005) also borrow a scale from Goggin (1986) that describes degrees of implementation. The three steps here are paper implementation (where the focus is on generating records indicating that implementation had taken place), process implementation (where new procedures are put into place), and performance

implementation (where the implemented policy has the intended benefit). This framework helps us categorize how deep the implementation has gone: for some teachers it is merely paper implementation. For the districts as a whole it is process implementation (based on the fact that observations and conferences are taking place, but with minimal apparent effect), and for only a few teachers and administrators it is performance implementation.

Desimone's approach of looking at effective implementation in terms of key qualities taken from Porter (1988, 1994) also illuminates our data. These qualities are specificity, power, authority, consistency, and stability. TESS's specificity seems to have indeed helped with the pace and fidelity of implementation, but not to have been appropriately balanced with teacher professionalism, flexibility, or limits on time for professional development. With the possible exception of the Fair Dismissal Act, many administrators tend to see it as consistent with other policies and practices, although teachers are less likely to see it that way. The stability issue also seemed to apply, since many teachers and administrators thought the policy would not last and this undermined the extent to which they wished to invest time and energy in the policy. The most telling (and likely most important) way that this framework informs our interpretation of our data is the dual issue of power and authority. Desimone argues that promulgating a policy change through power rather than authority is counterproductive, and that is what seems to have been the case with both TESS and LEADS. If Desimone is correct, implementing the system by building authority (through, for example, teacher decision making, buy-in, participation in networks and collaborative activities, and principal expertise) might have been more successful. Coburn (2003) makes a similar point about teacher buy-in and leadership as essential parts of promoting the quality of scaled up effective practices.

These vital elements, such as authority (in contrast to formal power in Desimone's sense), teacher input and buy-in, and the quality of building level leadership, are all heavily dependent on local context. The critical role played by these varying local elements in perceptions and outcomes lends credence to Honig's (2006) observation that it is too simple to ask *whether* a policy works. It is better to ask *under what conditions* it works, and suggests that Crowson and Goldring's (2009) observations that complex policies

require local knowledge, resources, and autonomy to operate effectively apply to state policy as well as federal policy. Similarly, Coburn's (2003) conceptualization of quality (which she maintains should be the primary way we look at bringing reforms to scale) requires that we not only attend to local contextual factors in evaluating policy implementation, it requires giving these local contextual factors a pivotal role in the implementation of policy. With TESS and LEADS implementation, the lack of buy-in, voice, and flexibility to adapt the processes to local circumstances have been major elements that we uncovered.

This point represents a significant departure in our findings from those of the prior team. Where Ashby et al. (2014) found that the state had given districts significant discretion in how to enact the mechanics of early stage implementation (i.e. how to implement the system), we find insufficient sensitivity to local context in the amount of flexibility given to districts about the design or requirements of the system (i.e. what to implement). The sense that some requirements were not aligned with actual needs and the perceived inability of teachers and administrators to prioritize their use of time and other resources because of the documentation requirements of TESS are significant problems with the systems. This, in turn, suggests that Datnow's (2006) sense-making/co-construction approach to understanding policy implementation offers considerable benefits.

Teacher and Principal Evaluation

The most significant point of contention within the teacher evaluation literature is whether formative and summative purposes can be served by the same system, or if these functions must be separate. Goe et al. (2012) argue that these purposes can be combined, while Danielson and McGreal (2000) claim that these can co-exist provided that the emphasis falls strongly enough on the developmental function. On the other side, Duke (1990) asserts that they need to be separate because the atmosphere must be safe for the experimentation and risk taking that is required for growth. Our administrator interviews strongly suggested that superintendents and principals believe that the emphasis of TESS is sufficiently on growth so that these purposes do not conflict. However, our teacher interviews strongly indicate that these functions are incompatible. The very possibility that the

outcome of TESS could have a negative impact on their jobs leads teachers to perceive the system as mainly summative. On the other hand, the lack of actual negative consequences for anyone because of TESS suggests that it is not effective as an accountability system either. Part of this is no doubt due to the tendency of evaluators to inflate scores in order to manage morale or maintain ongoing relationships (Kimball & Milanowski, 2009), but some doubt could simply be due to the fact that the system places less emphasis on the summative function. In trying to serve both functions, TESS does neither well.

This conclusion should not be taken as a refutation of Danielson's claim that the purposes can be combined, for she is careful to emphasize that making this work requires an adequate investment of time and other resources. For example, she contends that the collection of artifacts and other documentation as additional data sources beyond classroom observations can be done in a way that is not unduly burdensome. Perhaps it can be, but in the case of TESS our interviews clearly suggest that it was done in a way that imposed a cost in time and resources seen as excessive. Firestone et al. (2013) show that the issue of time intensity is not unique to TESS, but also occurred in an evaluation system in New Jersey and might be taken as an expected consequence.

There are a number of our findings that align with features recommended by the literature. Danielson & McGreal (2000), Darling-Hammond (2012), and Johnson and Fiarman (2012) note the need for clear and agreed standards of performance, and the TESS and LEADS rubrics are both widely perceived to represent good practice. Danielson & McGreal (2000), Darling-Hammond (2012), and Goe et al. (2012) all see the need to incorporate teacher (and principal) evaluation into a wider system of ongoing professional growth, and both systems extend beyond simply rating with a rubric. However, other parts of the system might suffer from the time intensity. Darling-Hammond (2011) and Goe et al. (2012) emphasize the need for high quality professional development processes that would themselves be time intensive, and none of our data suggests that any professional development, aligned with TESS or LEADS results, stands out as high quality.

A final area in which our findings resonate with themes in the literature is the credibility of and trust in the evaluator. Clipa (2011) notes that teachers prefer

outside evaluators, while Kimball & Milanowski (2009), as well as Sartain et al. (2011), claim that evaluators need training on deep impact teaching approaches in addition to a rubric and basic training. Johnson and Fiarman (2012) show how elements contributing to the credibility of Peer Assistance and Review (PAR) evaluators include being active teachers only temporarily out of the classroom, being highly trained, and being selected in a highly competitive process. Our data showed numerous issues with TESS evaluators, from concerns about bias and lack of expertise to concerns about conflicting incentives. Although we did not find a clear preference for outside evaluators, we did find a number of factors that limit the ability of the evaluators to observe and give feedback that is credible to teachers.

Social Organization of Schools

One of the key considerations about how TESS affects the social organization of schools is the degree of professionalism that characterizes the teaching role. Enhancing professionalism is an explicit goal of TESS, and one superintendent specifically mentioned that he perceived greater professionalism because the system encouraged membership and participation in professional organizations. Another way in which the system can be seen as having bolstered professionalism is by providing a common set of standards of teaching quality in the form of the rubric framework. Grossman (2001) and Lortie (1969) argue that one issue that keeps teaching from a professional status is a lack of shared norms and values, and the rubric takes a step in that direction.

However, by and large TESS has had a negative effect on the professionalism of teachers in the Jonesboro area schools. Lortie (1969) argues that in professions, practitioners control the monitoring, discipline, and rewarding of their own members. With TESS (and LEADS), the evaluation and allocation of related consequences is done in a traditional, hierarchical fashion. This means, it is judged and sanctioned from the outside, rather than internally. Whether or not principals are outside the teaching profession is certainly subject to debate, but teachers in Jonesboro do not typically see them as colleagues in that sense. This situation is exacerbated by the perception held by many teachers that the burden of proof is on them to show that they are competent. TESS and LEADS degrade the

professionalism of the teacher and principal roles by reinforcing control by non-practitioners, and lead some teachers at least to feel that their competence is being questioned. Moreover, while it is unclear how widely shared the perception that monitoring participation in professional organizations enhances professionalism, it is clear that it is not universal. One exasperated teacher noted, "If I join a professional organization because of TESS, it's going to be a union".

A crucial respect in which TESS has affected the social organization of schools is that it has further tipped the balance of the social dynamics of the school away from being a community and more toward being an organization. This distinction draws on Toennies' (2001) distinction between *gemeinschaft* and *gesellschaft*, between community and society. These authors insist that schools, as human service enterprises serving children, work best as communities. Sergiovanni (1994) argues that where organizations rely on formal authority and consequences, communities rely on referent power and intrinsic motivation, drawing on shared ideas and values. For Sergiovanni, in order to be communities, schools would need to invent new measures of quality and accountability. Beck and Foster (1999) extend this reasoning to claim that schools as organizations have a clear administrative role of commanding an organization along the most rationally efficient way of meeting a goal, but a community requires less control and might not take the most efficient path if another is more consistent with the community's shared values. TESS and LEADS are formal processes of measurement and control of a bureaucratic nature, and make the social dynamic less like that of a community. Many of the processes and interactions mandated in TESS and in LEADS might work in a community context, but formally mandating the system makes the enterprise more organization-like.

Collaboration is an activity that is often manifest in a community, and encouraging collaboration is a goal of these evaluation systems. TESS has a component encouraging interaction with colleagues, and LEADS has several domains about the role of leadership in fostering collaboration. Collaboration offers a number of benefits. Johnson (2015), Murphy (2013), and Scribner, et al. (2002), all argue that collaboration promotes teacher development and, therefore, student achievement. Unfortunately (and in keeping with the

implicit shift away from community), while participating in a professional community is part of the TESS rubric, TESS and LEADS do not seem to promote collaboration per se beyond incentivizing it as part of the rubric. Aside from the closer working relationship between evaluator and evaluatee, we were unable even to locate any explicit mechanism designed to foster collaboration directly related to TESS.

One last consideration about the impact of TESS and LEADS on the social organization of schools has to do with job/work design. Grant and Shin (2012) describe the impact that task significance and autonomy can have in promoting motivation, satisfaction, performance quality, and in reducing withdrawal behaviors. This sheds light on our findings in a crucial way. One puzzle has been why the attitudes of administrators and teachers towards TESS are so different. The task significance perspective is illuminating: the process helps administrators do their jobs more effectively so they view it more favorably than do teachers, for whom it does not offer help in doing their work. The time intensity of the systems and the increased monitoring can be seen as reducing autonomy, which also constitutes a problem in job design. Grant and Parker (2009) report that one type of social support for work is organizational support, where perceptions that an organization values and supports one's work are motivating. Insofar as TESS is seen as an imposition of additional responsibilities without additional resources, it is likely to have the opposite effect.

Section Nine: Recommendations and Conclusion

Our recommendations follow three general ideas: improving program implementation, separating the evaluation instruments' summative and formative purposes, and finding a more efficient use of resources.

Improving Program Implementation

Finding: Aspects of the TESS implementation have hindered the effectiveness of the policy.

The popular image of teachers leaving every day at three o'clock in the afternoon and having summers off from work is a myth. In addition to active classroom instruction, planning, communication with parents and students, grading, and other school duties permeate through almost every aspect of teachers' lives. Each new school policy and reform that is implemented adds to teachers' workloads. Teachers see TESS as an added onus; additionally, they do not see the benefit of the process.

One set of our recommendations can be thought of as "lessons learned" that can inform future state and district level implementation of education policy. One element of this set is the time cost to ground level implementers of new policy. We recommend that when new duties are added to teachers' and administrators' responsibilities, other duties need to be eliminated in proportion to the new ones added. When jobs are designed with insufficient time to complete all job requirements with an acceptable level of quality, the quality with which some responsibilities are met falls below an acceptable level. Unless a job is designed with more time than is needed to meet responsibilities, changes in the total time demands should only be done with great care. In some circumstances, new duties can be acceptably added if they enhance a worker's ability to perform existing duties. Therefore, the benefit of new responsibilities to workers' ability to achieve core goals must be made clear, especially if these new responsibilities represent a net increase in time obligations. Perhaps if TESS had been implemented with clearer and more convincing framing that it would improve teachers' practice and professional lives, teachers would have a better perception of the process.

Another key lesson learned regards phasing in new policies. We recommend phasing in the implementation of future policy changes in a manner that gives workers more time to adjust. This would give practitioners more time to find ways to meet their new duties well and reduce the extent of being "overwhelmed". As we have seen, the extra time Jonesboro used to phase in TESS appears to have had a positive effect on the teachers' perception of the system. Phasing in the system would also give policymakers time to adjust policies in light of evidence from the initial implementation about what is working as intended and what is not.

Taking a more phased-in approach can also improve ongoing implementation of TESS and LEADS. Rather than having new teachers undergo a full evaluation on all four domains as they begin teaching in districts, it would be more effective to have them initially concentrate on fewer domains. Through interviews, we learned that new teachers were overwhelmed by learning a new evaluation system on top of learning how to teach. Starting with only one or two domains might help new teachers learn the TESS process. Since interviewed

teachers and principals mentioned that they feel domains Two and Three were the most important, it might make sense to start formal evaluations on one or both of these domains instead of on all four. Of course, new teachers should be encouraged to remain mindful of the other domains on the TESS rubric. We also recommend that the school districts reduce other duties of new teachers and administrators as they move through their first TESS and LEADS cycles. As part of their induction to their new careers, teachers and principals should be able to focus on the most important parts of their jobs. Feiman-Nemser (2001) and Schwartz, Hernandez, and Ngo (2010) write about the importance of a reduced teaching load for new teachers.

A third lesson learned, that can help both future policy implementation and ongoing implementation of TESS and LEADS, is to be more mindful of how systems impact individuals. Training on and communication about TESS and how it affects teachers and administrators in general was a strength of implementation (this was less so for LEADS). However, there were significant differences in how the impacted on different individuals. One example is a lack of clarity about how the general rubric fit with particular teaching situations that did not match some of the rubric's implicit assumptions. Perhaps the most glaring example of this is the teachers' ongoing uncertainty about which track they are on, with 42.3 percent reporting they are still uncertain of their track placement. The issue of individualization also applies to alignment between teachers' identified needs and subsequent professional development. Individual and building level professional development needs to be better aligned with the highest priority developmental needs of teachers. At least in the case of building level professional development, the relationship between identified needs and selected professional development should be communicated more clearly. Ongoing implementation of TESS and LEADS can be improved by attending to these areas, and policymakers will do well in the future to bear in mind and prepare how new policies will affect individuals differently.

Separating Formative and Summative Functions

Finding: Disparate perceptions of how TESS and LEADS are used as summative and formative assessments compromise the

effectiveness of the process as a catalyst for professional improvement.

A central conclusion of our work is that combining the summative and formative functions into one system compromises the effectiveness of the system in achieving both goals. Both summative and formative evaluations are important. However, the dual nature of TESS has created unintended negative teacher perceptions about the tool. Ambiguity about whether negative job consequences can result from low TESS ratings has created a situation that in many ways is the worst of all possible scenarios. Administrators (and some teachers) doubt that the evaluation systems effectively remove low quality educators. However, the possibility of a teacher or administrator losing his/her job due to evaluation scores undermines the effectiveness of TESS and LEADS as development tools. Educators are hesitant to experiment with new methods and tend to stick with tried and true strategies in order to do well on their evaluations. The possibility of negative job consequences reinforces a "fixed" mindset in teachers; a "growth" mindset is needed for professional improvement.

The design of TESS attempted to avoid this by de-emphasizing the possibility of negative job consequences, but we find that the concern about the possibility of such consequences severely compromised the formative function. Within the bounds of the policy as it is designed, districts can attempt to address this concern by clarifying what would cause a teacher to receive negative job consequences, and then communicate this to teachers. If the possibility of negative job consequences is as remote as most administrators believe that it is, communicating this to teachers might lessen their concern about the possible cost of low ratings. However, if negative job action is as unlikely as it is perceived to be, this raises questions about the effectiveness of the summative aspect of the system. These questions are only reinforced by the low incidence of teachers receiving "unsatisfactory" or "basic" ratings.

Solving this problem more definitively would require policy redesign at the state level.

We recommend that at least the punitive aspect of the summative component of TESS and LEADS be removed entirely and replaced with a separate and effective accountability system. TESS and LEADS both try to de-emphasize the summative aspect of the systems so that they will be perceived as formative. However, in many

cases having any summative element at all fatally compromises the formative purpose. The district with the most critical ratings (Westside) has more negative perceptions of the system, supporting the notion that negative perceptions are driven by the possibility of adverse consequences. It is *that* possibility that makes the low ratings consequential. Moreover, with TESS (and perhaps also with LEADS), the accountability element is and looks likely to remain unrealized, so that the formative purpose is lost without the summative purpose being achieved. If the systems are to promote teacher and principal quality, they must be exclusively formative. We further recommend that any summative mechanism enhance rather than diminish the professionalism of those who ought to be accountable.

We recommend re-framing TESS and LEADS as developmental exercises. As such, it might make sense to not provide teachers with their scores above "one" or "two". Instead, evaluations should focus on areas of refinement and reinforcement. Interviews suggest that teachers are more focused on the numbers associated with their evaluations than they are with actual professional development. The TESS scoring system should still be used by evaluators for documentation, data collection, and perhaps also as an informal metric of whom to cultivate for later leadership positions. Teachers do not seem to benefit from knowing that they are a "three" in some areas or a "four" in others. However, teachers and administrators do need to know if their practice in certain areas is worse than what is expected. Therefore, low level ratings should still be communicated to teachers and administrators so that they know where to prioritize their improvement efforts.

We recommend training more teachers to conduct formative observations, and awarding those peer-evaluators a stipend. Peer-review would enhance teachers' sense of professionalism because it would be those within the profession making judgments about the performance of their colleagues. TESS results can help principals choose which teachers to train to become peer-evaluators. Teachers would see the training and added responsibility as a reward for excellent practice. We also feel that peer-evaluation would improve collaboration and community at schools. If observations are entirely formative (at least with respect to the absence of punitive measures), we feel that tensions surrounding

the process would be lessened, which would also lessen feelings of defensiveness.

Moreover, we recommend that teachers have the ability to choose their observer or evaluator. Teachers should be able to choose from peer-evaluators, principals, or outside content experts. It is important that teachers feel that their evaluators are credible and can offer insights into practices that can help teachers grow.

We recommend that all evaluators continue receiving instruction on observation and feedback practices as well as mentorship. Desired change is more likely to emerge from building a culture of collaboration and professionalism. This change will more organically stem from creating positive school professional cultures than by imposing a teacher quality system from the top down. In order to help this ethos of collaboration and community to emerge, school leaders should be trained about how to cultivate a collaborative and professional environment.

Finding a More Efficient use of Resources

Finding: The time demands associated with TESS hurt teacher perception of the process.

Another broad set of lessons for policymakers at the state and district level is to carefully consider the cost-benefit tradeoff and opportunity costs not only of the systems as a whole, but of all major elements of the system. In the case of TESS, we found that all perceived the system as time intensive, and while administrators tended to see this as partially justified (likely because the results of the process helped them to do their jobs better), teachers tended to view their TESS related activities as a poor use of time. Accordingly, we recommend that the state allow districts to use resources more efficiently, beginning with teacher time.

Our interviews show that both teachers and administrators tend to find the post-observation conference to be valuable, especially the comparison of the respective reflections on the lesson. This part of the system should be kept as it is. By contrast, we did not receive positive assessments of the other teacher requirements for TESS; therefore, these ought to be reconsidered.

We recommend that the requirement that teachers collect and submit artifacts be reduced or eliminated. This element exists to provide insight into aspects of teacher practice that observations do not capture, as well as to ensure that ratings are based on multiple sources of information. Nonetheless, in practice, we find that the

collection of artifacts is too time consuming and offers negligible value in promoting teacher growth; teachers report receiving little or no feedback on artifacts. If the submission of artifacts is *allowed* rather than required, teachers would still be able to use artifacts to demonstrate the quality of their planning and to help create a catalog of techniques and strong pedagogical ideas.

We further recommend that the requirement to conduct pre-conferences, and the documentation that goes with them, be scaled back or eliminated. There might be circumstances in which the observer needs to meet before an observation to understand what will happen in a class, and there might be circumstances where it would be beneficial to generate written records of these meetings. However, no interview subjects identified these processes or documents as being valuable for growth, and many described them as time consuming. As with artifacts, making them an option rather than a requirement would reduce the cost of TESS implementation without reducing the benefit.

TESS can be improved not only by streamlining the process, but also by reexamining the roles played by participants. One way to do this is to more effectively differentiate the process for different levels of teacher development, as Danielson (2011) recommends. TESS does this to a certain extent by having separate processes for novice and probationary teachers, and also notionally by having an intensive "support" track for teachers having trouble meeting expectations (the parallels between this and a traditional tenure structure are striking).

However, aside from this differentiation in intensity and frequency, the process itself remains largely the same irrespective of the level of teacher development. Our regression results show that the process is more likely to be seen as useful and as a good use of time and resources by that (relatively small) set of teachers who perceive (or recognize) that their current skills are not adequate to satisfy the demands of TESS (Only 1.1% of teachers "disagree" that they have adequate skills, and only 3.8% disagree more than they agree). This suggests that a formal process like TESS should be used in a way that prioritizes newer teachers or those who recognize their need for help. By narrowing the scope of TESS, more time and attention can be paid to making the process work, particularly by providing the meaningful feedback and support that all too often is currently displaced by observations and documentation.

For the strong majority of teachers who already see themselves as having the skills needed to meet the expectations of TESS (67.9 percent "agree" that they have these skills), it is unlikely that they will come to see the system as beneficial or worthwhile if they believe they already can do what the system purports to teach them. Even in the absence of the fixed mindset effects of the summative function, a formalized and still-hierarchical process is unlikely to convince the teachers who overrate their skills to improve. We agree with Danielson (2011) that teacher growth needs to be a career-long process; however, we also believe that for experienced teachers this is more likely to transpire as a function of it being a norm of a collaborative school culture and a value of a teacher professional community. Beyond ensuring that organizational structures and work processes are consistent with the emergence of such a dynamic, the aforementioned cultivation of leadership skills within the school community needed to foster that emergence is crucial.

Another way to redesign the roles of participants is to make better use of the expertise that is present within the existing set of teachers. We recommend that the school districts give more emphasis to or formalize the role of peer expertise in promoting professional growth. Many teachers mentioned that the guidance that they find to be most useful is what they receive from peers whom they respect. Several principals mentioned using TESS results to pair a teacher who has a specific identified need with another teacher who is strong in that area. This is the type of peer learning that happens in a well-functioning school community. Yet, TESS does not appear to have a mechanism for such activities to count toward the requirement that teachers have eighteen hours of professional development aligned to their PGP. We recommend that schools and districts consider finding ways to credit peer learning toward this requirement, in part to make the professional growth guidance that teachers receive more timely and better aligned with actual needs (and in a way more informed about local needs), while also further encouraging such peer learning. This approach would also further the broader goal of creating a culture of collaboration and professionalism within each school.

Conclusion

Implementing new systems of teacher and administrator evaluation is a

complex process that affects all stakeholders. Jonesboro School District, Nettleton School District, and the Westside Consolidated School District have all recently implemented the TESS and LEADS systems. In order for a policy implementation of this magnitude to be considered successful, careful attention needs to be paid to the mechanics of the implementation, the quality of the processes, and how the processes affect the culture at the individual schools.

While principals have a more positive view of TESS than teachers do, educator perceptions on the amount of time the evaluation system takes is highly problematic. The implementation of TESS has affected teachers' perceptions of professionalism and school culture. LEADS is less a part of their daily lives, and remains an afterthought to most principals.

We offer a series of recommendations focusing on improving implementation, improving the quality of the TESS process by separating the formative and summative purposes of evaluations, and finding a better use of resources. We believe these recommendations can help TESS and LEADS meet their potential as school improvement tools. We also feel that these slight modifications will improve the professional and overall culture in schools.

References

Ashby, J. Frank, R. and McClain. An Early Implementation Study of the Arkansas Teacher Excellence and Support System (TESS): Research, Analysis, and Recommendations (unpublished capstone project). Vanderbilt University, Nashville, TN.

Beck, L., & Foster, W. (1999). Administration and community: Considering challenges, exploring possibilities. In J. Murphy & K.S. Louis (Eds.), *Handbook of research on educational administration* (pp. 337-358). San Francisco: Jossey-Bass.

Cannata, M., Taylor Haynes, K. & Smith, T.M. (2014). Integrating Academic Press and Support by Developing student Ownership and Responsibility. Paper presented at the annual meeting of the American Educational Research Association in Philadelphia, PA, April 3-7, 2014.

CityData, Jonesboro, Arkansas. (2016). Retrieved January 3, 2016, from http://www.city-data.com/city/Jonesboro-Arkansas.html. Advameg, Inc.

Clipa, O. (2011). Teacher perceptions on teacher evaluation: The purpose and the assessors within the assessment process. *Procedural-Social and Behavioral Sciences*, *29*, 158-163.

Coburn, C.E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, *32*(6), 3-12.

Crowson, R., & Goldring, E. (2009). The new localism: Re-examining issues of neighborhood and community in public education. Yearbook of the National Society for the Study of Education 108 (1), 1–24.

Curtis, R. E., & Wurtzel, J. (2010). Teaching talent: A visionary framework for human capital in education. Cambridge, MA: Harvard Education Press.

Danielson, C. (2011). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, Va: Association for Supervision and Curriculum Development

Danielson, C. & McGreal, T. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies That Support Professional Development in an Era of Reform. *Phi Delta Kappan, 76,* 8, 597-604.

Darling-Hammond, L. (2012). Creating a comprehensive system for evaluating and supporting effective teaching. Stanford, CA. Stanford Center for Opportunity Policy in Education.

Datnow, A. & Park, V. (2009) Conceptualizing Policy Implementation. In Sykes, G., Schneider, B. L., Plank, D. N., & Ford, T. G.. *Handbook of education policy research*. New York: Routledge.

Desimone, L.M. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research, 72(3),* 433-479.]

Duke, D. L. (1990). Developing teacher evaluation systems that promote professional growth. *Journal of Personnel Evaluation in Education, 4*, 131-144.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York: Random House.

Feiman-Nemser, S. (2001). From preparation to practice: designing a continuum to strengthen and sustain teaching. *Teachers College Record, 103,* 6, 1013-55.

Firestone, W.A., Blitz, C. L., Gitomer, D.H., Kirova, D., Shcherbakov, A., & Nordon, T.L. (2013). New Jersey Teacher Evaluation, RU-GSE External Assessment: Year 1 Report. Retrieved from: http://www.nj.gov/education/AchieveNJ/resources/RUGSE11-12.pdf

Fixsen, D.L., Naoom, S. F., Blasé, K.A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Goe, L., Biggers, K., & Croft, A. (2012). Linking teacher evaluation to professional development: Focusing on improving teaching and learning. Research & Policy Brief. *National Comprehensive Center for Teacher Quality*.

Goggin, M. L. (1986). The "Too Few Cases/Too Many Variables" problem in implementation research. The Western Political Quarterly, 39(2), 328

Grant, A. M., & Parker, S. K. (2009). Redesigning Work Design Theories: The Rise of Relational and Proactive Perspectives. *The Academy of Management Annals : a Journal of the Academy of Management, 3,* 317.

Grant, A. M., & Shin, J. (2012). Work motivation: Directing, energizing, and maintaining effort (and research). In Ryan, R. M. *The Oxford handbook of human motivation*. New York: Oxford University Press.

Greenberg, E., Rhodes, D., Ye, X. & Stancavage, F. (2004, April 12-16). Prepared to teach: Teacher preparation and student achievement in eighth-grade mathematics. Paper presented at the American Educational Research Association, San Diego.

Grossman, M. (2006). Education and Nonmarket Outcomes. *Handbooks in Economics, 26,* 1, 577-634

Honig, M. I. (2006). Complexity and policy implementation: challenges and opportunities for the field. In M.I. Honig (Ed.), *New directions in education policy implementation: confronting complexity*. Albany: State University of New York Press.

Hull, J. (2013). Trends in Teacher Evaluation. Center for Public Education. Retrieved from http://www.centerforpubliceducation.org/Main-Menu/Evaluating-performance/Trends-in-Teacher-Evaluation-At-A-Glance/Trends-in-Teacher-Evaluation-Full-Report-PDF.pdf

Johnson, S. M. (2015). Will VAMS Reinforce the Walls of the Egg-Crate School?. *Educational Researcher, 44,* 2, 117-126.

Johnson, S. M., & Fiarman, S. E. (2012). The potential of peer review. *Educational Leadership, 70*(3), 20-25. Retrieved from http://www.ascd.org/publications/ educational-leadership/nov12/ vol70/num03/The-Potential-of-Peer-Review.aspx

Johnson, R. B. & Onwuegbuzie, A.J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14-26.

Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision-making within a standards-based evaluation system. *Educational Administration Quarterly*, *45*(1), 34-70.

Lee, V. E., Dedrick, R. F., & Smith, J. B.. (1991). The Effect of the Social Organization of Schools on Teachers' Efficacy and Satisfaction. *Sociology of Education*, *64*(3), 190–208. http://doi.org.proxy.library.vanderbilt.edu/10.2307/2112851

Lortie, D. (1969). The partial professionalization of elementary teaching. In Etzioni, A. (ed.). *The semi-professions and their organization: Teachers, nurses, social workers*. New York: Free Press.

Minor, E. C., , Porter, A. C., Murphy, J., Goldring, E. B., Cravens, X., & Elloitt, S. N. (2014). A known group analysis validity study of the Vanderbilt Assessment of Leadership in Education in US elementary and secondary schools. *Educational Assessment, Evaluation and Accountability, 26,* 1, 29-48.

Murphy, J., Hallinger, P., & Heck, R. H. (2013). Leading via teacher evaluation: The case of the missing clothes? *Educational Researcher*, *42*(6), 349-354.

Murphy, J. (2013). *The architecture of school improvement: Lessons learned*. Thousand Oaks, CA. Corwin

National Center for Teacher Quality (2011). State of the states: Trends and early lessons on teacher evaluation and effectiveness policies. Retrieved from http://www.nctq.org/dmsView/State_of_the_States_Teacher_Evaluation_and_Effectiveness_Policies_NCTQ_Report

National Council of Teacher Quality. (2012). State of the States 2012: Teacher Effectiveness Policies - Area 3: Identifying Effective Teachers; NCTQ State Teacher Policy Yearbook Brief. Retrieved March 15, 2016, from http://www.nctq.org/dmsStage/State_of_the_States_2012_Teacher_Effectiveness_Policies_NCTQ_Report

National Council of Teacher Quality. (2013). State of the States 2013 Connect the Dots: Using Evaluations of Teacher Effectiveness to Inform Policy and Practice. Retrieved March 15, 2016, from

http://www.nctq.org/dmsStage/State_of_the_States_2013_Using_Teacher_Evaluations_NCTQ_Report

National Commission on Teaching & America's Future. (1996). *What matters most: Teaching For America's future*. New York, New York.

New Teacher Project. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, N.Y.: The New Teacher Project.

Office of Educator Effectiveness. (n.d.). Arkansas TESS and LEADS Resources. Retrieved November/December, 2015, from http://www.arkansased.gov/public/userfiles/HR_and_Educator_Effectiveness/TESS/Handbook6-80968.0.pdf

Porter, A. C., Floden, R., Freeman, D., Schmidt, W., & Schwille, J. (1988). Content determinants in elementary school mathematics. In D. Grouws & Thomas Cooney (Eds.), Perspective on research on effective mathematics teaching. Reston, VA: National Council of Teacher of Mathematics.

Porter, A. C. (1994). National standards and school improvement in the 1990s: Issues and promise. American Journal of Education, 102, 421-449

Rockoff, J. E. (January 01, 2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review, 94,* 2, 247-252.

Sartain, L., Stoelinga, S. R., Brown, E., Luppescu, S., Matsko, K. K., & Miller, F.K. (2011, November). Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation. Chicago: University of Chicago Consortium on Chicago School Research. Retrieved from http://ccsr.uchicago.edu/publications/rethinking-teacher-evaluation-chicago- lessons-learned-classroom-observations-principal

Schmoker, M. (2004). Tipping Point: From Feckless Reform to Substantive Instructional Improvement. *Phi Delta Kappan, 85,* 6, 424.

Schwartz, R., & Ngo, J. (2010). Attracting and Retaining Strong Teachers. In M. Hernandez (Ed.), Teaching talent: A visionary framework for human capital in education. Cambridge, MA: Harvard Education Press.

Scribner, J. P., Hager, D. R. & Warne, T. R. (2002). The paradox of professional community: Tales from two high schools. Educational Administration Quarterly, 38, 45-76.

Sergiovanni, T. J. (1994). Organizations or Communities? Changing the Metaphor Changes the Theory. *Educational Administration Quarterly, 30,* 2, 214-26.

Sieber, S. D. (1973). The integration of fieldwork and survey methods. *American Journal of Sociology, 73,* 1335-1359.

Spillane, J., Reiser, B., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Researcher 72*(3), 387-431.

Stronge, J. H., & Tucker, P. D. (1999). The politics of teacher evaluation: A case study of new system design and implementation. *Journal of Personnel Evaluation in Education*, *13*(4), 339-359.

Taylor, E. S., & Tyler, J. H. (2011). The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers. NBER Working Paper 16877. Cambridge, MA: National Bureau of Economic Research.

Tönnies, F., & Harris, J. (2001). *Community and civil society*. Cambridge: Cambridge University Press.

U.S. Census website (2016). Jonesboro, Arkansas. American FactFinder - Community Facts. (n.d.). Retrieved February 16, 2016, from http://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml

U.S. Department of Education (2010). A blueprint for reform: The realignment of the elementary and secondary education act. Retrieved from http://www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf

USNews.Com. (2015). Westside High School Overview. Retrieved February, 2016, from http://www.usnews.com/education/best-high-schools/arkansas/districts/westside-cons-school-district/westside-high-school-1266

Wahlstrom, K., & Louis, K.S. (2008). How teachers perceive principal leadership. *Educational Administration Quarterly*, *44*, 498–445.

White, B.R., Cowhy, J., Stevens, W.D., & Sporte, S.E. (2012). Designing and implementing the next generation of teacher evaluation systems: Lessons learned from case studies in five Illinois districts. Consortium on Chicago School Research: Chicago, IL. Retrieved from

http://ccsr.uchicago.edu/publications/designing-and-implementing-next- generation-teacher-evaluation-systems-lessons-learned.

Appendix 1: Interview Protocol for Teachers

Interviewer:_____ Subject:_____

ICEBREAKERS

What do you teach?
How long have you been teaching?
What is something you've enjoyed about working in this district?

PERCEPTION

1. Since the implementation of the TESS evaluation system have you received yearly training? If so, how helpful do you feel the training has been?
2. What are the expectations of teachers in this evaluation process? What do you have to do?
   How have these expectations changed over the life of the program?

3. Do you feel prepared to meet these expectations?
4. Let's take a look at the rubric. What are your thoughts in general, about the rubric that is used to evaluate you? How well or poorly does it capture effective teaching practice?
5. Principals have a key role in the process….How well prepared do you feel a principal is to

observe, evaluate, and provide you with feedback through the TESS process? Has your principal grown more skilled and/or otherwise effective over the past three years?

6. Do you think an outside evaluator would be more objective? Would their lack of familiarity with you and the school context impair their ability to evaluate? Overall, do you think you would prefer an outside evaluator?

7. Do you feel free to experiment during evaluation to get feedback on how to improve? Do you play it safe and use a tried and true lesson for when you are evaluated?

8. So, let me ask you: Is there one particular part of the process you find most useful to you in your professional growth? Why do you think this?

9. How has TESS impacted your day-to-day work inside and outside the classroom?
10. How do you think the TESS teacher evaluation system has impacted your relationship with your principal?
11. How do you think the TESS teacher evaluation system has impacted your relationship with your fellow teachers?
12. How do you think the TESS teacher evaluation system has informed or impacted your professional development?
13. How do you think this new system has informed or impacted your curriculum?
14. How do you think the TESS teacher evaluation system has informed or impacted your instructional practices?
15. How do you think the TESS teacher evaluation system has impacted/informed students achievement?
16. What could be done to improve the usefulness of the teacher observation system?

CAPACITY

1. What sort of supports have you found to help you to be successful implanting this system? How have these supports changes over the past three years?
2. What sort of supports has the district provided in terms of extra time, resources, training? How have these changed over the past three years?
3. How would you describe the quality of the professional development you've received in preparing you for the TESS eval. Process? How has this changed over the past three years? Any examples?
4. What are some questions you still have about your role during the teacher evaluation process?
5. What are some challenges in terms of understanding the teacher evaluation procedures and expectations?
   In terms of following the process according to the specifics of the model, plan/procedures?
   Has the ease of understanding and implementing the process changed over the past three years?
6. So, let's talk about staff meetings at your school. How often do you have staff meetings at your school? What do you normally cover/discuss during these meetings?
7. Do you talk about TESS? What is the general focus of these conversations during faculty meetings? What are some of the topics or questions that teachers raise?
   Do you talk about TESS and the different steps and procedures required?
8. How frequently do you have professional conversations with your administrator and with the staff about teacher quality and student achievement?
   Has the frequency and quality of conversation increased under the new system?

Appendix 2: Interview Protocol for Administrators

Interviewer:_____ Subject:_____

ICEBREAKERS

How long have you been in education?
What is something you've enjoyed about working in this district?

PERCEPTION

1. After all of your initial training in regards to TESS, how successful do you feel the system has been?
   Has this level of success changed in the past three years, and if so how and why?
2. How would you describe the quality of the professional development you've received in implementing TESS? Has this improved in the past three years?
3. As a school administrator you have to balance many roles and met a great number of requirements. Successfully preparing teachers for an evaluation system requires a specific set of leadership characteristics and supports.
   How successful do you believe you have been in preparing teachers to continue using the TESS evaluation?
4. How have expectations placed on your for the TESS system changed over the past three years?
5. What aspects of TESS represent improvements in teacher evaluation practice over the old system?

What if any flexibility do you have to adjust implementation to fit your school/district and community?
What's your assessment of the new roles and expectations?
Can you assign it a grade (A to F)? Why?

6. How comfortable are you in using the TESS evaluation? Have you grown more confortable in the past three years?
7. Let's take a look at the rubric for a moment. What are your thoughts specifically about the rubric that is used to evaluate the teachers (probe: the one with the four domains)?
   How well or poorly does it represent effective teaching practice?
8. How well prepared do you feel, as a principal, to observe, evaluation, and provide teachers with feedback through the TESS process?
   Has this changed appreciably in the past three years?

9. How about your ability to use it correctly and in a timely manner?
10. Is there a tension between your responsibility to help teachers improve and to evaluate their performance?
    Do you think outside evaluators would find it easier to be candidly critical?
11. Which part/aspect of this process would be most useful to your teachers' professional growth?
    Why do you think this?
12. Tell me: How does Tess impact your day-to-day work? Has this changed over the past three years?
13. How has this new teacher evaluation system impacted your relationship with your teachers?
14. How has TESS changed what is expected of you as a principal?
15. How do you think this new system has impacted/informed student achievement?
16. What could be done to improve the usefulness of the teacher observation system?

Let's briefly talk about LEADS:

17. After all of your initial training in regards to LEADS, how successful do you feel the new system has been?
18. Do you feel that LEADS captures effective principal practice, and do you feel like the measurements used are effective?
19. How well or poorly placed is your evaluator to reach conclusions about what you do?
20. How well or poorly do you expect LEADS to help you develop professionally?
21. How would you describe the quality of the professional development you've received in implementing the LEADS rollout?
22. How has the implementation of the LEADs system affected your participation in the TESS teacher evaluation system?

<u>CAPACITY</u>

1. How well placed was your school to implement these systems in terms of the professional expertise of staff, ability to allocate time to implementation, and other resources or capacities?
   How have these changed over time?

2. What sort of support do you need to successfully use TESS?
3. What sort of support do you need to successfully respond to the requirements of LEADS?
4. What sort of supports is the district providing you in terms of extra time, resources, and

training to be an instructional leader? How have these changed over the past three years?
5. What are some questions you still have about your role during the teacher evaluation?
6. The TESS system is a fairly new system for teacher evaluation. What were some personal challenges in terms of understanding and following the teacher evaluation procedures and expectations?
7. How has the feedback you give to teachers changed since the implementation of TESS?
8. What contributes to or undermines the accuracy of your facilitation of the teacher observation system?

9. What are some supports you are getting to help ensure that you give *accurat*e teacher observation scores?
10. I have a technical question here.
What systems are in place to help you store and retrieve teacher observation data?
11. How do you plan to use teacher observation data to inform individual growth plans and professional development at your school?
12. How is the teacher observation system facilitating or impeding collaboration among educators in this district?
13. How do you feel the rollout of the LEADS evaluation system was similar or different to the rollout of the TESS system a couple of years ago?
14. How consistent are TESS and LEADS with other national, state, and local ed policies? In what ways? How durable do you expect TESS and LEADS to be as policies?

15. Any other issues that you would like to address that I didn't cover?

Appendix 3: Teacher Survey Protocol

Teacher

Q1 1. Select School District
- Jonesboro (1)
- Nettleton (2)
- Westside (3)

Q2 How many total years have you been in education?
- 1-3 (1)
- 4-6 (2)
- 7-10 (3)
- 10-20 (4)
- 21-30 (5)
- 30+ (6)

Q3 Please select your school's configuration from the following list:
- Elementary School (1)
- Intermediate School (2)
- Middle School (3)
- High School (4)
- Other (please specify) (5) _____

Q4 Which evaluation track are you on?
- 1 (1)
- 2A (2)
- 2B1 (3)
- 2B2 (4)
- Uncertain (5)

Q5 Have you had at least one formal evaluation with a pre-conferene and post-conference this school year?
- Yes (1)
- No (2)

Q6 I understand what is expected of me in each of the domains and sub-domains of the rubric.
- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q7 I can accurately describe to others the processes and procedures by which I will be evaluated (ie. The number of observations, artifact collection, and other related paperwork).
- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q8 I feel adequately informed about the TESS evaluation system
- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q9 Expectations have been communicated clearly and consistently.
- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q10 The overall quality of training I have received has been _____.

- Very Poor (1)
- Poor (2)
- Fair (3)
- Good (4)
- Very Good (5)

Q11 How many total hours of TESS training did you have during the 2013-2014 school year? Please include the online modules, district training events, Co-op training events, and other professional development opportunities for administrators.

- 0-10 (1)
- 11-20 (2)
- 21-30 (3)
- 31-40 (4)
- 41+ (5)
- I was not employed by this district during the 2013-2014 school year (6)

Q12 How many total hours of TESS training did you have during the 2014-2015 school year? Please include the online modules, district training events, Co-op training events, and other professional development opportunities for administrators.

- • 0-20 (1)
- • 21-40 (2)
- • 41-60 (3)
- • 61-80 (4)
- • 81-100 (5)
- • 100+ (6)
- I was not employed by this district during the 2014-2015 school year (7)

Q13 How many total hours of TESS training so far during the 2015-2016 school year? Please include the online modules, district training events, Co-op training events, and other professional development opportunities for administrators.

- • 0-20 (1)
- • 21-40 (2)
- • 41-60 (3)
- • 61-80 (4)
- • 81-100 (5)
- • 100+ (6)

Q14 I am prepared to carry out the following aspects of TESS:

	Strongly Agree (1)	Agree (2)	Uncertain (3)	Disagree (4)	Strongly Disagree (5)
Collect and document artifacts for each of the four domains (1)	•	•	•	•	•
Complete paperwork for pre-and post-conferences (2)	•	•	•	•	•
Develop lesson plans that incorporate principles from the "Planning and Preparation"	•	•	•	•	•
Implement instructional practices that effect principles from the	•	•	•	•	•
Create a classroom environment that reflects principles from the "Classroom Environment" domain (5)	•	•	•	•	•
Choose and fulfill the duties under the "Professional Responsibilities" domain (6)	•	•	•	•	•

Q15 Which of the following apply to you?

- I am a National Board Certified Teacher (1)
- I have undergone Pathwise Training (2)
- I serve/served as a Pathwise mentor (3)

Q16 What is the highest degree you have received?

- Bachelor's (1)
- Master's (2)
- Ed.S (3)
- Doctorate (Ed.D. or Ph.D.) (4)

Q17 Overall, I think the TESS evaluation system has had a positive impact on my own teaching practice.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q18 Overall, I think the new teacher evaluation system has had a positive impact on student achievement in my school.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q19 I believe the feedback given to me through the TESS process has helped improve my teaching

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q20 The following domains of the new evaluation system rubric accurately reflect effective teacher practices:

	Strongly Agree (1)	Agree (2)	Uncertain (3)	Disagree (4)	Strongly Disagree (5)
Planning and preparation	•	•	•	•	•
Classroom Environment (2)	•	•	•	•	•
Instruction (3)	•	•	•	•	•
Professional Responsibilities (4)	•	•	•	•	•

Q21 The new teacher evaluation system fits well with other school/district initiatives (such as implementing Common Core and other schoolwide curricular/policy changes).

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q22 The new teacher evaluation system consumes time and resources that could be better spent elsewhere.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q23 I believe that the obligations of TESS interferes with my ability to carry out other teaching responsibilities.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q24 There is a great deal of trust between administrators and teachers in this school.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q25 There is a great deal of teacher collaboration at our school

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q26 The new teacher evaluation system is helping me collaborate with my colleagues as part of a professional learning community

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q27 The quality and frequency of professional conversations with colleagues has increased under the new teacher evaluation system.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q28 Feedback from my teacher evaluation informs the professional development activities in which I participate.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q29 I have access to adequate support to improve areas of refinement identified in my teacher evaluations.

- Strongly Agree (1)
- Agree (2)
- Neither Agree nor Disagree (3)
- Disagree (4)
- Strongly Disagree (5)

Q30 Click to write the question text

	Strongly Agree (1)	Uncertain (2)
Have frequent, shorter observations rather than one long observation (1)	•	•
Use multiple raters and observers (2)	•	•
Incorporate students' standardized test scores (3)	•	•
Incorporate teacher Peer Ratings (4)	•	•
Incorporate student surveys	•	•
f. Incorporate parent surveys	•	•
g. Other (please specify) (7)	•	•

Q31 In terms of training and support with TESS, what are some ways you think the evaluation process can be improve

- Opportunities to observe a Level 4 teacher in your district (1)
- Having a district liaison assigned to school site for advising on TESS procedures (2)
- On line access to sample artifacts from other teachers' classrooms in your district (3)
- Face to face PD work sessions related to planning and preparation domain (4)
- Face to face PD work sessions related to instruction domain (5)
- Face to face PD work sessions related to classroom environment (6)
- Peer walk throughs at school site with debriefings to better understand scoring of formal evaluation (7)
- h. Other (please specify) (8) _____

Q32 Generally speaking, what BENEFITS have you encountered with the new teacher evaluation system this school year?

- Click to write Choice 1 (1) _____

Q33 The new teacher evaluation system fits well with other school/district initiatives (such as implementing Common Core and other school wide curricular/policy changes).

- Click to write Choice 1 (1) _____

Appendix 5: Administrator Survey Protocol

Admin Survey

Q1 Select School District
- Jonesboro (1)
- Nettleton (2)
- Westside (3)

Q2 How many teachers have you formally evaluated this year (including holding the pre- and post-conferences)?
- 0 (1)
- 1-5 (2)
- 6-10 (3)
- 11-15 (4)
- 16-20 (5)
- 20+ (6)

Q3 On average, how many hours each week do you spend on TESS-related duties?
- 0 (1)
- 1-3 (2)
- 4-6 (3)
- 7-9 (4)
- 10+ (5)

Q4 I can accurately describe to others the processes and procedures used to conduct teacher evaluations.
- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q5 The state of Arkansas has clearly and consistently communicated expectations about TESS.
- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q6 My district has clearly and consistently communicated expectations about TESS.
- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q7 I believe that I have received adequate training to perform my expected role under the new teacher evaluation system.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q8 The quality of training I have received has been _____. Click to write the question text

- Very Good (1)
- Good (2)
- Fair (3)
- Poor (4)
- Very Poor (5)

Q9 How many total hours of TESS training have you had during the 2013-2014 school year? Please include the online modules, district training events, Co-op training events, and other professional development opportunities for administrators.

- 0-20 (1)
- 21-40 (2)
- 41-60 (3)
- 61-80 (4)
- 81-100 (5)
- 100+ (6)

Q10 How many total hours of TESS training have you had during the 2014-2015 school year? Please include the online modules, district training events, Co-op training events, and other professional development opportunities for administrators.

- 0-20 (1)
- 21-40 (2)
- 41-60 (3)
- 61-80 (4)
- 81-100 (5)
- 100+ (6)

Q11 How many total hours of TESS training have you had so far during the 2015-2016 school year? Please include the online modules, district training events, Co-op training events, and other professional development opportunities for administrators.

- 0-20 (1)
- 21-40 (2)
- 41-60 (3)
- 61-80 (4)
- 81-100 (5)
- 100+ (6)

Q12 I am prepared to carry out the following aspects of TESS:

	Strongly Agree (1)	Agree (2)	Uncertain (3)	Disagree (4)	Strongly Disagree (5)
Accurately rate teachers using the TESS rubric	•	•	•	•	•
Accurately assess the suitability of artifacts for all four domains (2)	•	•	•	•	•
Conduct teacher conferences	•	•	•	•	•
Complete all TESS-related paperwork	•	•	•	•	•
Preparing or leading professional development at my school site (5)	•	•	•	•	•
Reviewing date from different classroom assessments across the school (6)	•	•	•	•	•
Attend outside professional development important for my growth as an administrator	•	•	•	•	•

Q13 To what extent has time spent on TESS-related tasks impacted the amount of time you have for the following:

	Greatly Impacted (1)	Somewhat Impacted (2)	Slightly Impacted (3)	No Impact (4)
Student discipline issues	•	•	•	•
Casual classroom walkthroughs unrelated to the TESS requirements (2)	•	•	•	•
Interacting with students (3)	•	•	•	•
Attending parent-teacher or other student-related conferences or meetings (4)	•	•	•	•
Completing other state or district required paperwork and tasks unrelated to TESS (5)	•	•	•	•
Time to reflect	•	•	•	•

Q14 1. I am confident in my ability to _____.

	Strongly Agree (1)	Agree (2)	Uncertain (3)	Disagree (4)	Strongly Disagree (5)
Provide substantive feedback (1)	•	•	•	•	•
Coach teachers on each of the four domains	•	•	•	•	•
Have critical conversations with teachers regarding their performance	•	•	•	•	•
Identify professional development and support for specific teachers based on their evaluations results (4)	•	•	•	•	•
Determine what type of professional development would be most beneficial for my school based on teacher effectiveness	•	•	•	•	•

Q15 How many years have you been an administrator?

- 1-3 (1)
- 4-6 (2)
- 7-10 (3)
- 11-20 (4)
- 21-30 (5)
- 30+ (6)

Q16 How many years of teaching experience did you have prior to becoming an administrator?

- 0 (1)
- 1-3 (2)
- 4-6 (3)
- 7-10 (4)
- 11-20 (5)
- 21-30 (6)
- 30+ (7)

Q17 Overall, I think the TESS system has had a positive impact on the quality of instruction in my school.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q18 Overall, I think the new teacher evaluation system has had a positive impact on student achievement in my school.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q19 The new teacher evaluation system fits well with other school/district initiatives (i.e. Common Core and other schoolwide curricular/policy changes).

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q20 The new teacher evaluation system consumes resources that could be better spent on promoting key district improvement initiatives (i.e. Common Core and other schoolwide curricular/policy changes).

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q21 I believe that the obligations of TESS interfere with my ability to support other programs and policies.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q22 The TESS system helps me to have better conversations with my teachers about effective instruction.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q23 I have resources that I can recommend and/or provide to teachers who need to improve their performance.

- Strongly Agree (1)
- Agree (2)
- Uncertain (3)
- Disagree (4)
- Strongly Disagree (5)

Q24 Administrators should be able to use teacher evaluation results in making decisions about _____.

	Strongly Agree (1)	Agree (2)	Uncertain (3)	Disagree (4)	Strongly Disagree (5)
Hiring (1)	•	•	•	•	•
Promotion (2)	•	•	•	•	•
Intra-District Transfers (3)	•	•	•	•	•
Teacher Pay (4)	•	•	•	•	•
Student Assignment	•	•	•	•	•

Q25 To what extent would you like more support and training around the use of teacher evaluation data in the following specific areas:

	In Depth Training (1)	Refresher Training (2)	No Further Training Needed (3)
Accurately assessing the suitability of artifacts for all four	•	•	•
Accurately rating teachers using the TESS rubric (2)	•	•	•
Conducting teacher conferences (3)	•	•	•
Coaching Teachers in aspects of each of the four domains (4)	•	•	•
Having critical conversations with teachers regarding their performance (5)	•	•	•
Identifying professional development and support for specific individuals based on their evaluations	•	•	•
Using teacher effectiveness data to determine what type of professional development would be most beneficial for your school (7)	•	•	•
Other (please specify) (8)	•	•	•

Q26 How are you primarily keeping track of artifacts and the observation cycles (i.e. Bloomboard, GoogleDocs, LiveBinder, Combination of Paper and Computer Records, Paper Records Only...)

- Click to write Choice 1 (1) _____

Q27 OPTIONAL: Generally speaking, what BENEFITS have you encountered with the TESS evaluation system this school year?

- Click to write Choice 1 (1) _____

Q28 Generally speaking, what CHALLENGES have you encountered with the TESS evaluation system this school year?

- Click to write Choice 1 (1) _____

Q29 How well selected are the measures used in LEADs?

- extremely poorly (1)
- poorly (2)
- fairly (3)
- well (4)
- extremely well (5)

Q30 Please rank domains from most important to least important to administrator practice

_____ Vision, Mission, and Goals (1)
_____ Teaching and Learning (2)
_____ Managing Organizational Systems and Safety (3)
_____ Collaborating with Families and Stakeholders (4)
_____ Ethics and Integrity (5)
_____ The Education System (6)

Q31 Which specific measures within the domains are most important?

Q32 Which specific measures within the domains are least important?

Q33 Which elements of the LEADS evaluation are easiest for your evaluator to gather data about your performance in answering?

Q34 Which elements of the LEADS evaluations are most difficult for your evaluator to gather data about your performance in answering?

Q35 Overall, is it easy or hard for your evaluator to gather sufficient and accurate data about your performance?

- Easy (1)
- Neutral (2)
- Difficult (4)

Q36 Evaluators collect a variety of data points in reaching conclusions about each aspect of their ratings. In combining these data points into an overall score, do you think they give appropriate weight to each data point, or do you think they overemphasize some things while not weighing others heavily enough?

- Weigh appropriately (1)
- Make errors in weighing (2)
- N/A or Don't know (3)

Q37 Has the LEADS framework facilitated better feedback about your perfamance from your evaluator?

- Yes (1)
- No (2)

Q38 Do you see the purpose of LEADs as being more formative or summative?
- Formative (1)
- Summative (2)

Q39 How has being evaluated by the LEADS process affected you you conduct evaluations in the TESS process?

Q41 Which LEADS activities are you most prepared to carry out?

Q42 Which LEADS activities are you least prepared to carry out?

Q43 Overall, LEADS will have a positive impact on my administrative practice
- Strongly Agree (1)
- Agree (2)
- Neither Agree nor Disagree (3)
- Disagree (4)
- Strongly Disagree (5)
- Click to write Choice 6 (6)

Q44 How many total years have you been in education?
- 1-3 (1)
- 4-6 (2)
- 7-10 (3)
- 10-20 (4)
- 21-30 (5)
- 30+ (6)

Appendix 6: Teacher Scale Items and Factor Loadings

Teacher Scale Items and Factor Loadings			
TESS Skill		*TESS Beneficial*	
Item	Factor	Item	Factor

	Loadings		Loadings
I am prepared to carry out the following aspects of TESS: Collect and document artifacts for each of the four domains	0.545	Overall, I think the new teacher evaluation system has had a positive impact on student achievement	0.659
I am prepared to carry out the following aspects of TESS: Complete paperwork for pre- and post-conferences	0.641	I believe the feedback given to me through the TESS process has helped improve my teaching	0.574
I am prepared to carry out the following aspects of TESS: Develop lesson plans and incorporate principles from the "Planning and Preparation" domain	0.791	The following domains of the new evaluation system rubric accurately reflect effective teacher practice: Planning and Preparation	0.836
I am prepared to carry out the following aspects of TESS: Implement instructional practices that effect principles from the "Instruction" domain	0.805	The following domains of the new evaluation system rubric accurately reflect effective teacher practice: Classroom Environment	0.821
I am prepared to carry out the following aspects of TESS: Create a classroom environment that reflects principles from the "Classroom Environment" domain	0.84	The following domains of the new evaluation system rubric accurately reflect effective teacher practice: Instruction	0.84
I am prepared to carry out the following aspects of TESS: Choose and fulfill the duties under the "Professional Responsibilities" domain	0.774	The following domains of the new evaluation system rubric accurately reflect effective teacher practice: Professional Responsibilities	0.844
		The new teacher evaluation system fits well with other school/district initiatives	0.642
TESS Training		*TESS Collab/Development*	
Item	Factor Loadings	Item	Factor Loadings
I understand what is expected of me in the domains and sub-domains of the rubric	0.673	There is a great deal of trust between administrators and teachers in this school	0.477
I can accurately describe to others the processes and procedures by which I will be evaluated	0.753	There is a great deal of teacher collaboration at our school	0.691
I feel adequately informed about the TESS evaluation system	0.766	The new teacher evaluation system is helping me collaborate with my colleagues	0.731
Expectations have been communicated clearly and consistently	0.675	The quality and frequency of professional conversations with colleagues has increased under the new system	0.671
The overall quality of the training I have received has been (?)	0.701	Feedback from my teacher evaluation informs the professional development activities in which I participate	0.595
		I have access to adequate support to improve areas of refinement identified in my teacher evaluation.	0.503

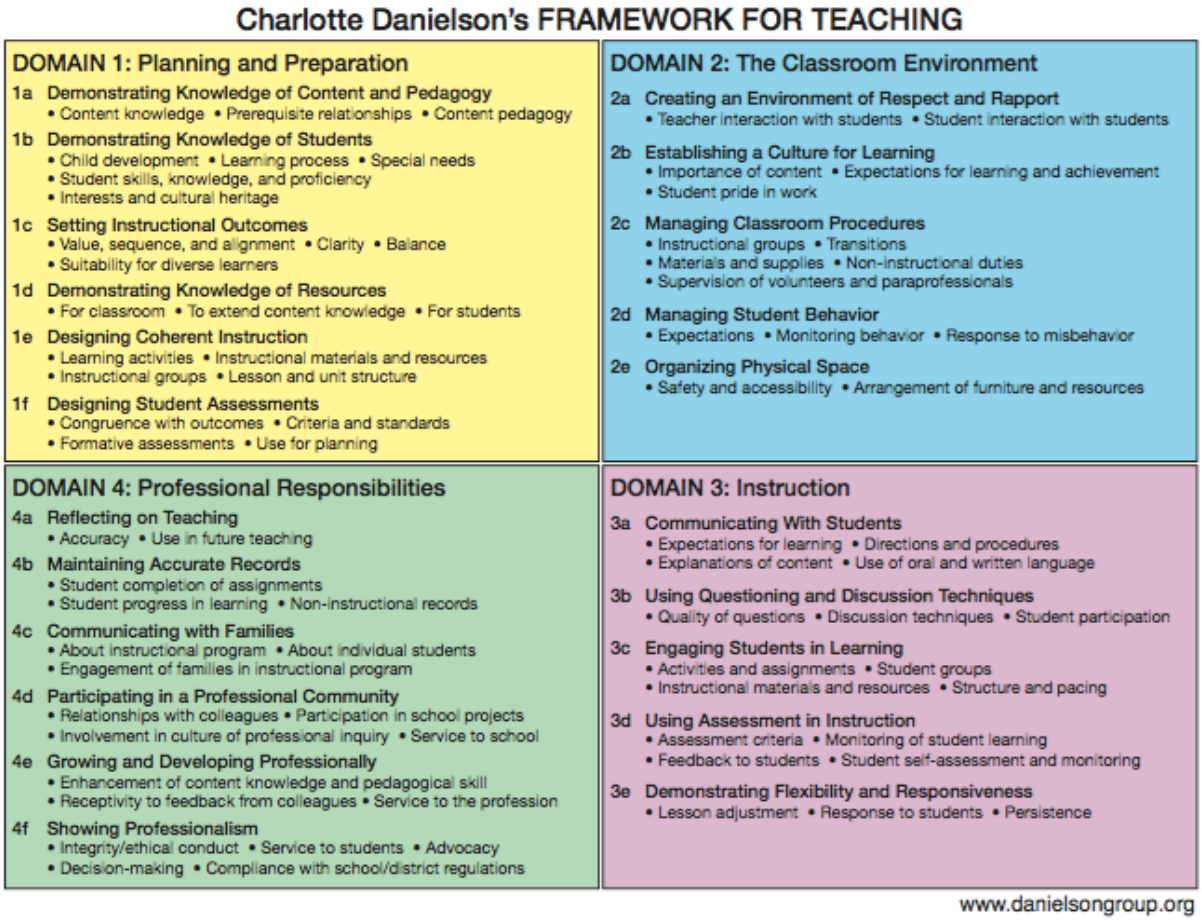
Appendix 7: Administrator Scale Items and Factor Loadings

Administrator Scale Items and Factor Loadings
TESS Capacity

Item	Factor Loadings	Item	Factor Loadings	Item	Factor Loadings
I am prepared to carry out the following aspects of TESS:-Conduct teacher conferences	0.933	1. I am confident in my ability to -Have critical conversations with teachers regarding their performance	0.8	1. I am confident in my ability to -Determine what type of professional development would be most beneficial for my school based on teacher effectiveness data	0.632
1. I am confident in my ability to-Coach teachers on each of the four domains	0.842	I am prepared to carry out the following aspects of TESS:-Accurately rate teachers using the TESS rubric	0.775	To what extent would you like more support and training around the use of teacher evaluation data: -conducting conferences	-0.524
I am prepared to carry out the following aspects of TESS:-Complete all TESS-related paperwork	0.834	1. I am confident in my ability to -Provide substantive feedback	0.766	I can accurately described to others the processes and procedures used to conduct teacher evaluations	0.498
I am prepared to carry out the following aspects of TESS:-Accurately assess the suitability of artifacts for all four domains	0.802	1. I am confident in my ability to -Identify professional development and support for specific teachers based on their evaluations results	0.704	I have resources that I can recommend and/or provide to teachers who need to improve their performance	0.449
Value of TESS					
Item	Factor Loadings	Item	Factor Loadings	Item	Factor Loadings
Administrators should be able to use teacher evaluation results in making decisions about: -intra district transfers	0.861	The TESS system helps me to have better conversations with my teachers about effective instruction	0.755	Administrators should be able to use teacher evaluation results in making decisions about: -teacher pay	0.617
Administrators should be able to use teacher evaluation results in making decisions about: -promotion	0.814	To what extent would you like more support and training around the use of teacher evaluation data: Using teacher evaluation data to determine what type of professional development would be most beneficial for your school	0.753	Administrators should be able to use teacher evaluation results in making decisions about: -student assignments	0.521
The new teacher evaluation system fits well with other school/district initiatives	0.777	The state of Arkansas has clearly and consistently communicated expectations about TESS	0.69	Overall, I think the new teacher evaluation system has had a positive impact on student achievement	0.492
Administrators should be able to use teacher evaluation results in making decisions about: -hiring	0.766				
Opportunity Cost of TESS					
Item	Factor Loadings	Item	Factor Loadings	Item	Factor Loadings
To what extent has time	0.818	To what extent has time spent	0.714	To what extent has time spent	0.653

spent on TESS-related tasks impacted the amount of time you have for the following:-time to reflect		on TESS-related tasks impacted the amount of time you have for the following-casual walkthroughs		on TESS-related tasks impacted the amount of time you have for the following-other paperwork	
To what extent has time spent on TESS-related tasks impacted the amount of time you have for the following-student discipline	0.768	To what extent has time spent on TESS-related tasks impacted the amount of time you have for the following-interacting with students	0.679	I believe that the obligations of TESS interfere with my ability to support other programs and policies	0.502
Desire for More Training and Support					
Item	Factor Loadings	Item	Factor Loadings	Item	Factor Loadings
To what extent would you like more support and training around the use of teacher evaluation data:-coaching teachers in aspect of each of the four domains	0.768	To what extent would you like more support and training around the use of teacher evaluation data:-accurately rate teachers using the TESS rubric	0.632	To what extent would you like more support and training around the use of teacher evaluation data: -Identifying professional development and support for specific individuals based on their evaluation results	0.565
To what extent would you like more support and training around the use of teacher evaluation data:-Accurately assessing the suitability of artifacts for all four domains	0.697	To what extent would you like more support and training around the use of teacher evaluation data:- Having critical conversations with teachers regarding their performance	0.609		

Appendix 8: Danielson's Framework for Teaching

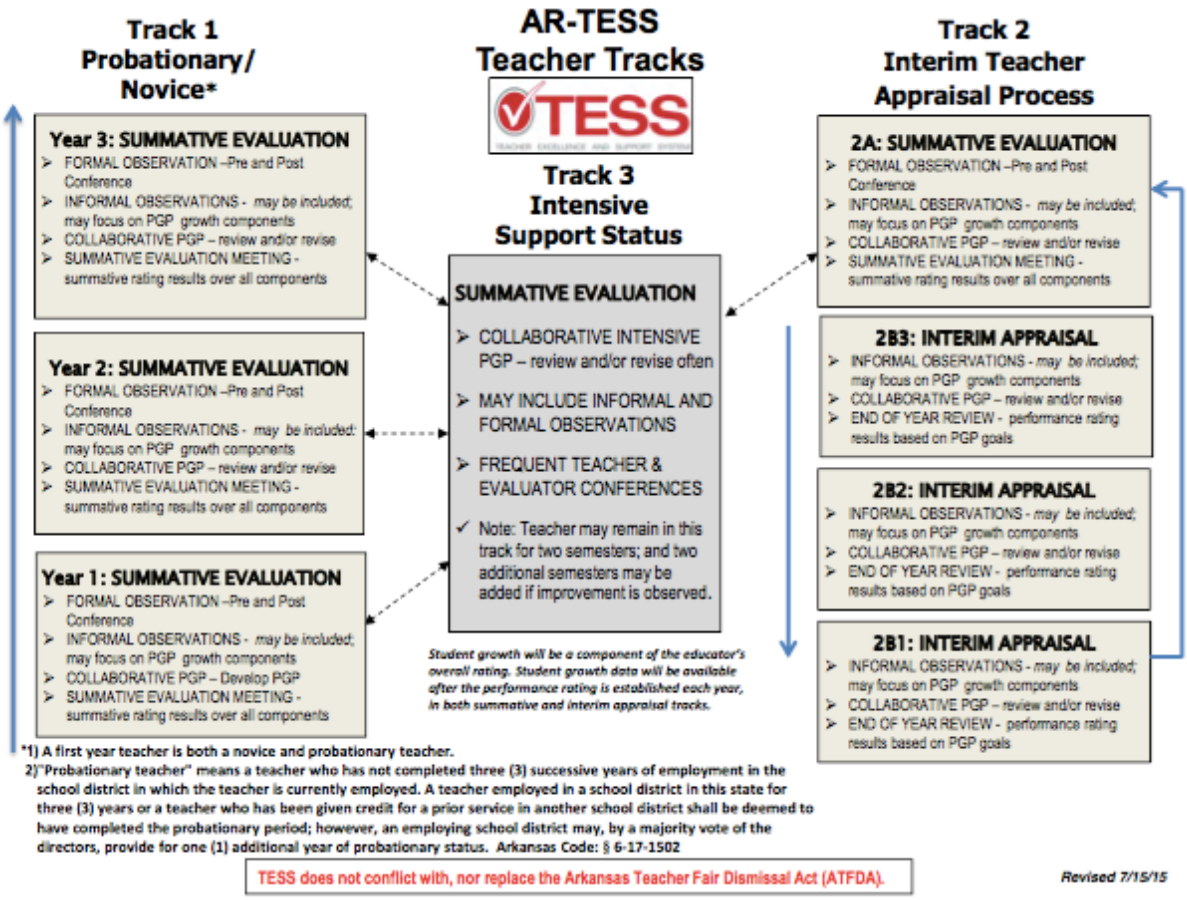


Appendix 9: TESS Performance Levels

Performance Levels

	Unsatisfactory	Basic	Proficient	Distinguished
3b: Using questioning / prompts and discussion	Teacher's questions are of low cognitive challenge, single correct responses, and asked in rapid succession.	Teacher's questions lead students through a single path of inquiry, with answers seemingly determined in advance.	While the teacher may use some low-level questions, he or she poses questions to students designed to promote student thinking and understanding.	Teacher uses a variety or series of questions or prompts to challenge students cognitively, advance high level thinking and discourse, and promote meta-cognition.
	Interaction between teacher and students is predominantly recitation style, with the teacher mediating all questions and answers.	Alternatively the teacher attempts to frame some questions designed to promote student thinking and understanding, but only a few students are involved.	Teacher creates a genuine discussion among students, providing adequate time for students to respond, and stepping aside when appropriate.	Students formulate many questions, initiate topics and make unsolicited contributions.
	A few students dominate the discussion.	Teacher attempts to engage all students in the discussion and to encourage them to respond to one another, with uneven results.	Teacher successfully engages most students in the discussion, employing a range of strategies to ensure that most students are heard.	Students themselves ensure that all voices are heard in the discussion.

Appendix 10: TESS Teacher Tracks



Appendix 11: TESS Intensive Support Track and Arkansas Teacher Fair Dismissal

TESS

ACA §6-17-2801 - 6-17-2809

Intensive Support Track and Arkansas Teacher Fair Dismissal

Required by Statute or Rules	District Options	Teacher Fair Dismissal Act ACA § 6-17-1501 et seq.
• **Teacher is Rated "Unsatisfactory" in any Domain**	Teacher has "Basic" or "Unsatisfactory" rating in a majority of components in any one (1) domain	Evaluation required ACA § 6-17-1504(a)
• **Evaluator Provides Written Notice** -Placement -Time Period -Notify Superintendent	Up to Two (2) Consecutive Semesters	Notice (aligned with ISS Plan) ACA § 6-17-1504(b)(1)
• **Develop Intensive Growth Plan** -Clear Goals and Tasks -Related to Intensive Growth Plan -Evidence Based Research -If Goals are related to Student Growth, use Formative Assessment -Ensure Necessary Support	Use IGP posted on ADE website or the BloomBoard PGP Process Number of Goals and Tasks Types of Assessment Types of Support	**Documentation of :** *"the efforts that have been undertaken to assist the teacher to correct whatever appears to be the cause for potential termination or nonrenewal"* ACA § 6-17-1504(b)(2)
• **Documentation** • **Intensive Growth Plan Evaluation** -Goals & Tasks Completed Written Notice Issued Removal From Intensive Support -Goals & Tasks Not Completed Time-line Extended Notify Superintendent Provide documentation to Superintendent	Two Additional Consecutive Semesters may be added if substantial progress is noted	
• **Superintendent Recommendation** -Recommendation for Non-renewal of Contract Pursuant to the Arkansas Teacher Fair Dismissal Act Code [Ann. 6-17-1501 et seq.]		Notice (aligned with ISS Plan) ACA § 6-17-1506(b)(2)

ADE 04.16.15

Appendix 12: TESS Non-Summative Tracks

TESS

ACA §6-17-2801 - 6-17-2809

Non-Summative Tracks

Tracks 2B1, 2B2 and 2B3

Required by Statute or Rules		District Options
• **Use TESS Role-Specific Framework and Rubric**		
• **Develop PGP**		Number of Goals and Components
• **Consider Appropriate Artifacts and Evidence**		Number of Artifacts
• **Provide a Performance Rating that is included in the Annual Overall Rating*(BloomBoard)** -Rating should be based on evidence collected for components aligned with PGP Goals		
• **Provide Ongoing Support**		Types of support
[*Annual Overall Rating will include Student Growth in the future subject to state and federal requirements.]		

Appendix 13: TESS Summative Tracks (1, 2A, and 3)

TESS

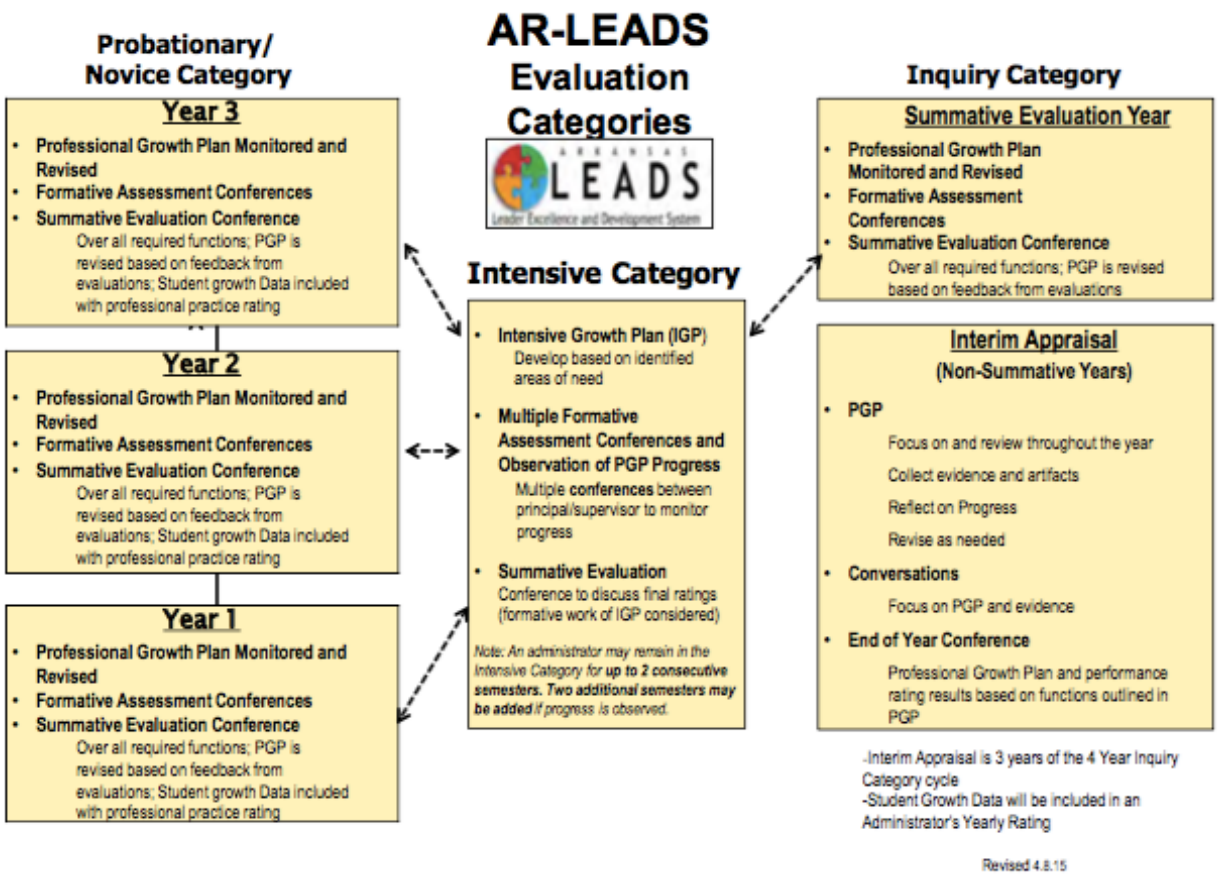
ACA §6-17-2801 - 6-17-2809

Summative Tracks

Tracks 1, 2A, and 3

Required by Statute or Rules		District Options
• **Use TESS Role-Specific Framework and Rubric**		
• **Develop PGP**		Number of Goals and Components
• **One (1) Formal Observation** -Announced -Pre-conference to discuss lesson plans and objectives -Observe 75% of the class period (The length of time for a formal classroom observation of a teacher teaching in a block schedule or in a class period lasting longer than sixty (60) minutes may be adjusted to allow for an observation for forty-five (45) minutes or more of the teacher's class period.) -Post conference for discussion and feedback -Ensure documented evidence of practice (BloomBoard)		Additional Formal Observations Informal Observations may be conducted.
• **Collect Evidence** -Observation documentation -Artifacts		Number of Artifacts
• **Provide a Performance Rating that is included in the Annual Overall Rating* (BloomBoard)** -Rating for Each Domain -Summative Rating for All Domains		
[*Annual Overall Rating will include Student Growth In the future subject to state and federal requirements.]		

Appendix 14: LEADS Tracks


AR-LEADS
Evaluation
Categories
ARKANSAS LEADS
Leader Excellence and Development System
Probationary/
Novice Category
Year 3
Professional Growth Plan Monitored and Revised
Formative Assessment Conferences
Summative Evaluation Conference
Over all required functions; PGP is revised based on feedback from evaluations; Student growth Data included with professional practice rating
Year 2
Professional Growth Plan Monitored and Revised
Formative Assessment Conferences
Summative Evaluation Conference
Over all required functions; PGP is revised based on feedback from evaluations; Student growth Data included with professional practice rating
Year 1
Professional Growth Plan Monitored and Revised
Formative Assessment Conferences
Summative Evaluation Conference
Over all required functions; PGP is revised based on feedback from evaluations; Student growth Data included with professional practice rating
Intensive Category
Intensive Growth Plan (IGP)
Develop based on identified areas of need
Multiple Formative Assessment Conferences and Observation of PGP Progress
Multiple conferences between principal/supervisor to monitor progress
Summative Evaluation
Conference to discuss final ratings (formative work of IGP considered)
Note: An administrator may remain in the Intensive Category for up to 2 consecutive semesters. Two additional semesters may be added if progress is observed.
Inquiry Category
Summative Evaluation Year
Professional Growth Plan Monitored and Revised
Formative Assessment Conferences
Summative Evaluation Conference
Over all required functions; PGP is revised based on feedback from evaluations
Interim Appraisal
(Non-Summative Years)
PGP
Focus on and review throughout the year
Collect evidence and artifacts
Reflect on Progress
Revise as needed
Conversations
Focus on PGP and evidence
End of Year Conference
Professional Growth Plan and performance rating results based on functions outlined in PGP
-Interim Appraisal is 3 years of the 4 Year Inquiry Category cycle
-Student Growth Data will be included in an Administrator's Yearly Rating
Revised 4.8.15