

East Tennessee State University

Digital Commons @ East Tennessee State University

ETSU Faculty Works

Faculty Works

11-1-2015

The Implications of Teacher Performance Assessment and the Impact on Teacher Decision Making

Renee Rice Moran

East Tennessee State University, ricemoran@etsu.edu

Follow this and additional works at: <https://dc.etsu.edu/etsu-works>



Part of the [Language and Literacy Education Commons](#), and the [Science and Mathematics Education Commons](#)

Citation Information

Moran, Renee Rice. 2015. The Implications of Teacher Performance Assessment and the Impact on Teacher Decision Making. *Language Arts*. Vol.93(2). 96-107. <http://www.ncte.org/journals/la/issues/v93-2>

This Article is brought to you for free and open access by the Faculty Works at Digital Commons @ East Tennessee State University. It has been accepted for inclusion in ETSU Faculty Works by an authorized administrator of Digital Commons @ East Tennessee State University. For more information, please contact digilib@etsu.edu.

The Implications of Teacher Performance Assessment and the Impact on Teacher Decision Making

Copyright Statement

Copyright 2015 by the National Council of Teachers of English. This document was published with permission by the publisher. It was originally published by *Language Arts*.

The Implications of Teacher Performance Assessment and the Impact on Teacher Decision Making

“Salary, promotion, tenure, and retention decisions should be tied to an effective evaluation system that includes peer review so that superior teachers can be rewarded, average ones encouraged, and poor ones either improved or terminated” (United States National Commission on Excellence in Education, 1983, p. 35). For three decades, this reform sentiment from the widely reviewed *A Nation at Risk* remained just that—a sentiment—until recently, and with unprecedented speed, this reform has become actualized into policy as many states (particularly those who received Race to the Top [Civic Impulse, 2015] grant funds) scramble to implement teacher assessment

programs. Some states have chosen to hold off on implementing these measures, but a few commissioners have jumped in feet first, and teachers in those states are already being scored using the new evaluation measures. As teachers try to make sense of these evaluation models, their scores, and the impact on students, administrators use these evaluations to make decisions about tenure, retention, and salaries.

In order to observe the implementation and impact of teacher evaluation policy more closely, I followed a small group of first-grade teachers in the Southeast—among the first to be evaluated on the new state teacher assessment system—as they navigated this new terrain (see Table 1 for specifics on

Table 1. *Specifics on the teacher evaluation model implemented*

Evaluation scores	
50% qualitatively based	50% quantitatively based
Classroom Evaluations based on the following broad criteria: <ul style="list-style-type: none"> • Instruction • Planning • Environment • Professionalism Teachers scored on a scale of 1–5 based on the following: <ul style="list-style-type: none"> • Motivation of students • Presentation of instructional content • Lesson structure and pacing • Activities and materials • Questioning • Academic feedback • Grouping of students • Teacher content knowledge • Teacher knowledge of students • Problem solving 	Teacher Evaluation scores based on the following student data: <ul style="list-style-type: none"> • 35% value-added data • For K–2 teachers, this is a schoolwide average of 3 years of scores. Remaining 15% of quantitative score based on one of the following: <ul style="list-style-type: none"> • Professional Development score • Value-added data* • Other forms of student assessment (could include DIBELS scores, running records, or other forms of assessment)
Teacher evaluation scores used for high-stakes decisions such as: <ul style="list-style-type: none"> • Promotion • Tenure • Hiring • Firing 	

*Schools with high value added scores often use their scores a second time in this 15%. In this way value added essentially becomes 50% of their score. Low-achieving schools tended to avoid this.

the evaluation model). My purpose was to examine how these teachers perceived the evaluation process and how their personal reactions influenced instructional decision making in their classrooms. In particular, I asked the following research questions: *How do the teachers studied make sense of the evaluation system based on their experiences? How did the evaluation process and teachers' experiences with the process influence the literacy instruction they provided in their classrooms?*

The assumption on the part of policymakers is that quantitative scores that measure a teacher's performance will increase teacher initiative and that, ultimately, low-ranked teachers will be fired. Additionally, the assumptions are that instruction will be improved and potentially more focused and that student test scores will be improved. It is imperative that we do not just stop at these assumptions, but consider the message of high-stakes evaluation and how teachers' experiences impact that message.

Review of Literature

Because this study focuses on the impact of a high-stakes policy on the lived experiences of teachers, I focused my literature review on previous studies that outlined the impacts of top-down policy. The implications of a top-down policy environment seem clear on the surface: those in authority create and/or implement a particular educational policy; teachers are informed or instructed as to how to execute the policy correctly; teachers change their classroom practice to fit the policy; and a change in student behavior and achievement should occur. In reality, the success of policy implementation is far more layered and complicated. When these complications are not taken into consideration, policymakers and administrators may be puzzled when well-intentioned reform efforts fail (Toll, 2001).

Why Does Top-Down Policy Often Fail at the School Level?

Despite the assumptions made about the implications of policy, research has suggested that even with official frameworks, curricula, and mandates, these measures often fail behind the closed doors

of classrooms for both intentional and unintentional reasons (Cuban, 1995). Various ideas have been posited as explanations for the deficiencies of many policy reforms. These include a) lack of teacher knowledge and professional development; b) distance of policymakers from the classroom; and c) differing discourses of policymakers and educators. Research mentioned here links to these three potential stumbling blocks.

First, in a 1984 study, Cuthbert interviewed classroom teachers about the implementation of federal and state policy initiatives in concordance with career, gifted, and special education. Her findings pointed to teachers' lack of understanding of the policy implementations, which resulted in the absence of proper classroom utilization.

The second concern was highlighted by multiple studies, including Wise (1987) who posited that the distance of policymakers from the actual classroom environment may impact the success of reform efforts. In concurrence, Eisenhart, Cuthbert, Shrum, and Harding (2001) noted that the lived world of policymakers, which revolves around approval from the public, budgetary requirements, and bureaucratic implications, may be far removed from the lived worlds of teachers and students who focus on curriculum and day-to-day classroom social interaction. These researchers argued that this disconnect may make the road from policy creation to policy fulfillment a long and arduous one.

The third explanation is supported by Toll's (2001) research, which found that the lack of success of policy reform had much to do with the differing discourses employed by policymakers and teachers. She contended that the two groups simply were not speaking the same language and that this division impeded policy implementation. In her study, she compared two opposing sources of documentation—interviews conducted with teachers and a policy document entitled the "National Reading Panel Report"—to demonstrate the

It is imperative that we . . . consider the message of high-stakes evaluation and how teachers' experiences impact that message.

contrasting language of teachers and policymakers. In multiple interviews with classroom teachers, Toll found that they based instructional decisions primarily on engagement with students and concern for children's affect. In addition, teachers reported the value of having control over their own choices in terms of classroom decision making. In contrast, her analysis of the National Reading Panel Report emphasized discourse that focused

In essence, reform failure may be directly related to the competing discourse between [teachers and policymakers].

little on the elements that were valued highly by the teachers in the study, suggesting instead a heavier emphasis on the following: the necessity of objective research to inform teacher judgment, the essential nature of student on-task behavior, and the notion of teachers in the passive role of consumers of information. In essence, reform failure may be directly related to the competing discourse between differing groups of stakeholders.

Lack of teacher understanding and professional development as well as differing discourses and goals of policymakers and educators may greatly impact how policy is enacted at the school level. Loeb (2012) notes that we should consider which entities or persons (federal, state, local) are most likely to make the best decisions, particularly when it comes to data-driven reform, and that we must consider local contexts when undertaking new curriculum or policies.

Are Top-Down Reforms Usually Implemented with "Fidelity"?

In his study, Spillane (1999) attended to the relationship between state and local policy in Michigan. Through a combination of qualitative and quantitative methods, he examined how teachers viewed, internalized (or did not), and enacted (or did not) policy reforms. The results of this study indicated that while all teachers claimed to have applied the reforms, observations demonstrated something very different. In fact, classroom change varied from nonexistent to modest to profound, and there appeared to be little relationship between the

use of innovative materials and a desire for change. Rather, Spillane (1999) argued that the success of policy implementation relied heavily on each teacher's zone of enactment, which he defines as the "space where reform initiatives are encountered by the world of practitioners and practice . . . [where] teachers notice, construe, construct, and operationalize the instructional ideas advocated by reformers" (p. 144). Based on this, he argued that the individual resources of the teacher matter, including prior knowledge, disposition, and beliefs, and that enactment zones can serve as mediating forces between policy implementation and classroom practice. Likewise, a study on the impact of curriculum reform and its relationship to change in classroom practice demonstrated variable levels of change, and the authors noted the difficulties teachers had with applying practices aligning with the reform curricula (Moyer, Cai, Wang, & Nie, 2011).

How Do Teachers Make Sense of Policy?

Coburn's (2001) work extended this notion of attention to teacher effect in terms of policy implementation by considering how teachers choose to implement or disregard a particular mandate. She contended that while some researchers suggest that policy influences teachers' work, it is more probable that the reverse occurs, and "teachers interpret, adapt, and even transform reforms as they put them into place" (p. 145). Coburn studied a California school system during the late 1990s in the midst of the state's sweeping reading accountability reforms to better understand these phenomena. The results of the study demonstrated that teachers often made sense of new policy through the process of collective sensemaking—in other words, through both formal and informal conversations and interactions with their peers. In this manner, messages from policymakers were reconstructed and then either attended to or disregarded. Coburn (2001) noted:

From a policymaker's perspective it may seem that schools and districts in reconstructing and reinterpreting policy messages are subverting the intent of policy or thwarting implementation. . . . But another way to look at it is that this sensemaking is both necessary and unavoidable. (p. 153)

This view acknowledges that teachers are confronted with a variety of messages from multiple sources and that they must find ways to translate these messages into workable, active classroom resources.

Coburn (2006) built on the issue of sense making with an examination of problem framing as well. In a continuation of her study of policy implementations in California, she found that the way teachers and administrators framed a particular problem often impacted the way a policy was carried out in terms of classroom practice. For example, through informal and formal dialogue as well as the influence of the principal, one California school reframed a policy implementation to focus on comprehension instruction rather than increased phonics as was originally intended in the state legislation. As a result, the faculty worked to direct classroom instruction as well as professional development toward comprehension rather than toward the original intention of the policy.

Coburn's work may point toward framing (2006) and collective sense making (2001) as potential components of successful reform efforts. As of yet, however, policymakers have not acknowledged the potential of stimulating collaboration in schools, though it would seem collaboration could be a viable means of making change. Nonetheless, based on Coburn's (2001,

2006) results, it is essential that these interactions are not stilted or contrived because that will simply work to undermine the effort of the reform.

Methods

My goal in this study was to describe the perspective of the individuals of a particular culture, in this case teachers, and to understand their daily experiences through the examination of observations, interviews, and artifact collection (Hatch, 2002). I wanted to know: *What is happening here, specifically? What do these happenings mean to the people engaged in them?* (Erickson, 1986, p. 124).

My selection of participants was purposive, relying heavily on the work of Merriam (1998), which presumes that the role of the investigator is to "discover, understand, and gain insight," making it important that he or she "select a sample from which the most can be learned" (p. 61). As a result, I selected individuals who fit the broad parameters of the research but would likely bring different perspectives or points of contrast to bear on my research. I chose to study three different school districts that differed in geographic location, size, and student makeup. Among the three districts, I chose eight first-grade teachers from four different schools whose student populations differed significantly in both socioeconomic levels and racial makeup.

NOW ACT!

Being an Active Part of the Conversation on High-Stakes Teacher Evaluation

You may reside in a state that has already implemented high-stakes teacher evaluation. If not, your state is most likely considering how the process will be addressed, if not applied. It is vital that those in the field of education have a voice as teachers' and students' daily lives are impacted.

- Consider beginning the dialogue at the school level with your coworkers and administrators.
- Work as a team to lay out the benefits and challenges, as well as the needs, of current and future teacher evaluations; this can be essential to the creation of a system that is valid and worthwhile.
- Share your results with district and state leaders; be vocal in your communities and with policymakers about what is occurring in the daily lives of teachers and students.

Remember, one of the keys to improvement in teacher evaluation is bridging the gap between the discourse of policymakers and what is occurring in classrooms.

Teacher perspective was a critical piece of this study, and I consider it to be pivotal in forming an understanding of classroom decision making. Therefore, the primary source of data for this study consisted of interviews with the participating teachers. My rationale in employing participant interviews was to “uncover the meaning structures that participants use to organize their experiences and make sense of their worlds” (Hatch, 2002, p. 91). I believed that classroom observations would serve as a valuable secondary piece of data

Teacher perspective was a critical piece of this study, and I consider it to be pivotal in forming an understanding of classroom decision making.

that might inform the interview process, as well as provide a context for the commentary of the teachers I was studying. In addition to transcribed interviews and field notes, I collected unobtrusive data (Hatch, 2002) that took the form of teacher artifacts, such as teacher observation report templates, lesson plans, and scores. I also collected district and school achievement data. I coded all data in 3 cycles (in-vivo, descriptive, and emotive) and engaged in analytic memoing (Saldaña, 2009) in order to record my own reflections. I then went through several cycles of categorizing and re-categorizing, beginning with eighty to ninety codes and eventually ending with eight themes. For the purpose of this article, I will focus on two themes—level of instructional change and subjectivity.

Findings

Varying Levels of Change in Instructional Practice

As I spoke with the interviewees about their personal changes as instructors in relationship to the evaluation model, two broad groups emerged. One group consisted of individuals whose instruction changed after the implementation of the evaluation model; the other group consisted of teachers whose instruction did not. Within the two broad groups of change versus no change, I found four distinct categories. Under the umbrella of change were

1) those teachers who connected to the model, internalized it, and believed that their instruction was improved as a result, and 2) those teachers who did not necessarily believe in the contents of the model but, due to fear of repercussions, felt forced into change and ultimately determined that their instruction was improved as a result. Under the umbrella of no change were 1) teachers who did not believe in the model and changed their instruction only on a very surface level, and 2) teachers who did not believe in the model and refused to change their instruction. Some teachers could be placed in multiple categories in that pieces of their instruction were subject only to surface-level changes while other aspects were altered on a deeper level. (See Table 2 for more detailed demographics and other relevant information on each participating teacher.)

Karen (0 years experience, Whiteside Elementary) reported the most positive reactions to the implementations. She viewed the evaluation system as a catalyst for improvement in instruction. She believed that she would benefit from classroom observations and the insights of others. Karen explained that “just having different people come in and [get] to see you teach” had had a positive effect thus far, “because sometimes what you think you do really well [may still prompt the observer to make] a suggestion for you of how to improve it.” Karen appreciated having individuals she viewed as knowledgeable observing her instruction and providing her with advice. She believed this improved the quality of her teaching. Additionally, she argued that the model encourages valuable reflection, which can spur good teaching:

Last year after all my post-conferences, I felt like I learned something about myself. They try and give you a positive and something that you can improve on. I felt like that really helped me the best.

Here, Karen notes that the post-conferences, which took place after each evaluation, encouraged her to reflect carefully on her own teaching. Karen’s own openness to new ideas and the positive filter through which her administrator seemed to broach improvements may have come in to play. She noted that the post-conferences were conducted in

Table 2. Teacher participant demographics

District/ School	Teacher/ Year	Economically Disadvantaged	Diversity	Eval. Score	35%	15%	Value- Added Score	Level of Admin. Support	Level of Change
Carson County White-side	Laura/4	74%	Diverse	5	Sch avg	Sch avg.	1.2	High	Forced change
	Karen/0	74%	Diverse	N/A	Sch avg	Sch avg	1.2	High	High level of change
	Allison/4	74%	Diverse	5	Sch avg	Sch avg	1.2	High	Forced change
Hampton Gary	Sally/12	53%	Little Diversity	2	Sch avg	STAR Literacy	-2	Low	No change
Booker Newton	Rebecca/2	48%	Highly Diverse	3	Sch avg	DIBELS	-1.7	Medium	Some change
	Joanne/12	48%	Highly Diverse	3	Sch avg	DIBELS	-1.7	Medium	Some real change, some surface level
Carry Anderson	Kendra/4	20%	Little Diversity	5	Sch avg	Sch avg	3	Medium	Mostly surface level change
	Jessica/4	20%	Little Diversity	5	Sch avg	Sch avg	3	Medium	Some change

a timely manner and integrated both her strengths and her areas of needed improvement. When asked if she found the evaluation model to be negative or stressful, she pointed out that it is simply a tool to move teachers forward:

I know a lot of teachers tend to stress out more about it now, but really it should be something that you're doing in your lessons every day. I don't think it is something that's there to scare you or to stress you out.

Karen's level of buy-in was certainly higher than any of the other participants. A variety of factors may have contributed to this buy-in, including the fact that it was Karen's first year of teaching and she may not yet have formed strong beliefs about her teaching practice. Additionally, all teachers at her school reported a supportive administrator who went above and beyond to provide them with training on how to implement the model. This level of support also seems to include the execution of thoughtful and positive post-conferences.

While many of the other teachers did not implement the evaluation so wholeheartedly at the beginning of the process, a few admitted that they considered it a catalyst for a type of "forced" change. In other words, while they may not have agreed with the premise behind the model, they

implemented it anyway because they felt they had no choice, and, as a result, they saw some improvement in their instruction.

Allison (4 years experience, Whiteside Elementary) is an example of a teacher who felt that she was forced into change, despite her misgivings. As an "apprentice" teacher, someone in the early stages of her career, she felt that she had to follow the requirements laid out by the evaluation model on a daily basis. She feared repercussions as a new teacher if she did not do what was asked of her. The requirements became a part of her daily teaching in large part because of the possibility of unannounced observations:

You had to be teaching that way all the time because you never [knew] when someone might come in to observe. It kind of forced you into making your teaching that way all the time so when someone walked in your classroom, you weren't trying to do something that wasn't natural to you and the kids weren't going, "We never do that." Six-year-olds would definitely say that.

For Allison, the implementation of the evaluation model was initially fear based. She began teaching in the required manner on a daily basis because of the possibility of an unannounced observation. Despite her reluctance, Allison reported the forced

change resulted in some positive outcomes in her instruction. She particularly cited improvements in terms of reflection and intentionality:

I fine-tuned the way . . . the facets in which we're supposed to teach, like the way we're supposed to teach our curriculum, and [it] just made us really focus on the why behind why you do certain things and just being really intentional about how you structure a lesson.

The cases of Allison and Karen demonstrate that deeper changes did occur in some classrooms on a daily basis as a result of the new evaluation model. It is important to note, however, that both Karen and Allison were working in what they deemed a very supportive environment. Their principal implemented numerous trainings in which to delve into the evaluation and offered continuing feedback and support. We must consider how viewing the policy through the lens of the context created at a particular school may impact teacher experience and reaction.

Other teachers studied only made superficial changes in their instruction. Teachers in these classrooms saw a chasm between what they viewed as effective instruction and the requirements of the new model, but they feared the repercussions of noncompliance. Thus, they implemented the requirements on a surface level. Laura (4 years experience, Whiteside Elementary) described this as putting on “a dog and pony show”—in other words, teaching in a manner for the evaluators that you would not on a regular basis. Rebecca (2 years

experience, Newton Elementary) explained that “there are complete ways to just almost create this fictitious image of who you are but not really be who you are.” For some, that might mean having a prepared lesson on hand “just in case” an evaluator walks in to do an unannounced observation, or it might entail creating a lesson for an announced observation that closely follows the rubric but looks nothing like the teacher’s everyday teaching.

While most participants said they did not engage in the process of surface-level change, a few found it to be a wise strategy. Joanne (12 years experience, Newton Elementary) relayed an incident in which a peer planned a lesson that she believed would ensure a good score on the evaluation rubric:

A teacher made the comment that when she knew her announced observation was coming up, she sat down and made sure that her lesson had every one of those things [on the rubric]. She normally doesn't do that. She didn't stick with what she normally does. She took out that rubric and actually made sure that she was doing those things. I never did that, but I thought that was pretty smart.

In her statement, Joanne demonstrates her belief in the need of teachers to find ways to manipulate the system. For her, it seems quite normal and even wise for a teacher to find ways to receive the highest score possible. Here we see evidence of policy that leads not to deeper levels of instructional change, but rather to learning how to “play the game.”

BY THE NUMBERS!

A Few Interesting Facts about the Current State of Teacher Evaluation

- Many states that accepted a portion of the **4 billion** dollars allocated from Race to the Top received these funds based on applications that demonstrated plans to revamp their teacher evaluation programs.
- In order to receive waivers from NCLB (extended to **43** states), states must submit reform plans that include teacher evaluation measures tied directly to student growth.
- Arne Duncan has allowed states to put off the mandatory high-stakes evaluation for one more year, until the **2015–2016** school year. (US News & World Report, August, 2014)

Additionally, it is important to consider context. Joanne works with an administrator who is neither highly supportive nor completely disconnected. She also works at a school with high poverty, high diversity, and low value-added scores. On several occasions, Joanne and her coworkers made reference to the impossibility of reaching the levels of student achievement required. This context may weigh heavily in Joanne's belief in surviving using means of superficial change.

For the final group, who refused to change their instruction regardless of consequences, the evaluation comes into direct conflict with what they view as effective instruction and best practices for children. Additionally, they may feel that surface-level change or "putting on a dog and pony show" is an unethical practice. Jessica (4 years experience, Anderson Elementary) explained, "I try to put the children first and the evaluation second." For her, the two are separate entities. Likewise, Rebecca (2 years experience, Newton Elementary) discussed her issue with surface-level change as she described teachers who follow the requirements of the model only if they know that an evaluator will be coming to watch them:

. . . and [they were] going to get a grade for [the lesson plan] and [they] had to turn it in. And so I don't understand that because that goes against every grain in my logic and thinking, but I've seen them do it.

She also explained her own reaction to this mentality:

They tell you "don't teach to the test," but at the same time, as a teacher, I didn't want to teach to the rubric, either. I didn't want to put on a show and do something that I wouldn't normally do every day. So I did pretty much what I normally did and let it kind of be my baseline from that point on, and I think that for me was a better perspective.

The teachers who took this stance tended to be willing to receive a lower score on their evaluation in order to continue teaching in the manner they viewed as best practice. Sally (12 years experience, Gary Elementary) described an environment at her school in which "most of the teachers keep doing what they're doing and carry on whatever happens."

Even Karen, who was an example of a teacher who internalized the model and had positive reactions to it, cited a scenario related to time management in which she would be willing to sacrifice a higher score for student understanding:

You've put it [the approximate length of time of the lesson] in the lesson, but things don't always go like you planned. . . . I feel like the most important thing is to make sure that the kids are getting it. You don't want to just move on through the lesson even if they're not getting it. I feel like it's more important to make sure that the kids are understanding than to actually get to that closing. So I feel [that you shouldn't] rush through just to get all the pieces in, [even though] you may not get as good of a score because you didn't get that closing in.

This analysis points to top-down policy that may force teachers to make very difficult choices between remaining in good standing at their jobs and making decisions about what they view as good practice in literacy instruction. These decisions are not simplistic, but rather can be confounding, even on moral and ethical levels for some. In this way, we see the teachers applying Coburn's (2001) theory of sense making as they navigate the policy through the lens of their own beliefs and the context of their school setting.

The Implications of Subjectivity in Scoring

Several participants also discussed the notion of subjectivity as problematic in terms of the evaluation model. While in many cases the school principal was the direct focus of commentary on subjectivity, teachers could be evaluated by various individuals, including lead teachers and central office staff. As a result, comments on evaluation extended past the administrator role. In this section, I highlight some teacher perceptions about the subjectivity of the model, instances they cited that they believed to be examples of it, and the relationship they saw between subjectivity and job security.

Laura may view the evaluation model through a slightly different lens than that of her peers because she is the only participant in the study who served in both the role of classroom teacher *and* lead teacher

(and, thus, evaluator). Laura noted that her position as an evaluator has given her an even clearer picture of the subjectivity of the model: “Even now that

Several participants also discussed the notion of subjectivity as problematic in terms of the evaluation model.

another lead teacher, and the curriculum facilitator were all observing one teacher and disagreed on how he should be scored. Laura noted that the other evaluators felt that the teacher spoke too harshly to his students when giving them feedback on their performance, while she argued that she had seen how well his straightforward manner worked with this particular group of students. She explained:

For example, I thought one person’s academic feedback was really good. Some people felt that it wasn’t good academic feedback. I saw it as good because sometimes I do that [speak candidly about their performance] with my kids, and all his kids love the snot out of him. So obviously they trust him enough for him to [be able to be that candid with them]. So, I’m sitting there thinking, well, hmmm, it could go either way on this.

Laura’s personal knowledge of this teacher’s abilities and interactions with his students impacted the way she scored him. While other members of the scoring team found his comments to be somewhat inappropriate, she regarded his manner of presenting academic feedback as both necessary and worthwhile. Laura argued that prior knowledge (or lack thereof) of an individual always affects the score he or she receives:

Let’s say Nancy Ross (central office employee), the head of all this evaluation stuff, comes in and watches me. She would probably give me a lower score than Mr. Whiteside (principal) would because he knows my background with these kids. He knows how hard I work with them. He knows that I’m at this point with these kids. They’re a lower class. It’s subjective. That’s why no one wants anyone from the outside world to come in and have to evaluate us—because of background

I’m a lead teacher, and I’m also observing, I see a lot of the subjectivity as we’re evaluating one person as a group and talking together.” She cited a particular incident in which she, her principal, a central office employee,

knowledge and also how you feel about the teacher. Everyone’s going to have people they like better than others. You’re going to sit in and evaluate someone that you’ve probably already pigeonholed into a certain type of teacher. You can’t go in there without some baggage.

Kendra (4 years experience, Anderson Elementary) also believed that scores on the observation component of the evaluation are impacted by subjectivity and by the beliefs and perceptions of the evaluator. Kendra stated that self-scores and individual evaluator scores may vary from person to person:

I did do a self-scoring . . . I read the rubric and, as clear as it is, there’s still so much room for interpretation. And I know from talking to other teachers [that] how he (principal) interprets things is different than how the woman who observed me last year interprets things; and what they see as more important is different as well. If I hit these three out of five, he’s going to give me a four, bump me up to a four. But it might be the other two bullets that I miss that she harps on.

Several teachers took issue with the notion that job security was tied to what they saw as a subjective scoring method. Both Kendra and Laura said that if the scores are going to have serious implications for job security, then they should not be subjective. Kendra explained, “If you’re going to score me on it, I want to know exactly how you would like me to do it.” Laura concurred, noting that she understands the presence of subjectivity, but takes serious issue with its impact on job security: “My thing is, alright, that’s fine, it can be a little subjective. Some things are. But if it’s my job on the line, I don’t want it to be subjective at all.”

This analysis points to the participants’ generally positive reaction to being observed. Overwhelmingly, they wanted someone to see them teaching and to offer them advice and guidance. They wanted the validation of what they were doing right as well as suggestions for ways to improve. Where the evaluation process seems to fail for these teachers is in its high-stakes nature. The implication of a score that can lead to hiring, firing, and/or tenure is a game changer. It moves the evaluation process from one of reflection and growth to one that must be adhered to in a rigid manner.

Discussion

The results of my study show that teachers take multiple paths when implementing policy mandates. These results add to that body of research by examining teacher perceptions of what it means to work under a high-stakes teacher evaluation system. As more and more states move to this model, it is critical that we carefully consider the implications on the lives of teachers and the impact on curricular content (Au, 2007). McGill-Franzen (2000) noted that the complex nature of teaching often results in unpredictable policy outcomes. Likewise, Spillane (1999) found that the amount of actual change in classroom practice as a result of policy mandates was variable and that instructional changes ranged from nonexistent to extreme. In this study, some participants demonstrated deep-level changes in their teaching, while others showed only surface-level change, and still others demonstrated no change at all in their classroom practice.

The reasons for these varying levels of change could be linked to Coburn's (2001) theory of sense making. She argued that sense making is a natural process in which teachers consider the messages from policymakers and then attend to or disregard them in their instructional practice. Likewise, the teachers in this study employed their own processes of sense making, and in doing so, elements of the evaluation model were either put into practice fully or partially, or were disregarded completely as teachers attempted to translate them into workable classroom adaptations (Coburn, 2001).

From a policymaker perspective, the message is clear: teach well, show student growth, or get out. This study clearly demonstrated, however, that there was not a single message. For some teachers, the message was that their jobs were on the line, and they needed to put on a show when someone walked in the room. Others saw an opportunity for improvement through the rubrics and observations. Still others saw the instrument as flawed and believed it interfered with their ethical decision making about what's best for kids.

Just as in Coburn's (2001) study, we see that policymakers cannot expect a basic input-output

equation when it comes to mandates. In other words, although those in decision-making roles expected the evaluation model to be carried out in a detailed and specific manner, this expectation does not mean that the implementation played out as planned at the classroom level. Spillane (1999) referred to teachers as the ultimate policy brokers; their belief systems are complex and varied, and beliefs always impact how teachers make sense of any mandate (Richards & Lockhart, 1994). Buehl and Fives (2009) posited that a teacher's willingness to embrace policy reform is almost always reliant on an individual's worldview and belief system.

Likewise, in this study, teachers' beliefs about the appropriate nature of classroom practice often impacted which measures of the policy were implemented and which were ignored. We see that the context of teachers' work matters as well. Administrators, central office staff, coworkers, school population, and other school factors may come in to play in how teachers receive policy messages. The triangulation of daily classroom observations, evaluation observations, and stated beliefs demonstrated that despite policy mandates, many teachers will continue making classroom decisions that they view as most appropriate. For example, Sally, Rebecca, and Jessica continued to teach in the manner they believed to be most appropriate, even if it meant a lower score on their evaluation. Other teachers such as Joanne and Kendra made instructional changes on a very surface level, but ultimately continued teaching in a manner closely tied to their beliefs about appropriate practice.

Kagan (1990) noted that in order to begin the difficult process of changing beliefs, individuals must be given "extended opportunities to examine, elaborate, and integrate new information into their existing belief systems" (p. 77). This study also validated this argument. Allison, Karen, and Laura all received ongoing professional development opportunities that allowed them to dialogue about

Policymakers cannot expect a basic input-output equation when it comes to mandates.

the policy implementation. Not surprisingly, this group of teachers was more willing to embrace the policy than were the other teachers.

Broader Implications for Teacher Evaluation

Speedy Implementations/“The Race”

Our current education system is heavily laden with the discourse of speed. In terms of education, we want things to be done well, but we also seem to want to do them quickly. Hence the current phrases: “Race to the Top” and “First to the Top” (see Tennessee First to the Top, 2012). In keeping with this notion of speed, the state in which I conducted

When teachers are given a say, policy has a greater chance of succeeding.

my study was anxious to be one of the first states to put in place a model of teacher evaluation that tied teacher evaluation scores directly to their students’ achievement. While state leaders can now claim the title of one of the first to implement this model of evaluation, the speed of the implementation appears to be the cause of many problems and concerns for teachers. As a result, teachers did not understand the model, and even those in leadership roles appeared to be improperly trained in the language of the model. This finding indicates that policymakers might be wise to slow down and gauge levels of competence before beginning to implement large-scale policy mandates. Field testing and more training could have curbed many of those initial implementation problems.

Teacher Autonomy and Giving Teachers a Say

The last and perhaps most important finding of this study encourages consideration of the issue of teacher autonomy. Various researchers have lamented the essentiality of teacher autonomy in terms of furthering professionalism, providing a stake in student learning, and realizing success of policy implementation (Allington, 2002; Carlone, Haun-Frank, & Kimmel, 2010; Weathers, 2011). This study validates those arguments. Teachers

expressed their feelings of helplessness in terms of policy implementation and, in many cases, found themselves engaging in practices they considered to be in direct conflict with what they believed to be effective instruction. When teachers are given appropriate autonomy, they are happier, more devoted to their profession, and have a greater stake in student outcomes (Allington, 2002; Malmberg & Hagger, 2009).

Several of the participants in this study expressed lack of knowledge about who was in charge, who was creating the policy. This may directly relate to teacher buy-in. Policymakers can consider a process in which teachers have some voice in the creation and implementation of an evaluation system. We have a long history of excluding teachers from the educational conversation (examples such as Bush’s Education Summit and The National Reading Panel spring to mind). When teachers are given a say, policy has a greater chance of succeeding (Hall, 2005), and a bridge might begin to be built between “they” (the policymakers) and “those in the trenches” (the teachers).

References

- Allington, R. (2002). *Big brother and the national reading curriculum: How ideology trumped evidence*. Portsmouth, NH: Heinemann.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36, 258–267.
- Bidwell, A. (2014, August 21). Duncan relaxes testing push, but teachers want more. *U.S. News and World Report*. Retrieved from www.usnews.com/articles/2014/08/21
- Buehl, M. M., & Fives, H. (2009). Exploring teachers’ beliefs about teaching knowledge: Where does it come from? Does it change? *The Journal of Experimental Education*, 77, 367–407.
- Carlone, H. B., Haun-Frank, J., & Kimmel, S. C. (2010). Tempered radicals: Elementary teachers’ narratives of teaching science within and against prevailing meanings of schooling. *Cultural Studies of Science Education*, 5, 941–965.
- Civic Impulse (2015). S. 844—112th Congress: Race to the Top Act of 2011. Retrieved from <https://www.govtrack.us/congress/bills/112/s844>

- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23, 145–170.
- Coburn, C. E. (2006). Framing the problem of reading instruction: Using frame analysis to uncover the microprocesses of policy implementation. *American Educational Research Journal*, 43, 343–379.
- Cuban, L. (1995). The hidden variable: How organizations influence teacher responses to secondary science curriculum reform. *Theory into Practice*, 34, 4–11.
- Cuthbert, A. M. (1984). A case study of policy-induced staff development in a local school district. *Dissertation Abstract international*, 45, 2025A. University Microfilms No. DA8421858.
- Eisenhart, M. A., Cuthbert, A. M., Shrum, J. L., & Harding, J. R. (2001). Teacher beliefs about their work activities: Policy implications. *Theory into Practice*, 27, 137–144.
- Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 119–161). New York, NY: Macmillan.
- Hall, L. (2005). Teachers and content area reading: Attitudes, beliefs, and change. *Teaching and Teacher Education*, 21, 403–414.
- Hatch, J. A. (2002). *Doing qualitative research in education settings*. Albany, NY: State University of New York Press.
- Kagan, D. M. (1990). Ways of evaluating teacher cognition: Inferences concerning the Goldilocks principle. *Review of Educational Research*, 60, 419–469.
- Loeb, S. (2012). In light of the limitations of data-driven decision making. *Education Finance and Policy*, 7, 1–7.
- Malmberg, L., & Hagger, H. (2009). Changes in student teachers' agency beliefs during a teacher education year, and relationships with observed classroom quality, and day-to-day experiences. *British Journal of Educational Psychology*, 79, 677–694.
- McGill-Franzen, A. (2000). Policy and instruction: What is the relationship? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research (Vol. III)* [Kindle version]. Mahwah, NJ: Lawrence Erlbaum Associates.
- Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass.
- Moyer, J. C., Cai, J., Wang, N., & Nie, B. (2011). Impact of curriculum reform: Evidence of change in classroom practice in the United States. *International Journal of Educational Research*, 50, 87–99.
- Richards, J. C., & Lockhart, C. (1994). *Reflective teaching in second language classrooms*. Cambridge, England: Cambridge University Press.
- Saldana, J. (2009). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage.
- Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies*, 31, 143–175.
- Tennessee First to the Top. (2012, May). *Tennessee educator acceleration model (TEAM) evaluation system handbook*. National Institute for Excellence in Teaching. Retrieved from <http://nietbestpractices.org/>
- Toll, C. A. (2001). Can teachers and policymakers learn to talk to one another? *The Reading Teacher*, 55, 318–325.
- United States, National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: A report to the nation and the Secretary of Education, United States Department of Education*. Washington, DC: The Commission.
- Weathers, J. M. (2011). Teacher community in urban elementary schools: The role of leadership and bureaucratic accountability. *Education Policy Analysis Archives*, 19(3), 1–33.
- Wise, A. E. (1987). *Legislated learning: The bureaucratization of the American classroom*. Berkeley: University of California Press.

Renee Moran is an assistant professor of reading education at East Tennessee State University and can be reached at ricemoran@etsu.edu.

Like "Language Arts Journal" on Facebook

