Evaluative Relationships: Teacher Accountability and Professional Culture

Rachel Garver
Montclair State University, garverr@mail.montclair.edu

Follow this and additional works at: https://digitalcommons.montclair.edu/educ-leadership-facpubs

Part of the Educational Leadership Commons

MSU Digital Commons Citation
Garver, Rachel, "Evaluative Relationships: Teacher Accountability and Professional Culture" (2019). Department of Educational Leadership Scholarship and Creative Works. 16.
https://digitalcommons.montclair.edu/educ-leadership-facpubs/16

This Article is brought to you for free and open access by the Department of Educational Leadership at Montclair State University Digital Commons. It has been accepted for inclusion in Department of Educational Leadership Scholarship and Creative Works by an authorized administrator of Montclair State University Digital Commons. For more information, please contact digitalcommons@montclair.edu.
Evaluative relationships: teacher accountability and professional culture

Rachel Garver

To cite this article: Rachel Garver (2020) Evaluative relationships: teacher accountability and professional culture, Journal of Education Policy, 35:5, 623-647, DOI: 10.1080/02680939.2019.1566972

To link to this article: https://doi.org/10.1080/02680939.2019.1566972

Published online: 22 Jan 2019.

Citing articles: 7 View citing articles
Evaluative relationships: teacher accountability and professional culture

Rachel Garver

College of Education and Human Services, Montclair State University, Montclair, NJ, USA

ABSTRACT
Research on recently adopted methods for teacher evaluation are largely focused on issues of validity and pay less attention to the consequences of implementation for the everyday practices of teaching and learning in schools. This paper draws on an ethnographic case-study to argue that the joint tasks demanded by neoliberal teacher evaluation policies structure interactions among teachers and between teachers and administrators in ways that erode professional culture. Implications for policymakers, school leaders, and teachers are considered.

ARTICLE HISTORY
Received 23 March 2018
Accepted 6 January 2019

KEYWORDS
Teacher evaluation; policy implementation; professional culture; accountability; value-added measures; ethnography

Introduction

Over the last ten years, there has been a marked attempt in the United States (US) to expand teacher evaluation systems to more precisely measure teacher quality and to provide information that drives improvement in teacher practice (Hallgren, James-Burdumy, and Perez-Johnson 2014). These efforts are backed by research that shows the critical importance of teachers in student outcomes (Darling-Hammond 2000) and reflect a shift in the approach to pursuing educational equity from desegregation to improving school quality (Superfine 2013). Policy incentives, investments from private foundations concerned with students’ access to effective instruction, and the increased availability of student performance data under federal accountability policies have bolstered these efforts and accelerated reforms to teacher evaluation. Following neoliberal, market-based policy trends in education since the 1990s, teacher evaluation, now also referred to as teacher accountability, focuses on measuring and rating teacher outcomes and impact (Amrein-Beardsley and Holloway 2017; Anderson and Cohen 2015; Ball 2016).

The reforms to teacher evaluation have been contentious. Some see the changes as unreasonable and unfair, while others think they are necessary to mending a rubber-stamp system of instructional oversight. The most controversial development has been the introduction of value-added measures (VAMs), which link the change in student test scores to teacher ratings. Scholars and teachers’ unions have raised concerns about the validity of student growth measures for judging teacher quality, especially when used in high-stakes contexts such as hiring-and-firing decisions (AERA 2015). School districts employ a range
of teacher evaluation tools to assess their instructional staff, and VAMs are typically used in combination with other measures, such as classroom observations and portfolios of lesson plans and student work.

With most attention in the literature to the methodological problems that threaten the reliability and validity of recently proposed and adopted teacher evaluation systems, our understanding of the implementation and implications of teacher accountability policies – how they are enacted by and shape the work of teachers and administrators – has been limited. Evidence that intensified teacher evaluation systems do not fulfill their intended outcomes of motivating instructional growth or dismissing weak educators (Amrein-Beardsly and Holloway 2017) begs the question of what these policies are doing instead. In this paper, I examine how teacher accountability policies impact professional culture through a one-year ethnographic study at a US public middle school. I draw upon organizational theory, policy enactment theory, and the sociology of professions to demonstrate the ways that recent reforms to teacher accountability color the relationships among teachers and between administrators and teachers. I argue that the joint tasks set forth by neoliberal teacher evaluation policies produce policy subjects (Ball et al. 2011) whose interactions are distanced and strained. Joint tasks, as I employ the term here, are sites of coordination required by or in response to policy mandates. They may be explicitly compelled by teacher evaluation policies (e.g. post-observation conferences between administrators and teachers), or they may be locally-forged, creative responses to policy directives (e.g. teachers must grade this standardized test). The erosion of professional culture within the joint tasks entailed by teacher accountability policies is concerning because of its implications for teacher retention, teacher recruitment, student outcomes, and professional growth (Johnson 2015; Kraft and Papay 2014) – undermining the very outcomes that these policies intend to support.

**Producing policy subjects through joint tasks**

Individuals relate to policies both as actors – those with agency in interpreting and implementing policy within their local context – and as policy subjects, individuals who embody and embrace the systems of meaning and modes of action promoted by the policy (Ball et al. 2011; Braun, Maguire, and Ball 2010). Policy subjects tend to behave in ways that reflect and rationalize the dominant institutional logic that motivates their policy environment (Ball et al. 2011; Holloway and Brass 2017; Thornton and Ocasio 2008). The accountability movement in US education has promoted a neoliberal logic that constantly surveils and measures teachers and has minimized a professional logic that respects educators’ expertise and autonomy (Goodrick and Reay 2011). Recognizing that there is debate about the term and how it is employed, I define neoliberalism in education as an institutional or policy logic that champions free-market economic principles. Neoliberalism is not the absence of but rather a particular form of governance that seeks to foster competition through ‘permanent vigilance, activity, and intervention’ (Foucault 2008, 132). With time, policies under the accountability regime have redefined teachers and teaching, and administrators and leading, to give rise to neoliberal subjects (Ball 2003). Institutional logics set ‘expectations for social relations and behavior’ (Goodrick and Reay 2011, 375) and impact how individuals ‘organize time and space, and provide meaning to their social reality’ (Thornton and Ocasio 2008, 101).
As Ball explains, policies work in ways that ‘do not just change what we do they also change who we are, how we think about what we do, how we relate to one another...[they are] in our heads and in our souls’ (2016, 1050). The logics behind policies shape how individuals think about their social world and the possibilities to act within it, such as what constitutes a problem and what solutions are deemed possible (Coburn 2004). Policy subjects are produced by and reproduce policy logics.

In this paper, I focus on the production of policy subjects as the means through which teacher accountability shapes professional culture within schools. One way that policies produce subjects who act in ways complementary to their internal logics is by structuring action and interaction. Although individuals have some agency as policy actors, the range of possible actions available to them is limited. The notion that teachers and administrators in schools have what Coburn calls 'bounded autonomy' (2004, 234) is derived from institutional theories that balance structural determinist and symbolic-interactionist perspectives (Thornton and Ocasio 2008). Embedded agency (Thornton and Ocasio 2008) and inhabited institutionalism (Hallet and Ventresca 2006) recognize that individuals have the freedom to act and make meaning to the extent that the organizational, cultural, and material tools available to them allow.

Policies structure action and interaction through day-to-day practices in schools:

[T]echnologies of reform typically do not confront us in the form of grand strategies but, rather, as mundane and practical changes in our everyday practices. They are embedded in new vocabularies of practice, new roles with new titles, and in grids, templates, mentoring relationships, annual reviews, evaluations and output indicators. It is these very practical and ordinary words and artefacts that present us with new ways of thinking about what we do, about our colleagues, and about ourselves. (Ball 2016, 1050)

I examine how policies organize interaction through mundane joint tasks, or sites of coordination required by or in response to policy mandates. In some cases, these joint tasks are compelled by policy, such as the explicit requirement for administrators and teachers to conference after a formal classroom observation. In other cases, while the ends of the joint tasks are made clear by policy (e.g. administer the test, rate the teacher), the means are not determined, and local actors must creatively forge collective approaches to comply with policy directives. Interactions within joint tasks, whether compelled or locally forged, are marked by institutional logics: ‘institutional logics shape and create the rules of the game, the means-ends relationships by which power and status are gained, maintained, and lost in organizations’ (Thornton and Ocasio 2008, 112). Because individual identities are created in part through interaction (Rodgers and Scott 2008), these joint tasks remake teachers and administrators. By drawing attention to the role of joint tasks in the production of policy subjects, I seek to identify the sites of subjectification under neoliberal evaluation policies where teachers and administrators are disciplined into particular ways of interacting.

The rise of teacher accountability

Reforms to closely monitor and ensure teacher quality have spread across the globe with countries’ efforts to train a workforce competitive in the worldwide marketplace (Tatto 2006) and with the dissemination of the technologies of ‘new public
management’ by international organizations (e.g. OECD) and policy entrepreneurs (Ball 2016; Echávarri and Peraza 2017; Flores and Derrington 2017; Tatto 2006; Verger and Curran 2014). Although there has been a general convergence around teacher evaluation policy internationally, individual countries have taken up the reforms in different ways and to different extents. Local conceptions of what constitutes a good teacher and good teaching, the strength and role of teachers’ unions, policymakers’ orientations to business and management, the extent of private-public partnerships, and/or the presence of established national or regional standardized tests that allow for comparison shape the variation in how countries have adopted international trends in teacher evaluation (Avalos-Bevan 2018; Echávarri and Peraza 2017; Englund and Frostenson 2017; Flores 2018; Klenowski 2012; La Londe 2017; Straubhaar 2017; Tatto 2006; Verger and Curran 2014). The attention to teacher accountability has brought about a range of practices (e.g. professional teaching standards, performance-based pay) and has led to greater centralization in some countries and decentralization in others (Tatto 2006).

Reflecting global patterns, the last ten years have seen an expansion and intensification of teacher evaluation policies in the US. Rubrics have been introduced or revised to precisely define effective instruction and to rate teachers on multiple components of their practice. Value-added measures (VAMs) link students’ growth on performance measures, such as standardized tests, to teacher effectiveness ratings. These reforms, which attempt to more precisely measure teacher quality, were spearheaded by the Obama administration, whose Race to the Top (RttT) initiative was the first major federal effort in the US to standardize teacher evaluation policies that have historically varied widely across states and districts. RttT offered financial incentives to incorporate student growth measures into teacher ratings. In addition, No Child Left Behind (NCLB) waivers granted under Obama also required states to strengthen their teacher accountability policies. These centralized initiatives were inspired and bolstered by several trends in education: the rising influence of econometric research in education that spoke to teachers’ impact on students’ academic achievement and future earnings (e.g. Chetty, Friedman, and Rockoff 2011; Hanushek 2011); an accountability policy paradigm in which effectiveness in all realms of education is measured by student outcomes on standardized tests (Murphy, Hallinger, and Heck 2013); the increased availability of student performance data; and advocacy groups and foundations concerned with the quality of US teachers (Amrein-Beardsley and Holloway 2017).

Most of the research on recent reforms to teacher evaluation has concentrated on the reliability and validity of detailed observation rubrics and VAMs. As the rise of VAMs is due to the increased influence of economists in education (Amrein-Beardsley and Holloway 2017), scholarship has responded in a similar language, focused on the technical, methodological core of the policies. Although VAMs are recognized as an improvement over status models that evaluate teachers according to their students’ performance in relation to an established threshold (AERA 2015), the list of concerns about the validity of VAMs is extensive. The array of critiques launched against VAMs include concerns about missing data; isolating the effects of a single teacher; excluding peer effects, school composition, and student attitudes toward the test; separating school and teacher effects; small sample sizes; bias against teachers with extremely low- and extremely high-performing students who show little growth on grade-level
tests; and using only one test to assess student knowledge (AERA 2015; Berliner 2013; Betebenner et al. 2012). Moreover, several scholars have found teacher ratings to be unstable from one VAM formula to another, one year to the next, or when switching grades or subjects, suggesting that VAMs are not a reliable gauge of instructional effectiveness (Berliner 2013; Betebenner et al. 2012). For all of these reasons, many have warned that VAMs lead to misinterpretation and misclassification and are nearly impossible to implement fairly with reliability and validity (AERA 2015; Amrein-Beardsley and Holloway 2017; Berliner 2013; Betebenner et al. 2012). Despite the laundry-list of concerns that educational researchers bring to the table, teacher ratings, like many neoliberal reforms that rely on highly technical measures of efficiency and effectiveness, emit an illusion of certainty, objectivity, and comprehensiveness that is perceived as legitimate (Schwarz 2015).

Recently, some scholars have called for research that looks beyond the technical issues of newly adopted strategies to measure and rate teachers toward the implementation of these policies and the unintended consequences of putting them to practice (Braun 2015; Hewitt 2015; Holloway-Libell and Collins 2014; Johnson 2015; Murphy, Hallinger, and Heck 2013; Riordan et al. 2015). As these reforms fall short of their intended outcomes (e.g. raising student achievement), the unintended outcomes are particularly important to consider (Johnson 2015; Murphy, Hallinger, and Heck 2013). According to Murphy, Hallinger, and Heck (2013), since the direct relationship between teacher evaluation and student achievement is questionable, we need to examine possible indirect relationships, including the role of professional culture in mediating policy’s impact on student learning.

Professional culture is the slice of school culture pertaining to teachers and administrators. Building on Cohen et al.’s definition of school culture, I define professional culture as ‘patterns of [administrators’ and teachers’] experiences of school life [that reflect] norms, goals, values, interpersonal relationships, teaching and learning practices, and organizational structures’ (2009, 182). A school’s professional culture is not simply a microcosm of the larger societal culture or a mirror image of the professional culture in other organizations and fields (e.g. business, finance) (Allaire and Firsatro 1984). Nonetheless, it is significantly shaped by the school’s environment, including the policy context (Allaire and Firsatro 1984; Cohen et al. 2009). Professional culture has been operationalized in previous research as a composite of several factors, including faith in leadership, peer collaboration, support for professional development, feelings of safety, trust in colleagues, and teacher voice (Kraft and Papay 2014; Riordan et al. 2015). Teacher-to-teacher and administrator-to-teacher relationships, which I examine in this paper, are the site for several of these components (e.g. trust, peer collaboration, leadership) (Cohen et al. 2009).

A positive professional culture is foundational for effective schooling. It is correlated to teacher retention, professional growth, and student achievement, all central goals of recent reforms to teacher evaluation (Cohen et al. 2009; Stoll et al. 2006). Kraft and Papay found that ‘teachers in schools with stronger professional environments experience greater returns to experience over time’ (2014, 489). As opposed to teacher evaluation systems that understand instructional quality as fixed and internal to individual educators, research shows that teacher development is a collaborative process. In schools with strong peer networks and administrative support, teachers improve, while teacher growth is stunted in schools with a poor professional environment (Johnson
A supportive professional culture can maximize and magnify a teacher’s abilities: ‘the school organization becomes greater than the sum of its parts, and in this way, the social capital that transforms human capital through collegial activities in schools increases the school’s overall instructional capacity and, arguably, its success’ (Johnson 2015, 119). Serving as the foundation for social capital, staff relationships have a significant impact on professional improvement and student achievement.

**Teacher accountability in practice: the impact on professional culture**

The sociology of professions, which examines how professions, professionals, professionalism, and professionalization are shaped by their historical, social, political, and economic contexts, suggests how neoliberal education policy may impact professional culture in schools. Looking across several fields of work, Evetts proposes that market-driven reforms have brought about what she refers to as a new-professionalism characterized by ‘increasing levels of managerialism, bureaucracy, standardization, assessment and performance reviews’ (2011, 407). New-professionalism allows practitioners little control over their work as the ‘managerial demands for quality control and audit, target setting and performance review become reinterpreted as the promotion of professionalism’ (Evetts 2011, 412). Practitioners have seen a loss of discretion and an increase in oversight of their performance based on individual output aligned to organizational goals (e.g. number of clients visited, number of traffic tickets distributed). They are no longer trusted to exercise professional judgment independently and are subjected to constant surveillance, shifting the relationship between supervisor and employee: ‘the traditional relationship of employer–professional trust is changed to one necessitating supervision, assessment and audit’ (Evetts 2011, 416). Contracts replace professional commitment and competence as the mechanism to ensure work quality: ‘sociality and social relations are replaced by informational structures. We all know and value others by their outputs rather than by their individuality and humanity’ (Ball 2016, 1054). Supervisors take on the role of interpreting policy and insist on consensus around their interpretation, leaving little autonomy to their employees (Ball et al. 2011). These changes in the ‘ways and means of organizing work and workers’ (Evetts 2011, 409) have facilitated individualization in the workplace, encouraged competition, and discouraged collaboration and collegiality.

Applying Evetts’ concept of new-professionalism to the field of education, Anderson and Cohen (2015) explain how the expansion of market-based reforms – such as school choice, increased reliance on test scores, data-driven management, and expanded teacher evaluation systems – have produced new teacher identities. Neoliberal policies influence the behavior of teachers from the outside by, for instance, increasing the number of tests that teachers must prepare their students for, but also from the inside, normalizing market-based logics in the ways that teachers and administrators reason about and address daily decisions. Reduced to their student achievement scores and under the constant surveillance of measurement, teachers expend significant energy on ‘impression management’ (Anderson and Cohen 2015, 13). Under pressure to attain particular performance targets, teachers focus on compliance rather than on professional learning and growth (Ball et al. 2011). Teachers experience anxiety and may lose motivation in their careers, as developing meaningful connections with students, parents, and colleagues with similar commitments is trivialized in comparison to...
more easily measurable aspects of their job. Buchanan explains the misalignment of teacher accountability to teacher motivation: ‘The current reform context overemphasizes the technical and rational components of the profession, and devalues the emotional, personal, and relational aspects of teaching, which is at the core of many teachers’ commitments’ (2015, 703). By changing what it means to be a good teacher, neoliberal education policies attempt to mold teachers into subjects who interact in ways that support and reproduce their market-based logics.

The literature proposes a few ways in which teacher evaluation has impacted professional culture or, more specifically, professional relationships within schools. All the studies reviewed in this section foresee that recent reforms to teacher evaluation will negatively impact professional culture, specifically by decreasing collaboration and increasing anxiety and litigiousness.

**Increased individualism and decreased collaboration**

Several scholars suggest that reforms that aim to more precisely measure teacher quality will lead to increased competition and individualism (Ball 2003; Collins 2014; Flores and Derrington 2017; Hewitt 2015; Johnson 2015; Sachs 2001). Some trace this change back to the comparative design of teacher rubrics and student growth measures, which are intended to differentiate teachers rather than reflect their effectiveness (Amrein-Beardsley and Holloway 2017; Hewitt 2015). VAMs may incentivize competition between teachers, especially when financial bonuses are offered to teachers whose students show the most growth (Collins 2014). Teachers are discouraged from sharing expertise and successful strategies and from providing each other with constructive feedback. Teachers are encouraged to take sole ownership over their own students rather than to feel accountable to the entire student body at their school (Johnson 2015). By enhancing the traditional egg-crate culture of schools in which teachers operate independently in their classrooms, neoliberal teacher evaluation policies may undermine student growth and professional development, both of which rely on teacher collaboration (Johnson 2015).

**Increased anxiety and decreased morale**

Teachers may feel increased anxiety about their evaluations as the formulas for teacher ratings are made more complex. They find themselves stuck between the pressure to achieve high ratings and the difficulty to understand how their scores are calculated, feeling fearful and a loss of control over their career (Balch and Koedel 2014; Collins 2014; Conley and Glasman 2008; Flores and Derrington 2017; Leonardatos and Zahedi 2014). Teachers have several concerns about how VAMs are determined, such as how their ratings are impacted by student attendance, school closings, and working with students with disabilities or low-income students (Balch and Koedel 2014). They are increasingly insecure and lose confidence in their practice as they feel constantly surveilled and measured (Ball 2003; Leonardatos and Zahedi 2014). Anxiety can foster teacher practices that undermine student learning, such as gaming the test, cheating, and focusing on the students who are most likely to improve (Conley and Glasman 2008). On the other hand, proactive efforts by states to deter these practices, such as New York’s extensive spending on test security and the requirement that teachers do not score their own students, communicates policymakers’ distrust in teachers and further lowers morale (Leonardatos and Zahedi 2014). Reinforced by public opinion and media coverage that denigrate educators, reforms to teacher
evaluation may lead teachers to leave the profession or retire early at higher rates rather than put up with the stress of having their practice constantly measured and assessed (Hewitt 2015; Leonardatos and Zahedi 2014).

**Increased litigiousness**

Some legal and education scholars have forecasted an increase in lawsuits filed by teachers and teacher unions or by families and community advocates. The lack of transparency and the questionable methodologies of student growth measures open school districts up to being sued by teachers and teacher unions (Baker, Oluwole, and Green 2013; Holloway-Libell and Collins 2014). Teachers may challenge their dismissal on several legal bases, considering the problems with VAMs’ reliability and validity:

Overly prescriptive, rigid teacher evaluation mandates, in our view, are likely to open the floodgates to new litigation over teacher due process rights...despite the fact that much of the policy impetus behind these new evaluation systems is the reduction of legal hassles involved in terminating ineffective teachers. (Baker, Oluwole, and Green 2013, 18–19)

Sawchuk (2015) documents many such suits brought by teacher unions to federal and state courts, based on a range of complaints, including lack of clarity in or fidelity to policy and lack of validity and reliability in the formulas utilized. In addition to legal actions initiated by employees, teachers, schools, and school districts may be charged with educational malpractice now that incompetent and inadequate instruction is allegedly possible to precisely and objectively identify: ‘A new public policy favoring accountability and the advent of a statistical model that purports to identify ineffective – possibly incompetent – instruction may combine to build a path to recognizing educational malpractice as a viable remedy for students who allegedly suffered an academic injury’ (DeMitchell, DeMitchell, and Gagnon 2012, 259). DeMitchell, DeMitchell, and Gagnon (2012) worry that increased litigiousness could encourage teachers to adopt practices that ensure higher ratings but do not support student learning, such as gaming or teaching to the test.

**Limitations of the literature**

Our understanding of the impact of teacher accountability on professional culture has been limited by the range of methodologies used. Some of the studies build on theory, law, and the logic of neoliberal policies to forecast how teacher accountability may impact professional culture (e.g. Anderson and Cohen 2015; Baker, Oluwole, and Green 2013; Braun 2015; DeMitchell, DeMitchell, and Gagnon 2012). Other studies are empirically based but rely on interviews and surveys that elicit teacher and administrator perceptions of new evaluation policies and how they affect professional interactions (e.g. Collins 2014; Hewitt 2015; Reddy et al. 2018). While these studies offer important understandings of the perceived validity of teacher accountability, they are vulnerable to social desirability and cannot be understood as direct reflections of behavior within schools. None of these studies include observations within schools that substantiate or disconfirm theoretically-based projections or practitioners’ self-report. Moreover, few studies go beyond issues of policy design to trace the implementation of teacher accountability. I examine the joint tasks entailed by teacher accountability in action. Relying largely on ethnographic observations, I offer data to probe
theoretical projections and self-report-based claims. The focus on ‘joint tasks’ bridges the bodies of literature reviewed above by illuminating one mechanism through which neoliberal policy logics construct teacher subjectivities and, thereby, shape professional culture. I document how policy subjects are created intersubjectively, through the interactions entailed by teacher evaluation. While studies based on surveys and interviews have been largely focused on teacher-teacher relationships alone, I also consider administrator-teacher relationships.

**Methodology**

**Data collection**

This paper is derived from a one-year ethnography of Germaine Middle School (GMS), a racially and economically segregated public middle school located in a large urban district in the US. GMS was selected for its diverse, high-needs student body, whose academic achievement is more reliant upon teacher quality than student populations of higher socio-economic status (Berliner 2006; Kainz and Pan 2014; Owens, Reardon, and Jencks 2016). Asian, Black, and Latinx students each constituted more than 20% but less than 50% of the student body. Approximately 15% of students were classified as English learners (ELs) and 15% as students with disabilities (SWDs), and the entire student body qualified for free lunch.

An ethnographic approach allowed me to capture how the day-to-day enactment of accountability policy was embedded in the local organizational and social context (Ball et al. 2011; Coburn 2005, 2006; Spillane 2006). I conducted over 600 hours of participant observation and 73 semi-structured interviews with teachers, school staff members, school and district administrators, parents, and students during the 2014–2015 academic year. I interviewed every school administrator and fifty-three teachers, which represented more than 85% of the instructional staff at GMS.

This paper draws heavily from my observations of more than 100 staff meetings, including departmental meetings, whole staff meetings, committee meetings, and administrative meetings, and more than 80 lessons across all content areas and grade levels. I collected documents used by participants at staff meetings, such as tables presenting student performance data and district guidelines for grading exams. As tools that in part structured interactions, these documents assisted in creating a holistic record of joint tasks. On days when I trailed school leaders through their regular routines, I witnessed formal tasks required by teacher evaluation policy, such as administrators’ observations of teachers in classrooms and pre- and post-observation meetings between administrators and teachers.

**Data analysis**

With the use of the qualitative analysis program MAXQDA, I initially analyzed observation fieldnotes and interview transcripts by open coding, producing an extensive set of emergent themes (Glaser and Strauss 1967). During this first round of analysis, I also assigned descriptive codes that indicated, for instance, the grade level, content area, month, category of actor (e.g. student, teacher, parent, administrator), to the excerpted sections of data. These descriptive codes allowed me to analyze how findings
varied across organizational structures, actors, social groups, locations, and time. After
the first round of analysis was complete, I consolidated, hierarchized, and precisely
defined the inductively-derived themes to create a codebook that was then used to
recode the entire data set (Miles, Huberman, and Saldaña 2014).

The data captured by the codes ‘teacher evaluation,’ ‘mandates to teachers from the
school or district,’ and ‘value-added measures’ were central to my analysis, and, when
I found professional relationships to be significant across these themes, I broadened my
analysis to include codes that indicated relational aspects of teacher evaluation, such as
‘school culture,’ ‘perceptions of teachers and teaching,’ ‘perceptions of administrators
and educational leadership,’ and ‘teacher resistance.’

Findings
At the time of my study, GMS’s teachers were evaluated by 1) classroom observations
conducted by school administrators and guided by a detailed rubric and by 2) a value-
added measure that averaged student growth on two standardized tests – one designed
by the state and one provided by the district. The state test had been previously
administered for measuring student achievement and making student promotion deci-
sions, and it had only recently been integrated into educators’ evaluations. The inclu-
sion of the district test was an optional addition to the state test, and the teachers at
GMS had consistently voted to include it in their VAM. The state test was administered
once a year in the Spring, while the district test consisted of a pre-test in the Fall and
a post-test in the Spring. The results of district test did not have consequences for the
students, only for the teachers’ VAMs.

Although teacher accountability promotes an individualized logic in which
teachers are held accountable for their own performance and those of their
students alone, several of the steps needed to fulfill the policy demands and to
produce a teacher’s rating were collective. The teacher evaluation system required
the staff at GMS to carry-out multiple joint tasks compelled by or in response to
policy mandates. Below, I highlight several of these joint tasks and provide
evidence to show how they structured relationships in ways that eroded the
professional culture at GMS. Professional interactions both directly and indirectly
related to assessing instructional quality were strained by neoliberal teacher eva-
luation policy. Feelings of anxiety and vulnerability marked teachers’ interactions
with each other, while administrator-teacher interactions were distanced, juridical,
and technical. I distinguish between the character of administrator-teacher and
teacher-teacher relationships because teacher evaluation policies work differently
upon school actors depending on their location in the institution; they delineate
the roles of administrators and teachers and infuse a power differential between
the two. Moreover, teacher evaluation policy demands different joint tasks be
carried out by teachers and administrations (e.g. formal observations, feedback
conferences) and among teachers (e.g. establishing norms to grade the tests,
grading each other’s students), establishing distinct contexts that structure
interactions.
Teacher-teacher relationships

Teachers at GMS had to work together to accomplish several steps required by the teacher evaluation system. Many practices, such as proctoring and grading the tests of other teachers’ students, made teachers more reliant upon each other. In some cases, interdependence was explicit, such as when two different teachers relied on the same test scores from the same students for their student growth measures. The joint tasks of evaluation were sites that fostered anxiety, distrust, horizontal surveillance, and vulnerability in teachers’ relationships. As teachers were enlisted in the process of measuring their peers, they became sources of ‘uncertainty and instability’ (Ball 2003, 220) for each other. Below I present three joint tasks required by student growth measures, from classroom instruction to the administration and grading of exams, that strained teacher relationships.

Sharing students, sharing student test scores

In cases where there was no standardized test offered in a teacher’s content area or when core subject and intervention teachers worked with the same students, the same student test scores factored into the student growth measure of multiple teachers’ ratings. For teachers, this meant that the measure of their professional quality was subject to the instructional practices of their colleagues. Such interdependence required the trust that one’s colleagues were committed and effective educators, or at least preparing their students well for the tests. Teachers expressed feelings of vulnerability, anxiety, and injustice about their reliance on colleagues, which was reified and made high-stakes by evaluation policy.

Teachers outside of the four primary content areas (math, English language arts, science, and social studies) were particularly vulnerable to the performance of their colleagues’ students. Elective teachers (e.g. gym, music, art) were completely dependent on their colleagues in other content areas when it came to the student growth aspect of their rating. For example, the gym teachers were evaluated using the English language arts (ELA) and math scores of the students in their classes. This disconnect between what a teacher taught and the assessment used in their evaluation was perceived as unfair across the entire staff. One assistant principal critiqued the state’s suggestion to increase the percentage of a rating reliant on student growth measures because of this misalignment. In an interview, he rhetorically asked, ‘you think that’s fair...[that] the gym teacher’s rating is dependent on how the ELA teachers and the math teachers teach?’ The district had sought to minimize this misalignment by introducing assessments in more subject areas than math and ELA in recent years. These efforts only addressed social studies and science teachers, and they increased the number of standardized tests for students and the time spent on grading summative assessments that do not inform instruction for teachers. Improving the validity of teacher ratings came at the expense of the expansion of testing, which past research has repeatedly shown is detrimental to student learning by, for example, narrowing the curriculum (Jennings and Bearak 2014). Here is one instance in which it becomes clear that efforts to attend to the validity of current methods, specifically VAMs, are insufficient because they neglect the consequences for the everyday work of teaching and learning in schools.
Although less dependent on their colleagues' practice than elective teachers, core subject and intervention teachers who shared the same students were assessed according to the same student test scores, even though they did not have full control of the students' learning in their content area. GMS pulled many students out of their classes for a range of intervention programs or services. Some of these sessions were initiated by teacher referrals or justified by student test scores the previous year, and others were mandated by SWDs' Individualized Education Programs (IEPs). GMS sought to schedule supplementary supports, such as speech therapy and ELA and math intervention, during elective periods so that students did not miss their core classes. Teachers who provided math or ELA intervention services for students one or two periods a week and the ELA and math core teachers relied on each other to support students concurrently. One intervention teacher explained: 'I'm depending a lot on the other teachers to also hopefully get those kids moved up...if the core teachers aren't making that move for me and I'm not helping them make that move then I'm kind of screwed a little.' The two teachers' fates were tied together, which had the potential to promote collaboration between the core and intervention teachers, although in the absence of allotted time and resources for discussing students and sharing strategies, I observed no formal collaboration and limited informal communication between core and intervention teachers over the year.

Core subject teachers in inclusive classrooms and special education teachers who pulled students out for some or all of the minutes allocated to math or ELA were similarly dependent upon the same student tests scores. GMS sought to integrate SWDs as much as possible in inclusive or general education classrooms (in the 'least restrictive environment'). SWDs whose IEPs recommended smaller instructional settings were placed in inclusive, team-taught homerooms and accommodated with pull-out sessions for part or all of their math and ELA instruction. This design meant that a teacher may only instruct a student in ELA a third of the time and be reliant on another teacher to provide quality instruction for the remaining periods dedicated to the same subject. In reviewing her students' data at a faculty meeting, one math teacher worried aloud that her performance was being evaluated in part by the performance of students who were pulled out for part or all of their math instruction. A district special education administrator who worked with GMS explained how student growth measures heighten the need for trust among teachers:

If I'm going to be rated on the success or outcomes of a test for my children, I solely would like to know that I know every in and out about that child. If I take that professional onus, I have to trust my colleague has that same professional culture that I have. I trust you that if I send my kid there for 30 minutes, they're going to work as hard as I am working with your kids for 30 minutes. That's the breakthrough. Teachers don't have that trust.

Distrust and feelings of unfairness emerged from the way that the evaluation system made teachers interdependent.

**Proctoring tests for other teachers' students**

According to test guidance, teachers were not allowed to proctor the test for their own students. They were reliant on their colleagues to administer the test. In the preparation
for administering the state standardized tests, I observed the staff express anxiety about their interdependence and pressure each other to conform to testing regulations. A math teacher who also served as GMS’s test coordinator spent weeks preparing for the multiple exams that needed to be administered in the Spring. He worked with attention to detail to ensure that he managed the tests according to regulation. Reflecting on the gravity of his responsibilities, he explained: ‘I don’t want to mess up because it’s [the teachers’] ratings.’ At one staff meeting, he insisted that teachers proctor the tests with an appreciation for the impact on their colleague’s ratings: ‘there is no texting, no cell phones...no iPads...you can’t just sit down...this is another teacher’s rating.’ The next day, the test coordinator explained to me why he had been so forceful with his colleagues:

It affects my rating, it is not fair if another teacher does not actively proctor while my kids take the test. You can say to the whole class, “pay attention,” “make sure you stay on task,” “make sure you answer all of the questions,” “don’t leave any blank.” You can tell a sleeping student to wake up but you cannot say anything specific about the question they are on. You have to walk around for an hour and a half.

In order to meet this teacher’s expectations for being an accountable colleague, teachers had to actively monitor the class, encouraging students to complete the test to the best of their ability without providing direct assistance, which would qualify as cheating. During the district assessments a few months later, which also factored into teachers’ student growth measures, the test coordinator was similarly emphatic that his colleagues proctor with vigilance, even though the official guidelines were more lenient than with the state exam, such as allowing teachers to spread the test over two days if necessary. He insisted in a staff meeting that the teachers understood how the extent to which they encouraged students to remain attentive could have serious consequences for their colleagues. Despite the actual variability in teacher ratings because of proctoring practices, the test coordinator encouraged teachers to feel vulnerable to each other, to worry that others might not act in their benefit. Proctoring for colleagues strained professional relationships not only by introducing anxiety and suspicion, but also by encouraging teachers to forcefully pressure each other to behave in ways that maximized student scores.

Although I was unable to observe teachers proctoring during standardized exams, I witnessed GMS pilot student surveys for the district, which was exploring the possibility of incorporating student opinions into teacher ratings. Like the tests, these surveys had to be administered by a colleague of the teacher being evaluated. In one classroom, the students struggled to comprehend, or even in some cases to read, the questions. In response, the teacher paraphrased each question in her own words for the class, allowing them to complete the survey while also shaping student understandings and, thereby, responses. Despite her efforts to keep the students engaged, the teacher struggled to manage student behavior. Many of the students were wandering around the classroom. Some explicitly refused to complete the survey, and others expressed that it was boring. This observation suggests that concerns about variability in how teachers administered assessments for teacher evaluation may have been warranted.

Grading other teachers’ students

Teachers spent a majority of their weekly allotted collaborative time preparing for, norming, and grading tests in the Spring. As mentioned above, the guidelines for the
administration of the district tests were more relaxed than the strict regulations for the state exams, which were graded externally. Teachers could grade their students’ district pre-tests in the Fall, however the post-test had to be graded by a colleague. Teachers relied on their colleagues to grade fairly and according to how each department had normed with the rubric provided by the district.

The teacher leaders were put in charge of norming the test with their departments and overseeing grading sessions. One teacher drew attention to how both quality of instruction and grading practices affected scores as she prepped the ELA department for marking the post-test: ‘there is no teacher who got less than effective last year...no teacher that this was a negative thing for them...part of it is that we do a good job [teaching] during the year...[and it] could be we score ourselves low in the fall and our buddies score us better [in the Spring].’ There were grave concerns in the ELA department about including some teachers in the grading. The year before my study, two teachers in the ELA department were excluded from grading because they were not trusted to grade fairly. An ELA teacher explained her anxieties about one of the teachers: ‘she is disgruntled...she scores harshly...she gives someone who should have gotten a 3 a 1.’ Scoring one’s colleagues as leniently as the rubric allowed was understood to be the appropriate approach and was necessary to being perceived as a good colleague. The administration allowed the teacher to grade the year of my study, but her colleagues were vigilant to put structures in place to intercept any grading issues that may arise with her or others. One ELA instructional leader explained that she would work with two other teachers to double-check how their colleagues scored the post-test: ‘[Three of us] are going to serve as back-scorers...I’m going to be very strict with this stuff...I’m going to say this is really serious stuff.’ The ELA department leaders intentionally heightened the pressure on and surveillance of teachers during their collaborative time when they normed and graded the assessments so as to communicate the importance of the task. ‘[It] has to be a big formal thing...[It] has to be a really formal process, because these are people’s ratings,’ one teacher lectured her colleagues in a department meeting.

My observation of grading sessions not only revealed horizontal surveillance, but also vulnerability as teachers examined the performance of their colleagues’ students. Teachers’ discomfort and nervousness at the transparency required by grading manifested itself as a combination of self-deprecating humor and setting low expectations at grading sessions. In the social studies meeting, one teacher explained apologetically to her colleague: ‘for my class you are not going to have all 3s and 4s...it will be mostly 2s and 3s.’ At the math meeting, one teacher apologized to a department leader, ‘I’m sorry if I make you look bad.’ In response, the department leader teased the teacher, who was known as a strong instructor and had been asked to facilitate professional development sessions for his colleagues: ‘I’ll have to put you on my list, I’ll be in your room every week next year [to observe you].’ ‘[My students] don’t remember anything about math, I am a horrible teacher,’ the teacher lamented. The math teachers’ playful banter sought to lessen any embarrassment or disappointment that might come from their colleagues seeing their students perform poorly. Moreover, grading other teachers’ tests facilitated comparison and constructed teachers’ reputations as weak or strong educators, making grading sessions a high-stakes site where collegial respect could be lost or gained.
Administrator-teacher relationships

Teacher accountability policies introduced policy artifacts into GMS that administrators and teachers had to negotiate or employ together in the evaluation process. In this section, I discuss joint tasks entailed by two artifacts: 1) the teachers’ contract that outlined the components of the evaluation system and each party’s rights in the process and 2) the rubric that set forth a standardized definition of effective teacher practice. Teachers and administrators had to organize their behavior so that they fulfilled but did not overstep the agreements stipulated in the contract, and they had to employ the rubric to assess instructional practice. Both of these joint tasks negatively impacted the administrator-teacher relationship, making it juridical, technical and distant.

Upholding but not overstepping the contract

The teachers’ contract with the district fixed the terms of the evaluation system. It set guidelines for the frequency of observations, what should be accomplished at each stage of the evaluative process, and the responsibilities and rights of each party involved. The contract was a live document in GMS. Administrators and teachers made reference to it in order to define and compel appropriate professional behavior in ways that made administrator-teacher relationships at GMS juridical and technical.

The contract in place at the time of my study had only recently gone into effect. The district had been delayed in receiving the state’s RttT funds because it had previously failed to reach an agreement with the teachers’ and administrators’ unions. Out of contract, the evaluations had been inadmissible and carried no consequence. The legal agreement finally reached represented the outcome of a contentious process in which each stakeholder had jockeyed to advance their own interests. According to one assistant principal at GMS, the teachers’ union had leveraged the pressure the state was under to implement reforms required by RttT to secure more favorable conditions. Administrators succeeded in simplifying their responsibilities, which reduced their workload; they no longer needed to provide artifacts, such as lesson plans, with their observation write-up’s, and the rubric used to assess teacher practice was downsized by the elimination of some categories. I learned details about the negotiation process from administrators and teachers as they worked out the appropriate norms of interaction with each other, suggesting that the agonistic foundation of the contract continued to imbue professional relationships at GMS.

Despite their actual knowledge of the agreement’s terms, administrators and teachers frequently made reference to ‘the contract’ in making sense of their professional responsibilities, rights, and expectations of each other. The administration was careful to act in accordance with the contract to avoid the risk of a teacher filing a grievance. Administrators debated what they were allowed and not allowed to demand from teachers. ‘Can we collect lesson plans?’ one assistant principal asked, and another replied, ‘yes, but you have to give it back the same day, their plans are their property.’ The principal and assistant principals were wary of officially counting the observations they conducted while I shadowed them. The principal explained, ‘I’m probably not allowed to do a rated observation with you in the room…I don’t want to open myself up to being grieved.’ The administration at GMS knew the potential consequences of a grievance. They occasionally made reference to a dispute involving a former teacher
who had contested her dismissal from GMS. The administrators had to go to the central
district office to testify for the case, which had been ongoing for several months. At the
end of the year, the administration developed a schedule to differentiate instruction
better for ELs in the next school year, but this plan was never put into practice because,
according to one administrator, it violated the teachers’ contract: ‘I came up with this
really complicated genius program, [but the teachers did not like it], there were ELA
teachers with six different groups and contractually that is not allowed.’ Upholding but
not overstepping the contract took precedence in a process that would ideally be driven
by the goals of supporting and improving teaching and learning.

The administration also invoked the contract to direct teacher behaviors and to
justify their authority to do so. During staff orientation before the school year began,
GMS’s leadership dedicated several hours to review the district’s regulations around
everything from lesson plans to bathroom passes. Teachers’ instructional obligations
were given equal attention to all other legally mandated guidelines, such as which kinds
of cleaning products educators could independently use in their classrooms. GMS’s
administration distilled a list of approximately forty regulations on a document that
teachers were required to review at home and sign to confirm that they acknowledged
that they would be held accountable if they acted out of compliance. One teacher
reported to me that she thought the beginning of the school year was particularly
‘disorganized’ because the new contract had reduced due process requirements and had
made it easier for the administration to dismiss teachers: ‘there were some teachers here
the first week who have disappeared, the contract has changed due process so that
issues with teachers are dealt with more quickly.’ According to this teacher, some
teachers had been excessed in the first week, which meant that everyone’s teaching
assignments had been shifting around the first month of the school year.

Teachers referenced the contract less frequently than administrators. When they did,
it was typically evoked to make sense of their rights and entitlements. One teacher
explained that the new agreement meant that the teachers would be back-paid for the
years in which they were working out of contract. The teacher who served as the union
representative led the way at GMS in redefining teacher issues as potential breaches of
the agreement. One day, when she discussed with another teacher how to find resources
for instruction, she noted that the contract required the state and district to provide
lesson plan support for the common core learning standards in ELA and math and that,
because they failed to provide this support, no teacher ratings should be negative:
‘you’re not going to get anything from the district...which is why they shouldn’t give
any ELA teachers [a rating of] “ineffective” because they don’t have [any resources] up
online.’ In one staff meeting, she pressed the administration as to whether teachers
would get a negative note in their file if they brought in cleaning wipes to use in their
classrooms because the contract forbid substances unapproved by the maintenance staff
to enter the building. She advocated for the rest of the teaching staff in clarifying
teacher responsibilities and defending teacher rights, and, of all teachers, she most
frequently invoked the contract in negotiating interactions with administrators.

Linking teacher practice to a rubric
In addition to student growth measures, the detailed operationalization of effective
instruction through an evaluative rubric sought to precisely measure teacher quality.
Similar to the use of the contract at GMS, the rubric created to assess instructional practice was invoked by teachers and administrators to negotiate interactions and to compel particular behaviors. Utilized as a mediating device, the rubric created distance between administrators and teachers. The administrators made use of the rubric liberally, applying it to interactions with teachers that were only peripherally related to evaluation. In this way, the rubric gained broad power and reproduced the corrosive culture entailed by teacher accountability policy throughout the day.

Administrators employed the rubric to mediate interactions, which allowed them to be less explicit about exercising their authority over teachers. In one instance, administrators brainstormed about how to phrase a letter to some teachers that required them to attend a workshop on classroom management. Anticipating the sensitivity of singling out particular teachers, especially veteran teachers, for support, they decided to draw upon the language in the teacher evaluation rubric: ‘you are growing in the area of [rubric category], this is an opportunity for you to work on your management that we have provided for you.’ They leveraged the rubric to issue directives indirectly, disconnecting them from teachers rather than opening lines of communication necessary for effective instructional supervision (Reddy et al. 2018).

The administration was especially prone to utilizing the rubric in interactions that were tense, such as when they gave constructive or critical feedback. The principal explained that he had internalized the rubric and no longer needed it in front of him when writing his post-observation reports, however he brought the rubric with him to post-observation meetings with teachers to facilitate giving face-to-face feedback: ‘I can go and give feedback without the rubric, although it is helpful to have the rubric… especially if what you are going to be saying is not going to be reinforcing all the positives.’ The administration hoped that the rubric helped teachers to not take personal interactions personally. Although they sought to avoid potentially negative interactions with teachers, the administrators did not consider how using the rubric as a crutch undermined this goal, since the distance it created eroded professional relationships at GMS.

The rubric simplified the content discussed and vocabulary used in pre- and post-observation conferences between administrators and teachers. The conversations were typically limited to rubric categories and subcategories. Teachers defended their practice and administrators offered feedback through the rubric’s terminology. Both parties seemed to find predictability and thereby safety by constraining their dialogue within the boundaries of the rubric. Using the rubric as a mediating device to avoid confrontation was given priority over deep reflection about and support of instructional practice.

The administration also employed the rubric as a tool to compel teachers’ compliance with their directives in contexts indirectly related to evaluation. One assistant principal asked the teachers on the EL committee she supervised to create binders to document their work for accountability purposes, linking their participation to aspects of the rubric. She suggested that another assistant principal take up the practice for his committee of teachers working on school climate: ‘Why don’t you give them a binder… explain this is for evidence of your professional work…you can use it for [their] rating [too]…it goes under “professional responsibility” for [the teacher rubric].’ In another administrative meeting, she suggested that the other administrators adopt her practice
of requiring teachers to hand in a log of parent outreach every semester to hold them accountable for contacting families during the time designated for this purpose each week. The principal then proposed that the log should be collected each month. Another administrator suggested that the internet-based program GMS used to contact parents should also be employed to monitor teachers, showing how the administration transformed a tool for instruction and parent engagement into an instrument for surveillance and evaluation. Implied in their efforts was the idea that the increased pressure that came with linking a task to the formal evaluation system would either motivate or compel teachers to comply. Research on teacher motivation, however, has shown that external pressures and incentives are less effective than connecting to a teacher’s sense of professional purpose (Nieto 2003). At a staff meeting, one teacher indirectly commented on the ever-expanding realm of tasks that were measured and monitored by joking that the principal should rate the maintenance staff with the teacher rubric after they did not paint the classrooms as requested. Strategically adopting the rubric to further local priorities, the administration’s ‘school-based policy elaboration’ was central to incorporating the policy into GMS’s culture (Braun, Maguire, and Ball 2010), allowing distance to infiltrate teacher-administrator interactions throughout the school day.

**Discussion and implications**

Teacher evaluation systems that purport to precisely measure a teacher’s quality and impact on student learning warrant concern beyond the validity and reliability of their methodologies. By documenting the day-to-day enactment of teacher accountability at GMS, I show how the joint tasks entailed by policy structured interactions in ways that eroded the professional culture. Sharing student test scores and proctoring and grading for colleagues’ students fostered vulnerability and anxiety in teacher-teacher relationships. Administrators’ relationships with teachers were made juridical, technical, and distant as they liberally utilized the contract and the rubric to mediate interactions and to issue directives.

This paper contributes an empirical examination of projections regarding the impact of teacher evaluation reforms on professional culture; a more comprehensive understanding of the opportunities and costs in attempting to precisely measure teacher quality as a means to ensure students receive an adequate education; and an opportunity to identify, and thereby interrupt, the joint tasks required by teacher evaluation that are particularly corrosive to professional culture.

**Contributions to the literature**

The case of GMS suggests that research on the consequences of neoliberal reforms to teacher evaluation needs to be broadened. It is important to pay attention to the role of teachers and administrators in the implementation of teacher evaluation policies, to examine both how possibilities for action and interaction are structured by policy and how actors uniquely negotiate the spaces that they are given.

The current literature on implementation, which relies on theory or self-report, envisages that teacher accountability increases individualism, anxiety, and litigiousness among school professionals. My findings at GMS confirm some of these prognoses,
disconfirm others, and illuminate mechanisms that link policy and professional culture. Within the joint tasks required for evaluation, teachers at GMS did not become more individualistic or competitive. In fact, teachers were made more interdependent and had to coordinate to accomplish several steps of the process that led to their ratings. Coordination did not always entail collaboration, or working together with trust to support student learning and professional growth. The weekly time set aside for teacher collaboration was often consumed with the joint tasks of norming and grading exams, leaving less time for planning together, sharing resources, or providing each other with feedback. The time dedicated to enacting teacher accountability may help explain why these reforms accompany decreased teacher collaboration. My findings at GMS did confirm that neoliberal evaluation policies increase anxiety for teachers and suggest several possible sites that may generate these feelings. Teachers’ reliance on their colleagues’ instruction, proctoring, and grading for their student growth measures and administrators’ use of evaluation rubrics and the contract to ensure compliance in all realms of a teacher’s day may lower staff morale. Although I found no evidence of active lawsuits brought by the community or employees relevant to teacher accountability, I did witness how understandings of the contract that regulated teacher evaluation shaped administrator and teacher interactions and served to define behaviors through a juridical lens.

**Prioritizing professional culture**

Teacher evaluation has a twofold mission: to ensure instructional quality and to support professional growth. Recent reforms to teacher evaluation have thrown these two priorities off balance. No matter how much policymakers and administrators frame teacher evaluation as being a developmental tool, teachers experience it as surveillance and constant measurement. Warren and Ward explain: ‘teachers feel the pressures of accountability, and experience a disconnect between the promise of the policy and what it actually delivers’ (2018, 15). How can we reclaim teacher evaluation so that it supports teachers? Below, I consider how the case of GMS suggests we can mitigate the negative impact of teacher evaluation on professional culture, which is integral for professional growth (Kraft and Papay 2014).

**Implications for policymakers**

Educational policy within the accountability movement has not adequately attended to school climate (Cohen et al. 2009). With the greater flexibility afforded states under ESSA (Amrein-Beardsely and Holloway 2017; Sawchuk 2016) and a turn away from teacher accountability by large foundations (Loewus 2017), policymakers have the opportunity to rethink teacher evaluation systems and to roll-back aspects that undermine professional culture. Efforts by policymakers to respond to recent research by improving the reliability and validity of measures of teacher quality are shortsighted because they do not consider the broader impact of these policies. In fact, such efforts would only reinforce norms of interaction that undermine a positive professional culture, which supports teacher retention and student achievement. The impact on professional culture is one more reason that policymakers should be wary of VAMs and rubrics that attempt to quantify teacher quality. Lawmakers should avoid policies that
require teachers to expend large portions of their collegial time preparing students for, administering, and grading tests. Teachers should not be asked to work on aspects of teacher evaluation that are purely for accountability and do not contribute to their professional development. Moreover, accountability should shift toward holding districts and states responsible for the equitable distribution of quality educators and away from judging individual teachers, which has questionable validity and erodes professional culture. Considering teacher shortages and high rates of teacher turn-over, policymakers need to be more attentive to professional culture.

**Implications for teachers and school administrators**

Through joint tasks, teacher evaluation policy constructed policy subjects at GMS whose interactions were strained and distanced. The character of these relationships can reify over time such that teachers and administrators behave in ways that facilitate neoliberal reforms. Indeed, Holloway and Brass (2017) found that teachers in the US have become invested in the ways that market-based education reforms define their role and their value as educators, in contrast to the advent of accountability policies in the early 2000s which met extensive teacher resistance. However, the construction of policy subjects is not total (Ball et al. 2011). Individuals are granted ‘embedded agency’ within the range of possible action allowed by the policy environment: ‘while institutions constrain action they also provide sources of agency and change’ (Thornton and Ocasio 2008, 101). Teachers and administrators are also policy actors (Ball et al. 2011). The market-based logic of neoliberal teacher evaluation policies is one of several institutional logics available to teachers and administrators for making sense of their professional roles and their possibilities for action: ‘The contradictions inherent in the differentiated set of institutional logics provide individuals, groups, and organizations with cultural resources for transforming individual identities, organizations, and society’ (Thornton and Ocasio 2008, 101). I offer school actors two approaches to countering a market-based logic and to mitigating teacher accountability’s corrosive impact on professional culture: segmenting logics and fostering communities of connection.

**Segmenting logics to contain the influence of teacher accountability.** Goodrick and Reay proffer that multiple logics can coexist ‘by segmenting their impact’ (2011, 379). At any given time, all institutional logics are arranged in a particular constellation, with some more dominant than others (Goodrick and Reay 2011). At GMS, the neoliberal, market logic of teacher accountability was dominant in multiple sites across the working day. Alternative logics, such as a professional logic that values the expertise and judgement of teachers, were present but subordinate. Even though professional and market logics may be in competition with each other, they can coexist if they are made dominant in different aspects of work (Goodrick and Reay 2011). School leaders who seek to support professional growth and student learning should compartmentalize teacher accountability policy within their schools and take steps to contain its impact, such as limiting their use of the contract and evaluative rubric to compel teacher compliance. Administrators have disproportionate influence over the structure of professional interactions in schools and can play a large role in ‘re-imagining the teacher as an intellectual, rather than as a technician or as a bundle of skills and competencies’ (Ball 2016, 1056). Shaped by multiple often contradictory logics, actors already shift between logics throughout their day.
unconsciously (Ball et al. 2011). School administrators and teachers can apply an understanding of their malleability to switch between logics with intention.

**Forging communities of connection.** Interactions construct policy subjects and are necessary to deconstruct them. Identity is shifting and relational, such that new modes of interaction over time can create new subjectivities (Rodgers and Scott 2008). Seemingly mundane interactions in schools can cumulatively shape and reshape organizational norms (Allaire and Firsirotu 1984). In this way, teachers and administrators can resist the corrosive impact of teacher accountability by the simple act of connecting. Everyday efforts to humanize one’s colleagues is a form of radical opposition to the professional distance and strain that support neoliberal education policies. Agency to resist teacher accountability cannot be attained individually. A reform of teacher and administrator identities relies on a community of practice that fosters connection: ‘[a]s a locus of engagement in action, interpersonal relations, shared knowledge, and negotiation of enterprises...communities hold the key to real transformation’ (Wenger 1998, 85). Teachers and administrators can transform their schools by interacting in ways characterized by support, connection, and mutual professional respect. Finding connection in professional interactions is a mechanism of policy reform from the ground up.

**Acknowledgments**

This article is based on research that was made possible in part by financial assistance from the Ruth Landes Memorial Research Fund, a program of the Reed Foundation. Thank you to Dr. Sofia Chapparo and to the reviewers for the 2017 Sociology of Education Association conference for their thoughtful feedback on earlier versions of this paper.

**Disclosure statement**

No potential conflict of interest was reported by the author.

**Notes on contributor**

Rachel Garver is an assistant professor of Educational Leadership at Montclair State University. Her research examines the implementation of equity-oriented policies by teachers and administrators, school discipline and safety practices, and the preparation of educational leaders committed to social justice.

**ORCID**

Rachel Garver [http://orcid.org/0000-0001-7504-8103](http://orcid.org/0000-0001-7504-8103)

**References**


