How Middle School Special and General Educators Make Sense of and Respond to Changes in Teacher Evaluation Policy

Alisha M. B. Braun
University of South Florida

&

Peter Youngs
University of Virginia
United States

Abstract: In this multiple case study, we apply sensemaking theory to examine and compare how middle school special and general educators perceive and respond to teacher evaluation reform, including formal classroom observations, informal walkthroughs, and student growth measures. Our findings reveal that special educators experience conflict between the policy's main elements and their understandings of how to effectively teach students with disabilities. Furthermore, special and general educators held contrasting beliefs regarding the appropriateness of evaluation. Our findings illustrate the importance

of acknowledging differences in special and general educators' roles and responsibilities and encourage policymakers to reconsider uniform teacher evaluation policies.
Keywords: teacher evaluation; accountability; educational reform; sensemaking theory

Cómo los educadores especiales y generales de la escuela intermedia tienen sentido y responden a los cambios en la política de evaluación docente

Resumen: En este estudio de caso múltiple, aplicamos la teoría de la sensemaking para examinar y comparar cómo los educadores especiales y generales de la escuela intermedia perciben y responden a la reforma de evaluación docente, incluidas las observaciones formales en el aula, los recorridos informales y las medidas de crecimiento de los estudiantes. Nuestros hallazgos revelan que los educadores especiales experimentan conflictos entre los elementos principales de la política y su comprensión de cómo enseñar efectivamente a los estudiantes con discapacidades. Además, los educadores especiales y generales tenían creencias contrastantes con respecto a la idoneidad de la evaluación. Nuestros hallazgos ilustran la importancia de reconocer las diferencias en los roles y accountability de los educadores especiales y generales y alientan a los encargados de formular políticas a reconsiderar las políticas uniformes de evaluación docente.
Palabras clave: evaluación docente; accountability; reforma educativa; teoría de la sensemaking

Como os educadores especiais e gerais do ensino médio compreendem e respondem a mudanças na política de avaliação de professores

Resumo: Neste estudo de caso múltiplo, aplicamos a teoria do sensemaking para examinar e comparar como os educadores especiais e gerais do ensino médio percebem e respondem à reforma da avaliação de professores, incluindo observações formais em sala de aula, orientações informais e medidas de crescimento dos alunos. Nossas descobertas revelam que educadores especiais enfrentam conflitos entre os principais elementos da política e seus entendimentos de como ensinar efetivamente os alunos com deficiência. Além disso, educadores especiais e gerais mantinham crenças contrastantes sobre a adequação da avaliação. Nossas descobertas ilustram a importância de reconhecer diferenças nos papéis e accountability de educadores especiais e gerais e incentivar os formuladores de políticas a reconsiderarem políticas uniformes de avaliação de professores.
Keywords: avaliação de professores; accountability; reforma educacional; teoria do sensemaking

How Middle School Special and General Educators Make Sense of and Respond to Changes in Teacher Evaluation Policy

Within the shifting policy landscape of accountability and teacher evaluation reform, increased attention is being paid to the use of classroom observation tools and student growth measures (Bill and Melinda Gates Foundation, 2013). The use of classroom observations to evaluate teachers is common practice, yet there is much inconsistency in the focus, duration, and frequency of classroom observations (Pianta & Hamre, 2009). Moreover, the observational instruments that are used in practice are rarely consistent across states, districts, or schools (Cohen & Goldhaber, 2016). To ensure evidence-based teacher evaluation policy and practice, it is important that principals and school districts use classroom observation tools that have demonstrated reliability and validity as part of their teacher evaluation process (Marx, 2014; Pianta, 2012).

In addition, while there is evidence that some teacher evaluation systems are valid and reliable when used with general educators (e.g., CLASS, FFT), questions remain about the technical properties of many widely-used systems when employed with special education teachers (Jones & Brownell, 2014). This lack of research on the use of general education teacher evaluation methods with special educators includes both classroom observation instruments and student growth measures (Jones, Buzick, & Turkan, 2013). While there may be practical and philosophical value to having school leaders use common evaluation practices for general and special educators, it is important to evaluate this empirically to guide policymaking and to ensure that all teachers are evaluated effectively.

In this multiple case study, we compare the perceptions and experiences of special and general educators with classroom observation instruments and student growth measures as part of a relatively new statewide teacher evaluation system in Virginia. We draw on data from one-on-one interviews with teachers (two special educators and two general educators) working at a Virginia middle school. We focus on mathematics and language arts middle school teachers because of the strong expectations they face under recent accountability policies to promote student achievement in these core academic subject areas. Special and general education middle school teachers alike are subject to individual value-added data accountability requirements, yet there is a relative lack of research comparing their experiences with these new approaches to teacher evaluation. Our analysis reveals that classroom observation practices were not uniform or standardized across multiple teachers within this middle school, nor were school administrators trained to use research-based observation tools. Further, we found that in comparison to general educators, special educators felt that the use of student growth measures to assess teacher performance failed to evaluate a significant component of their job, namely their role as a case manager.

In this paper, we begin by reviewing relevant educational policy and literature on teacher evaluation to contextualize our analysis. In the second section, we present the conceptual framework, sensemaking theory, that undergirds our research design and analysis. Next, we describe our sample, research methods, and analytical strategies. In the fourth section, we present our findings. We close the paper with a discussion of limitations, policy and practice implications, and recommendations to help move teacher evaluation forward.

Policy Context

Teacher accountability reform in the U.S. has been significantly shaped by key federal education policies. The No Child Left Behind Act (NCLB), enacted in 2002, represented a reauthorization of Title I of the federal Elementary and Secondary Education Act (ESEA).

First enacted in 1965, Title I is the main federal policy in the U.S. and is reauthorized by Congress every five years or so; NCLB was the version of Title I that was reauthorized under President George W. Bush. NCLB mandated annual student testing in grades 3 through 8 to hold schools and teachers accountable for students' academic achievement. Accordingly, this testing requirement strongly impacted the practices and experiences of middle school leaders, teachers, and students. The federal government required that states design and administer statewide student assessments in reading and mathematics that are aligned with the state's curriculum standards.

During the Obama administration, President Obama and U.S. Secretary of Education Arne Duncan initiated Race to the Top (RTTT) in 2009, which provided significant financial incentives to states to make changes in their teacher evaluation policies. As a result, many states began requiring school districts to use valid, reliable classroom observation instruments and state student test data to evaluate teachers on an annual basis. In addition, Obama and Duncan offered states Title I ESEA waivers, which relaxed the Adequate Yearly Progress provision of NCLB in exchange for states making changes to teacher evaluation policies and other policies emphasized in RTTT.

The most recent reauthorization of the federal Title I law, the Every Student Succeeds Act (ESSA), was passed in December 2015 and has maintained the requirement that states annually administer reading and math assessments in grades 3 through 8. While the testing requirement has persisted, ESSA returns some autonomy to states to decide how and to what extent they will or will not weigh student assessment data and other components, such as teacher observations, in their revised teacher evaluation policies and systems. We collected the data for this study in 2014-15; i.e., during the NCLB era of accountability.

In response to federal accountability reform, Virginia significantly changed its teacher evaluation system on July 1, 2012 when the *Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers* were issued by the state's department of education. The guidelines outline seven performance standards teachers must be evaluated on: *Professional Knowledge, Instructional Planning, Instructional Delivery, Assessment of/for Student Learning, Learning Environment, Professionalism,* and *Student Academic Progress.* Standard Seven, *Student Academic Progress,* amounts to 40% of a teacher's overall evaluation score, while every other standard contributes only 10%.

In Virginia, the Standards of Learning (SOLs) describe the state's expectations for student learning in core content areas. SOLs are state-level guidelines in Virginia (i.e., standards) for teachers' curricular and instructional practices. Teacher performance is rated as "exemplary," "proficient," "developing/needs improvement," or "unacceptable" individually for each of the seven standards, as well as cumulatively for an overall evaluation summary. Local school districts in Virginia have flexibility regarding how to implement this policy with special educators; many Virginia districts, including the district we focus on in this study, apply the policy in the same way to both general and special education teachers. In addition, the policy does not specify who (e.g., principals, district administrators) is responsible for evaluating special education teachers. Further, this policy is similar to that enacted in many other states in the wake of the federal 2009 Race to the Top initiative (Steinberg & Donaldson, 2016).

In turn, the district where we conducted this study modified its teacher evaluation system to be aligned with the state reform. The evaluation procedures, instruments and materials that were adopted by the district were consistent with the state guidelines and recommendations (Sun, Mutcheson, & Kim, 2016). Consistent with state guidelines, principals collected evidence on teacher performance from three sources: classroom observations, student achievement data, and student surveys. We focused our study on teachers' understanding of and experiences with the first two sources. The evaluation procedures for probationary teachers (similar to non-tenured teachers) included at least three classroom observations per year conducted by school administrators, a midyear review, and a summative annual evaluation. In contrast, teachers with continuing contracts

(similar to tenured teachers) were observed at least once per year and received a summative annual evaluation. In this study, we focused only on teachers with continuing contracts (i.e., experienced teachers). The same evaluation procedures were applied to special educators and general educators.

It is important to highlight that although our study was conducted during the NCLB and RTTT era, our analysis of teachers' perceptions of and responses to the abovementioned policy changes during this time can inform state and district policy decisions as they revise their teacher evaluation systems under the new wave of policy change under ESSA. Since ESSA was enacted in 2015, states and school districts have place notable emphasis on teacher professional development activities and formative teacher evaluation. At the same time, most states continue to require districts to use classroom observation instruments and student growth measures to summatively evaluate teachers on a regular basis.

Literature Review

A growing body of conceptual literature has examined whether teacher evaluation measures originally designed for general educators are appropriate for special educators (Johnson & Semmelroth, 2014; Jones & Brownell, 2014; Jones et al., 2013). At the same time, there has been less empirical research on special educators' perceptions of or experiences with new approaches to teacher evaluation. In this section, we first review scholarship on the appropriateness of classroom observation instruments and student growth measures in evaluating special educators. Then we review literature on general education teachers' experiences with new teacher evaluation measures.

Middle school special educators' work responsibilities and assigned students differ from those of general educators in key ways that have important implications for teacher evaluation. For example, special educators tend to focus their work on individual students with disabilities (SWDs) as opposed to whole classes of students and are typically responsible for such tasks as creating and enacting Individualized Education Programs (IEPs), adapting curriculum and assessments, and overseeing student progress (Jones, 2016). Additionally, special educators frequently provide explicit, teacher-directed instruction to individual students (Johnson & Semmelroth, 2014; Jones & Brownell, 2014). It is important to be mindful of such differences in work responsibilities between special and general educators when evaluating research on how each type of teacher is evaluated.

Researchers have shown that prominent classroom observation instruments, such as the Classroom Assessment Scoring System (CLASS) and the Danielson Framework for Teaching (FFT) are generally appropriate for use with middle school general educators. CLASS reflects a conception of instruction emphasizing three domains of teacher-student interactions: instructional support, classroom organization, and emotional support (Pianta & Hamre, 2009). The FFT addresses four domains: planning and preparation, classroom management, teaching for student learning, and professionalism. While some aspects of CLASS and the FFT are consistent with the work of special educators, researchers have noted that their emphasis on constructivist teaching is inconsistent with special educators' focus on direct instruction (Johnson & Semmelroth, 2014; Jones & Brownell, 2014). In addition, both instruments are designed to assess whole-group instruction (or small-group work within larger classes) while many special educators provide a lot of instruction one-on-one or in very small groups.

Principals and school districts are increasingly employing student growth measures in teacher evaluation; such measures document change in the scores of individual students or groups of students as they move from one grade to the next. Student growth measures can be distinguished from status models, which simply provide data on student performance at a single point in time, which is often compared with a previously determined goal. Value-added models (VAMs) are a form of student growth measure that account for student and school characteristics in assessing an

individual teacher's contribution to the learning of their students (Braun, Chudowsky, & Koenig, 2010; Harris & McCaffrey, 2010).

Researchers have raised a series of concerns about the use of student growth measures with middle school special educators. First, given that many SWDs have testing accommodations, when such accommodations are incorrectly assigned or ineffective, student growth measures can contribute to measurement error and threaten the validity of inferences from growth measures (Johnson & Semmelroth, 2014; Jones et al., 2013). For example, if an SWD needs additional time to take state tests, but such time is not provided, the measure of their growth can contribute to measurement error. Second, a significant percentage of SWDs perform extremely low on state assessments; very low scores due to test difficulty can also lead to measurement error and limit the use of growth measures (Johnson & Semmelroth, 2014; Jones et al., 2013).

Third, it is unclear what constitutes a reasonable rate of academic growth for most SWDs (Johnson & Semmelroth, 2014) on student growth measures. That is, if SWDs begin a given grade performing below grade level, it may be difficult to determine an appropriate expectation for their growth during that year. Fourth, given that many special educators work with SWDs on social and behavioral goals (as well as academic ones), the exclusive focus in student growth measures on academic gains may not capture key aspects of their work. Finally, the nature of special educators' work requires them to collaborate frequently with general educators and to share responsibility for SWDs' academic performance. While student growth measures seek to attribute student achievement gains to individual teachers, this may be less appropriate when middle school special and general educators share responsibility for SWDs' learning and performance.

In contrast to the considerable conceptual scholarship on the use of new teacher evaluation approaches with special educators reviewed above, there is a notable lack of empirical studies on this topic. In one study, Jones and colleagues (2018) examined the use of the Framework for Teaching with 80 special education teachers from Rhode Island and Idaho; they reported that average scores for the instructional domain for the teachers in the sample "were consistently low across all components, with average scores ranging from 1.78 to 2.30" (p.25). The authors note that such consistently low scores in the instructional domain pose a threat to validity. In another analysis, Jones and Gilmour (2018) reviewed state policies related to special educator evaluation; they found that 22 states provided at least some guidance to school districts in this area; of these states, 18 provided directions regarding the use of student outcomes and 12 states provided guidance regarding measures of instructional practices as part of evaluation.

The majority of research on the use of new teacher evaluation systems focuses on general educators, which we review next. In the area of teacher evaluation for general educators, several studies have examined the strengths and limitations of using classroom observations and student growth measures for summative assessment purposes (e.g., Cohen & Goldhaber, 2016; Donaldson & Cobb, 2016; Kim & Youngs, 2016; Rice & Malen, 2016). For example, in a conceptual essay, Cohen and Goldhaber (2016) compared these two types of measures. The authors contrasted what can be learned from each measure type, what conditions are necessary for the accuracy of each one, and the most important sources of error for each. Cohen and Goldhaber pointed out that substantial research on value-added measures has illuminated a variety of concerns about their limitations and biases and may have pushed practitioners to more highly value observation-based measures, when in fact observation instruments face many potential sources of inaccuracy and bias that have not been investigated with the same intensity.

In an empirical study, Donaldson and Cobb (2016) investigated general education teachers' experiences in 14 Connecticut school districts that enacted a new classroom observation instrument as part of the state's new teacher evaluation system. Teachers indicated that they were observed more often under the state's new system than previously and they were experiencing more

accountability pressure; about two-thirds of the sample of 684 teachers reported participating in two formal observations in the first year of the new system. The state's classroom observation rubric provided teachers and principals with a common language for discussing instruction, and 57% of survey participants stated that post-observation conferences with school leaders were somewhat or very valuable (Donaldson & Cobb, 2016).

In another empirical study, Rice and Malen (2016) interviewed more than 100 teachers at two elementary schools and one middle school in a large mid-Atlantic district to explore teachers' experiences with classroom- and school-level VAMs. Teachers reported experiencing accountability pressure associated with teacher evaluation and generally did not believe that "student standardized test scores were a valid measure of either student growth or educator effectiveness" (p.35). In particular, many teachers felt such measures did not accurately represent the academic progress of students with disabilities. Additionally, many did not understand how the VAMs in their district worked and few felt it was reasonable to try to isolate the impact of individual teachers on student learning (Rice & Malen, 2016).

In summary, several empirical studies have examined general educators' experiences with classroom observation instruments and student growth measures. In addition, numerous scholars have written conceptual essays about the strengths and limitations of these measures for use with special education teachers. At the same time, few studies have explicitly compared the experiences of special and general education teachers with new teacher evaluation components. To address this gap in the literature, in this study we compare the perceptions and experiences of middle school special and general educators with a new teacher evaluation system in Virginia.

Sensemaking Theory

To investigate the meanings that special and general educators make of the new statewide teacher evaluation policy and apply to their practice, we draw on sensemaking theory for our theoretical framework. According to sensemaking theory, educators' practices are based on how they construct understandings of information they are exposed to in their environments and how they act based on these understandings (Coburn, 2001; Youngs, Jones, & Low, 2011). The meaning of information about teacher evaluation and effective teaching is often unclear, especially in the context of changes in state and district policy; thus, middle school general and special educators must develop their own interpretations of the expectations placed on them associated with teacher evaluation. Such teachers, then, concentrate on policy messages in their environments and construct understandings of them based on their prior beliefs and practices (Coburn, 2001; Spillane, 2000).

Experienced middle school teachers' attempts to make sense of the expectations they face associated with teacher evaluation are social in a few ways. First, they construct understandings of their roles through social interactions and negotiations. For veteran middle school special educators, for example, through interactions with principals, district administrators, special education colleagues, and general education colleagues, they learn which students they are expected to teach, whether and how they are to modify the general education curriculum for their students, the amount of time they are to work in general educators' classrooms as well as their own classrooms, and how they are to handle student assessment. Second, sensemaking is also social in that it reflects beliefs in the special education profession about the inclusion of students with special needs, use of diagnostic assessments, and appropriate expectations for their academic progress. Such beliefs can shape how experienced special educators make sense of and respond to changes in teacher evaluation.

In this study, we draw on sensemaking theory to compare how middle school special and general educators perceive and make sense of classroom observation instruments and student growth measures as part of their new statewide teacher evaluation system. Sensemaking theory

suggests that experienced teachers will construct understandings of changes in teacher evaluation based on their prior professional experiences and their interactions with teacher colleagues and administrators. Therefore, we drew on sensemaking theory to ask questions in the interviews about study participants' professional backgrounds, opportunities to learn about the new teacher evaluation system, interactions with teacher colleagues about teacher evaluation, experiences with professional development, and ways in which their experiences with teacher evaluation had led them to make changes in their teaching practices.

While both special education and general education middle school teachers must rely on others in their schools for help determining their responses to these changes, they often differ with regard to their work responsibilities and prior professional experiences. Thus, the theoretical framework underlying our study posits that differences in special and general education teachers' professional relationships, roles, and prior experiences shape their understandings of and responses to new teacher evaluation policy.

Methods

District/School Sample

Our research is a multiple case study of four teachers (two general educators and two special educators) at one middle school in a small, rural Virginia school district (Yin, 2014). The district contains a total of six schools (one high school, one middle school, and four elementary schools). As mentioned above, the teacher evaluation procedures, instruments and materials adopted by the district are consistent with state guidelines and recommendations (Sun et al., 2016), so this district serves as a desirable context to examine how teachers respond to changes in state teacher evaluation policy. Furthermore, this small district has a fair amount of resources and is high performing based on its state standardized test scores, with only one of six schools not meeting the Annual Measurable Objectives (AMO) during the year we conducted our study. Accordingly, problematic aspects of the teacher evaluation process experienced by special and general education teachers in this context are likely challenges in response to the new teacher evaluation reform, as opposed to being confused with additional challenges encountered in a high-poverty or lower-achieving district.

In 2014-15 (the year of data collection), the middle school in this study served 907 students; 32% of whom were eligible for free/reduced lunch. Additionally, 78% were White, 12% were African American, and 4% were Hispanic. At the time of the study, the school was accredited by the state based on its state standardized test scores and had met Annual Measurable Objectives (AMO) for all subjects for federal accountability measures. As noted, we focused on the experiences of middle school special and general educators due to (a) the accountability pressures faced by both groups of teachers and (b) the relative lack of research on middle school teachers' experiences with new approaches to teacher evaluation.

Teacher Sample

In our multiple case study design, we interviewed four classroom teachers (two general educators and two special educators) from this middle school to compare the experiences of these two groups of teachers. In selecting teachers to participate in the study, we focused on middle school special and general educators who were responsible for academic instruction in core content areas (mathematics and language arts) assessed in the new teacher evaluation reform. To protect the confidentiality of our participants, we use pseudonyms herein. Particularly, we refer to the two general educators as Ms. Clark, an eighth-grade mathematics teacher with 10 years of teaching experience, and Ms. Lee, a sixth-grade language arts teacher with nine years of teaching experience. The two special educators are Ms. Taylor, who taught language arts in collaborative and self-

contained classrooms and had seven years of teaching experience, and Mr. Harris, who taught seventh-grade mathematics in collaborative and self-contained classrooms and had 34 years of teaching experience. All four teachers held continuing contracts, having worked at this school for at least two years prior to the 2014-15 school year and had multiple years of teaching experience (see Table 1). It is useful to include veteran teachers (i.e., teachers who have job security and who have established their approach to teaching) in a study of teachers' experiences with teacher evaluation as opposed to including early career teachers who are still developing their approach to teaching, who are new to the school, and/or who might leave teaching at the end of the year. The school principal recommended the teachers for interviews based on our inclusion criteria (i.e., experienced, core content area special and general educators) and all of the teachers agreed to participate in our study. In summary, we were interested in representing and comparing the perspectives of experienced middle school teachers in special and general education in core content areas.

Table 1
Participants' Years of Experience and Teaching Assignments

Pseudonym	Years experience	General/special education	Subject	Grade level
Ms. Clark	10	General education	Mathematics	8
Ms. Lee	9	General education	Language arts	6
Ms. Taylor	7	Special education	Language arts	7
Mr. Harris	34	Special education	Mathematics	7

Data Collection

Data collection during the 2014-15 school year involved the authors interviewing two middle school special educators and two middle school general educators twice each (winter 2014 and spring 2015). We interviewed each participant twice to capture their perceptions of the new teacher evaluation process over time as the various evaluation components unfolded across the academic year (i.e., student achievement data collection and analysis, classroom observations, delivery of evaluation feedback). Each one-on-one interview occurred after each teacher in the study had participated in an observation and feedback session, was approximately 60-minutes in duration, and occurred in a quiet location of the participant's choosing (most often in their classroom before or after school hours). In the interviews with both pairs of teachers, we used the same interview protocols to allow for comparison. The interview protocols used are included in Appendix A. We probed to learn about the participants' professional backgrounds and teaching assignments, and the curricular and instructional expectations they experienced. We also asked about the primary goals and key components of their district's teacher evaluation system, including the ways it assessed their instruction and took account of student achievement gains in assessing their performance. In particular, we asked both pairs of participants about the nature of the feedback they received on their teaching and opportunities for professional development. We also probed to learn about how participants' experiences with the formal teacher evaluation process influenced their instruction.

Analytical Strategies

For each round of qualitative data collection (winter 2014 and spring 2015), the authors wrote a detailed analytic memo immediately following each audiotaped interview that described the tone and meaning discerned at the time of the interview. We also transcribed each interview verbatim in its entirety. The authors used NVivo software to analyze data from the interviews to generate initial codes (see Appendix B for our lists of initial and final codes). By grouping categories and using the constant comparative method (Corbin & Strauss, 2015), we moved to higher levels of abstraction and eventually derived the following codes: effects on general educators; effects on school; effects on special educators; general educators' views of classroom observations, walkthroughs, feedback; general educators' views of student growth measures; policy context; proposed changes; special educators' views of classroom observations, walkthroughs, feedback; special educators' views of student growth measures (Corbin & Strauss, 2015; Miles, Huberman, & Saldana, 2014).

Once we established the final codes, the two authors separately coded all eight interviews (*n*=2 special educators x 2 interviews; *n*=2 general educators x 2 interviews). Our separate coding efforts were compared and in cases of disagreement, we discussed the codings until we reached consensus. Prior to discussion of our separate coding efforts, we agreed on 73% of the codings.

For the second stage of qualitative data analysis, we compiled case reports by teaching assignment (i.e., special and general education) and identified emergent themes regarding (a) goals and components of teacher evaluation; (b) feedback and professional development; and (c) influences of teacher evaluation on instruction. We then used these themes to explore possible connections among (a), (b), and (c). This technique is recommended by Corbin and Strauss (2015) as a way to reveal a number of relationships in one's data; in our case, these relationships included those between and among teacher evaluation components, feedback, professional development, and/or changes in instruction. At the same time, several linkages between additional factors and outcomes were evident, including ones between special educators' views of how well the teacher evaluation system took account of their roles and the learning ability of their students.

In ascertaining and describing educators' perceptions of and experiences with teacher evaluation, teaching position was our key analytical unit (i.e., special or general education). Thus, our data analysis involved looking across the expectations associated with teacher evaluation placed on the two special educators; we then repeated this with the two general educators. Similarly, in developing an account of the effects of teacher evaluation on participants, we also grouped them by teaching assignment. This enabled us to analyze the varying role that the principal played for special and general educators and, in particular, to better understand how the principal's approach to teacher evaluation played out differently for the two groups of teachers.

When it became evident that the teacher evaluation process did not take sufficient account of the special educators' roles or the students with whom they worked, we created additional tables by selecting and rearranging our codings of relevant texts from special and general educators about their roles to confirm these patterns in our data while remaining attentive to disconfirming evidence (Miles et al., 2014). In this way, we were able to analyze the ways in which the focus on student achievement gains in teacher evaluation and the inattention to special educators' roles as case managers were associated with how the special educators responded to the new teacher evaluation policy. This also permitted us to consider how the general educators' experiences with teacher evaluation differed markedly from those of the special educators (e.g., better match with their roles, more appropriate feedback, greater access to relevant professional development).

Procedures for Establishing Credibility of Interview Data

In this study, we took three main steps to establish the credibility of the interview data reported on in this manuscript. These included member checks, a multiple case design, and peer review and debriefing (Glesne, 2006; Miles et al., 2014). In terms of member checks (Miles et al., 2014), we met with the four interview participants at the end of the 2014-15 school year to share and get feedback on emergent themes regarding (a) perceptions of and experiences with teacher evaluation, (b) opportunities for feedback and professional development, and (c) effects of teacher evaluation on instruction.

Second, by including special and general educators in our interview sample, we incorporated a multiple case design featuring replication logic (Yin, 2014). More specifically, the inclusion of general educators in our sample enabled us to examine if the two groups of teachers experienced contrasting expectations associated with teacher evaluation and whether special educators' roles and the students with whom they worked were fully accounted for by new approaches to teacher evaluation.

Finally, we received external feedback on our research design and initial findings from faculty colleagues at our universities in the areas of teacher evaluation and special education (Glesne, 2006). In particular, these colleagues encouraged us to restrict the focus of the analysis to middle school teachers to capture nuanced differences between special and general educators' experiences related to teacher evaluation reform in core subjects; if we had included elementary school teachers, the set of responsibilities of special educators would be markedly different than general educators' because both special and general education teachers in middle schools focus on one or two subjects, whereas in elementary school they focus on additional subjects not assessed by the reform.

Results

Perceptions of and Experiences with New Teacher Evaluation System

Data analysis revealed considerable differences between the perceptions and experiences of special and general educators concerning the new teacher evaluation system. One primary difference was that special educators perceived that the new teacher evaluation system only evaluated half of their work responsibilities. As one special educator, Ms. Taylor,[3] explained, "We're evaluated just on the SOL classes we teach, or just on our teaching responsibilities when really, we have two jobs because we're a teacher and we're a case manager, so we have a lot of legal requirements on us." Regarding the case manager component of the position, she went on to explain, "We write legal documents. We have meetings after meetings. We have testing. We have training. We have data. We have progress reporting, so there's a lot of components. It's a whole other job and we're not evaluated on that really at all." The other special educator, Mr. Harris, felt similarly, commenting on how the teacher evaluation system disproportionately emphasized his role as a core content area teacher over his role as a special educator, "I'm more evaluated as a math teacher than I am a special ed teacher or person who's responsible for IEPs as well as SOLs."

The additional case managerial professional duties and responsibilities placed on special educators appeared to be outside of the scope of what was assessed by the new teacher evaluation system, and this was perceived by special educators as being inappropriate. Both special educators explained that they were evaluated based on student growth data and classroom observation of their SOL instruction in the same way as the general educators. However, Ms. Taylor and Mr. Harris felt that being evaluated in the same way as their general education colleagues was not appropriate due to the uniqueness of their profession.

Furthermore, both special educators felt like the half of their responsibilities that were evaluated by the teacher evaluation system were evaluated in ways that were not appropriate for special education students. Once again, this calls into question the credibility of the new teacher evaluation system for special education. As Mr. Harris explained, "Our students are special ed for a reason, they're not going to hit those SOLs as well." From his perspective, the unique learning challenges experienced by his students limits their ability to reach the same academic growth targets as their general education peers. Ms. Taylor further elaborated on this point: "I don't know that it [the student growth measure] is appropriate for special ed…they're expected to show the same amount of growth as a regular student, which they have disabilities so it's not appropriate. I mean, they might not be able to make the same target goal as another kid because they have processing difficulties or decoding or reading."

In contrast to special educators, general educators perceived that the new teacher evaluation system was generally appropriate. That is, the recent changes in teacher evaluation seemed to fit with general educators' instructional practices. However, it is important to point out that this sentiment was limited to teachers in core content areas. Both general educators expressed some reservations about the degree to which the teacher evaluation system was appropriate because not all teachers were assessed in the same way across content areas. Both general educators felt it was inappropriate for the teacher evaluation system to assess teachers delivering content in non-SOL classes such as art, foreign language, physical education, and health differently than that those teaching core content SOL classes including mathematics, science, and language arts. For example, a general educator, Ms. Lee, commented: "There is no computerized test for PE or art so those subjects get to make their own benchmarks and the teacher gets to plan it so that's unfair. Other aspects of it [the new teacher evaluation system] are fair."

Nature of Feedback

A major critique of the new teacher evaluation system by both general and special educators at this school was how their evaluation largely depended on the building administrator who was administering it. All four teachers did not feel that their evaluation was reliable or consistent due to the subjectivity and individual differences of the respective administrator overseeing their evaluation. In other words, the administrators in this study represented the new policy, but they did not enact it in consistent ways. For example, a general educator, Ms. Clark, referred to the personal element of the evaluation in the following comment: "When you have more than one administrator in the building…each person could get a different evaluation regardless. And not that necessarily personality is going to play into it, but what one person perceives as being good, another may not see it as being superior as another. I mean, yeah there's guidelines and that kind of thing, but there's always a certain amount of personal error that can play into it." All four participants consistently expressed their concern that different administrators might evaluate the same thing in a different way. Participants openly shared that they heard many teachers complaining the evaluation system was neither uniform within nor across schools due to individual variation across school administrators. A common proposed change recommended by participants was that the same administrator needs to oversee the evaluation throughout the entire school building to ensure consistency.

In addition to critiquing the reliability of feedback from evaluators, teachers also challenged the usefulness of administrator feedback in relation to adequately assessing all seven performance standards outlined in the new teacher evaluation system. For example, Ms. Lee expressed her concern that, "Everyone's time is limited in schools and sometimes administrators do not have the time that it takes to really observe each of those seven domains for each teacher." Teachers sympathized with the time constraints of administrators, acknowledging that it was not feasible to

expect administrators to spend a sufficient amount of time conducting classroom observations to adequately evaluate all seven performance standards as dictated by the policy. They cited common instances of administrators intending to observe for a full class period, but the intended observation time may be reduced to 20 minutes or even to five minutes due to a variety of practical constraints.

Similar to general educators, special educators also questioned the ability of school administrators to adequately evaluate them on all seven professional standards, particularly given their lack of training and expertise in the field of special education. For example, Ms. Taylor commented, "As we understand it, the middle school administration is in charge of teacher evaluation in the middle school. It does not matter what department you're with. We don't have any idea about how this division-level special ed administrator comes into play in our evaluation. He has never observed me in an academic setting. He has observed me in IEP meetings." This relates to the previous finding that both special educators felt that the new teacher evaluation system failed to assess all components of their professional responsibilities.

This held true in the observation component of the evaluation system in that both special educators were evaluated and observed by school administrators who were not trained in their field. For example, Mr. Harris explained that during classroom observations, "They're [school administrators] not looking at, 'Are you approaching it from the special ed standpoint that you need to?'" In a practical sense, it was difficult or nearly impossible for middle school administrators to evaluate *Professional Standard One: Professional Knowledge* for special educators when they did not have the necessary training or background in that said profession. The professional qualifications and practices of the special educators and their school administrators were misaligned, and special educators questioned how district-level special education administrators (whose professional qualifications were aligned) might be incorporated into their evaluation process to enhance credibility.

The general consensus across all teacher participants was that the new teacher evaluation system was not helpful to their professional development. Instead, teachers perceived it as being primarily procedural in nature. From a sensemaking perspective, all teachers reported that their schemas had not changed (i.e., they did not consider themselves to be "better" teachers) now compared to when the teacher evaluation system was put into place. Teachers acknowledged the procedural nature of the teacher evaluation system by not only directly stating that the evaluation process is procedural, but also by explaining they would be observed twice a year and during their evaluation meetings they would go through scripted questions such as, "Did you do that student evaluation form? Did you have SOLs posted around the room? Did you keep data on the SOLs?"

The emphasis on the procedural nature of the teacher evaluation system is also reflected in comments such as general educator Ms. Clark's, "We may be better at collecting paperwork, but we're not better teachers." Similarly, special educator Ms. Taylor stated, "I am not getting feedback on, 'Am I meeting the students' needs?'" This illustrated her desire to move beyond procedural feedback to what she perceived as her primary role as a teacher. In general, all four teachers felt the evaluation system did not add any value to determining a given teacher's competence. Instead, participants explained that evaluators know if a teacher is a "good" teacher based on their general reputation, stating that it is obvious to administrators who are "good" versus "bad" teachers based on hearsay from students, parents, and colleagues.

It is important to point out that no research-based observation instruments such as the CLASS or FFT mentioned earlier in the literature review were used by evaluators at any point during the evaluation process. Common classroom observation practice consisted of an administrator writing a running script documenting everything that the teacher and his/her students did in real time during a portion of a class period or during a five-minute walkthrough on their tablet, and then sending that script to the teacher once they exited the room. Each of the two times a teacher was

observed, he/she then had a post-observation conference with their evaluator where they provided context to the lesson and went through the procedural questions such as the examples stated above.

Apparent Effects of New Teacher Evaluation System on Practice

General educators reported that the new teacher evaluation system had not significantly changed their practice. Both general educators acknowledged the value that feedback from classroom observation and student data could have for practice. At the same time, they indicated they were already using such feedback to inform their practice prior to the new evaluation system being put into place. For example, Ms. Lee stated, "I've always done that, it's just that now it has to be done partway through the school year instead of the end…It [the new teacher evaluation system] hasn't affected my work as a teacher. I'm self-driven. I just do my job and don't worry about the rest of it." Likewise, Ms. Clark explained the new teacher evaluation system had not notably changed her practice because she has always done pre-assessments and post-assessments with her students, although she did not previously go into the same level of detailed analysis by asking, for example, how many points as an aggregate her students increased. This general lack of impact on practice is nicely summarized in Ms. Lee's comment: "I feel like I've always been a good teacher. I'm just proving it now."

In contrast to general educators, special educators reported that the new teacher evaluation system had affected their beliefs and practice by focusing instruction on grade-level SOL content and teaching to the test. They both cited the fact that under the new evaluation system, 40% of a teacher's overall evaluation score was based on *Professional Standard Seven: Student Academic Progress*, which had created concern among teachers and encouraged teaching to the test. Special educators criticized such emphasis on student growth data for being too narrow in scope and for restricting classroom instruction to focus only on SOL content. Ms. Taylor explained, "It's a very narrow band measure and if we put too much weight on it, then it forces teachers to gear toward that narrow banding. And that's why I've changed what I did this year to try and get my kids to pass that SOL test."

This effect of the new teacher evaluation system on practice was not perceived in a positive light, as illustrated by the following remark by special educator Mr. Harris, "We're not supposed to teach to the test but how can you not teach to the format of the test? I've changed how I teach out to the group to try and get them to get through the test. But am I improving my education to the kids? I don't think so." The following comment by Ms. Taylor mirrors this sentiment, "It's so hard because I feel like I'd be a better teacher if I could teach to [my students'] individual needs, not necessarily get through all this grade-level content so they can take the test." The professional beliefs and identities of both special educators appeared to be in conflict with the effects of the policy on their practice, while the professional beliefs and identities of the general educators appear to be more aligned with the effects of the policy.

Discussion

In this multiple case study, we examined four experienced middle school teachers' perceptions of and experiences with a new teacher evaluation system in Virginia. The purpose of our study was to compare two special educators and two general educators with regard to their perceptions of the new system, their experiences receiving feedback on their teaching, and ways in which the new system seemed to affect their practice. In this section, we discuss our main findings in relation to prior research on teacher evaluation; discuss implications for policy, practice, and future research; and identify some limitations.

The two special educators in the study reported that the new teacher evaluation system did not take account of their roles developing and managing IEPs and both felt that the student growth measures on which they were evaluated were not appropriate for students with disabilities. In contrast, the two general educators viewed the new system as appropriate. This finding extends prior conceptual literature suggesting ways in which special educators' work responsibilities differ from those of general educators (Johnson & Semmelroth, 2014; Jones, 2016; Jones & Brownell, 2014) by identifying associations between these differing roles and responsibilities and the divergent perceptions of special educators compared to general educators regarding the appropriateness of teacher evaluation methods. Additionally, the finding that the new teacher evaluation system failed to acknowledge the learning needs of students with disabilities empirically compliments conceptual critiques of prominent observation instruments and VAMs for failing to address the needs of such students (see, for example, Jones et al., 2013).

Additional findings highlighting the unique contribution of our study of explicitly comparing the experiences of special and general education teachers concern apparent differences in the impact of the new teacher evaluation system on the instructional beliefs and practices of the two groups of teachers. The two special educators reported changing their instructional practices in response to the new evaluation system. In particular, they oriented their instruction towards grade-level SOL content even though this emphasis was not consistent with their beliefs about effective instruction for SWDs. In other words, there was a conflict between the main elements of the new policy and the beliefs of these special educators regarding effective teaching; this conflict left the special educators very critical of the new evaluation system. In contrast, the general educators indicated they had already been teaching in the ways emphasized by the new evaluation system. Thus, while the special educators seemed to change their instruction in response to the new teacher evaluation system and experienced conflict between the new policy and their beliefs about effective instruction, the general educators neither changed their instructional practices nor their beliefs.

In this study, we investigated the nature of the feedback that teachers received as part of teacher evaluation. All four participants raised concerns about variation in the implementation of classroom observations across building administrators. In addition, they felt the evaluation process focused on paperwork and questioned the ability of their administrators to evaluate them with regard to all seven of Virginia's professional teaching standards. While research on general educators has shown that useful feedback on instruction, especially feedback from principals, can promote pedagogical improvement (Donaldson & Cobb, 2016; Sun et al., 2016), there was little evidence that the general or special educators in this study received such feedback.

Implications

Our findings lead to a series of potential implications for policy, practice, and future research. First, as states revise their teacher evaluation policies and systems under ESSA, it is important to acknowledge differences between special and general educators with regard to their roles and responsibilities and to consider whether uniform teacher evaluation policies are appropriate for both groups of teachers. To ensure effective teaching to meet the needs of all students, policymakers might consider the unique roles and responsibilities of special and general educators throughout the teacher evaluation process. For example, policymakers may decide to require the use of different evaluation methods to assess special and general educators that are specifically designed to evaluate their different roles and responsibilities. Continued research on the utility and effects of using a uniform versus differing teacher evaluation approach for special and general educators is needed to guide these policy changes.

Second, while scholars have made many conceptual arguments about the importance of integrating feedback and professional development with teacher evaluation, there are fewer empirical

studies that have documented instances of principals providing useful feedback to teachers to help improve their teaching. More research in this area is warranted to inform how best principals can provide useful feedback to teachers. Building upon the previous implication and recommendation, just as policymakers might consider mandating the use of different evaluation methods for special and general educators, they might also consider encouraging school leaders to provide different recommendations for teacher professional development that are not only aligned with their evaluation feedback, but also designed to address identified areas of improvement which are uniquely applicable to the professional roles and responsibilities of the two types of teachers.

Third, the use of sensemaking theory (Coburn, 2001) in this study helps explain conflicts that special educators experience between their beliefs regarding effective instruction and the demands of new teacher evaluation policies. It seems important for policy actors to support the differing ways that special and general educators make sense of and respond to teacher evaluation policies to reduce the likelihood that such conflicts will arise. To be proactive in this regard, it may make sense for district and school leaders to look for ways that special and general educators can work together to align their professional identities, teacher evaluation expectations, and overall school goals and mission to jointly support student learning and success.

There were a few limitations to note in this study. First, the study only included four teachers at one middle school in a small town. While this constrained focus provides a unique opportunity to examine a very particular experience, it is important to acknowledge that it is not possible to generalize the findings from this study to a larger number of middle schools or to middle schools in other contexts. Second, the analysis relied primarily on interviews with the four study participants; we did not observe the participants during instruction or when they met with administrators as part of the evaluation process. Finally, we did not collect data on teachers' evaluation ratings, so were unable to link teachers' responses to our interview questions with ratings of their effectiveness. Future work aimed at understanding the effect that summative teacher evaluation ratings may have on teacher perception of the new evaluation system over time could be a valuable extension of this research.

Conclusion

Our study addresses the lack of research on the use of general education teacher evaluation methods with special educators by comparing the perceptions and experiences of special and general education middle school teachers with Virginia's statewide teacher evaluation system. While there may be practical and philosophical value to having common evaluation practices for general and special educators, our findings raise concerns about such practices and policies. In this study, the special educators experienced conflict between the main elements of the teacher evaluation policy and their beliefs about effective teaching for SWDs. In contrast to general educators, they did not feel that the teacher evaluation methods were appropriate. As the teacher evaluation policy landscape continues to be reshaped as part of ESSA, it is especially important for policymakers to address such concerns regarding the use of classroom observation tools and student growth measures to ensure that all teachers are evaluated effectively.

References

Bill and Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study.* Author.

Braun, H., Chudowsky, N., & Koenig, J. (2010). *Getting value out of value-added: Report of a workshop.* National Research Council Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Accountability.

Coburn, C. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23*(2), 145-170. https://doi.org/10.3102/01623737023002145

Cohen, J., & Goldhaber, D. (2016). Building a more complete understanding of teacher evaluation using classroom observations. *Educational Researcher, 45*(6), 378-387. https://doi.org/10.3102/0013189X16659442

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research, 4th edition.* Sage.

Donaldson, M.L., & Cobb, C.D. (2016). Implementing student learning objectives and classroom observations in Connecticut's teacher evaluation system. In J.A. Grissom & P. Youngs (Eds.), *Making the most of multiple measures: The impacts and challenges of implementing rigorous teacher evaluation systems.* Teachers College Press.

Glesne, C. (2006). *Becoming qualitative researchers: An introduction* (4th ed.). Pearson.

Harris, D. N., & McCaffrey, D. (2010). Valued-added: Assessing teachers' contributions to student achievement. In M. Kennedy (Ed.), *Teacher assessment and teacher quality: A handbook.* Jossey-Bass.

Johnson, E., & Semmelroth, C.L. (2014). Special education teacher evaluation: Why it matters, what makes it challenging, and how to address these challenges. *Assessment for Effective Intervention, 39,* 71-82. https://doi.org/10.1177/1534508413513315

Jones, N. (2016). Special education teacher evaluation: An examination of critical issues and recommendations for practice. In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 63-76). Teacher's College Press.

Jones, N., Bell, C., Brownell, M., Qi, Y., Peyton, D., Pua., Fowler, M., & Holtzman, S. (2018). *Using classroom observations in the evaluation of special education teachers. Paper* presented at the annual meeting of the American Educational Research Association, New York.

Jones, N.D., & Brownell, M.T. (2014). Examining the use of classroom observations in the evaluation of special education teachers. *Assessment for Effective Intervention, 39*(2), 112-24.https://doi.org/10.1177/1534508413514103

Jones, N., Buzick, H., & Turkan, S. (2013). Including students with disabilities and English language learners in measures of educator effectiveness. *Educational Researcher, 42*(4), 234-241. https://doi.org/10.3102/0013189X12468211

Jones, N., & Gilmour, A. (2018). Special education teacher evaluation: Examining current practices and research. In J. B. Crockett, B. S. Billingsley, & M. L. Boscardin (Eds.), *Handbook of leadership and administration for special education.* Routledge.

Kim, J., & Youngs, P. (2016). Promoting instructional improvement or resistance? A comparative study of teachers' perceptions of teacher evaluation policy in Korea and the United States. *Compare, 46*(5), 723-744. https://doi.org/10.1080/03057925.2015.1057478

Marx, R. W. (2014). Reforming again: Now teachers. *Teachers College Record, 116,* 1-16.

Miles, M. B., Huberman, A. M., & Saldana, J. M. (2014). *Qualitative data analysis: A methods sourcebook, 3rd Edition.* Sage.

Pianta, R. C. (2012). *Implementing observation protocols: Lessons for K-12 education from the field of early childhood education.* Center for American Progress.

Pianta, R., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109-119. https://doi.org/10.3102/0013189X09332374

Rice, J. K., & Malen, B. (2016). When theoretical models meet school realities: Educator responses to student growth measures in an incentive pay program. In K. Kappler Hewitt & A. Amrein-Beardsley (Eds.), *Student growth measures in policy and practice: Intended and unintended consequences of high-stakes teacher evaluations* (pp. 29-47). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-53901-4_2

Spillane, J. P. (2000). Cognition and policy implementation: District policymakers and the reform of mathematics education. *Cognition and Instruction, 18*(2), 141-179. https://doi.org/10.1207/S1532690XCI1802_01

Steinberg, M. P., & Donaldson, M. L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy*, *11*(3), 340-359. https://doi.org/10.1162/EDFP_a_00186

Sun, M., Mutcheson, R. B., & Kim, J. (2016). Teachers' use of evaluation for instructional improvement and school supports for such use. In J. A. Grissom & P. Youngs (Eds.), *Improving teacher evaluation systems: Making the most of multiple measures* (pp. 102-15). Teacher's College Press.

Virginia Department of Education. (2012). Guidelines for uniform performance standards and evaluation criteria for teachers. Author. Retrieved from www.doe.virginia.gov/teaching/regulations/2011_guidelines_uniform_performance_standards_evaluation_criteria.pdf

Yin, R.K. (2014). *Case study research: Design and methods, 5ᵗʰ Edition.* Sage.

Youngs, P., Jones, N., & Low, M. (2011). How beginning special and general education elementary teachers negotiate role expectations and access professional resources. *Teachers College Record, 113*(7), 1506-1540.

Appendix A

Interview Protocols

Implementation of Virginia New Teacher Evaluation System Teacher Interview Protocol: Winter 2014

This interview is part of a research study on the new teacher evaluation system in Virginia. The goal of the research is to better understand factors that influence the ways in which principals and teachers implement changes in teacher evaluation.

1.   Please describe your background in education, your current teaching position, and how long you have been teaching at this school.

2.   What is the primary goal(s) of the new teacher evaluation system?

   - Improving instruction
   - Identifying effective teachers
   - Identifying ineffective teachers
   - Improving student learning
   - Other goal(s)

2.   How does the teacher evaluation system assess your instructional practices?

3.   Do you have any questions or concerns about this part of the process as it applies to your work as a (e.g., special education teacher, ELL teacher, high school English teacher, etc.)?

5.   How does the teacher evaluation system consider student learning gains in assessing your performance?

6.   Do you have any questions or concerns about this part of the process as it applies to your work as a (special education teacher, ELL teacher, high school English teacher etc.)?

7.   In what other areas does the system assess your performance?

8.   Do you have questions or concerns about these assessments of your performance as a (special education teacher, ELL teacher, high school English teacher etc.)?

9.   How did you learn about the teacher evaluation system?

10.   Think of the teachers in this school with you discuss professional issues most often. How often do you talk with them about curriculum, lesson planning, and/or student assessment? How often do you talk with them about teacher evaluation?

11.   Did you receive feedback in 2013-14 from classroom observations or other parts of the evaluation process? If so, when do you receive the feedback?

12.   How do you receive it and to what extent is the feedback timely and useful?

13.   What questions and concerns do you have about this part of the process?

Implementation of Virginia New Teacher Evaluation System Teacher Interview Protocol: Spring 2015

This interview is part of a research study on the new teacher evaluation system in Virginia. The goal of the research is to better understand factors that influence the ways in which principals and teachers implement changes in teacher evaluation.

1.   At this point in the school year, how would you describe the primary goal(s) of the new teacher evaluation system?

- Improving instruction
- Identifying effective teachers
- Identifying ineffective teachers
- Improving student learning
- Other goal(s)

2.   Think of the teachers in this school with you discuss professional issues most often. How often do you talk with them about curriculum, lesson planning, and/or student assessment? How often do you talk with them about teacher evaluation?

3.   Did you receive feedback in 2014-15 from classroom observations or other parts of the evaluation process? If so, when do you receive the feedback?

4.   What options do you have for appealing a rating of your performance?

5.   What professional development and other kinds of assistance, if any, are readily available to help teachers address areas of concern that may be identified in the evaluation process?

6.   How useful is the professional development and other assistance to teachers in your area?

7.   How has your involvement in the teacher evaluation system affected your work as a teacher, if at all?

8.   What changes, if any, have taken place in your school that are attributable to the teacher evaluation system?

9.   Overall, do you consider the teacher evaluation system to be fair and helpful to teachers and to students? Why or why not?

10.   Looking at your experiences, what changes in the teacher evaluation system would you suggest?

Appendix B

List of Initial and Final Codes

Initial codes	Final codes
1. Policy Context; Teacher Evaluation Process	1A. Policy Context; Teacher Evaluation Process
2. Teachers' Views of Teacher Evaluation Process	1B. How Teachers Learned About the New Teacher Evaluation Process
2A. Gen Ed Teachers' Views of Classroom Observations, Walkthroughs	2. Teachers' Views of Teacher Evaluation Process
2B. Spec Ed Teachers' Views of Classroom Observations, Walkthroughs	2A. Gen Ed Teachers' General Views of the Teacher Evaluation Process
2C. Gen Ed Teachers' Views of Student Growth Measures	2B. Spec Ed Teachers' General Views of the Teacher Evaluation Process
2D. Spec Ed Teachers' Views of Student Growth Measures	2C. Gen Ed Teachers' Views of Classroom Observations, Walkthroughs, Feedback
2E. Gen Ed Teachers' Concerns	2D. Spec Ed Teachers' Views of Classroom Observations, Walkthroughs, Feedback
2F. Spec Ed Teachers' Concerns	2E. Gen Ed Teachers' Views of Student Growth Measures
3. Effects on Teachers and School	2F. Spec Ed Teachers' Views of Student Growth Measures
3A. Effects on Gen Ed Teachers	2G. Gen Ed Teachers' Views of Other Standards
3B. Effects on Spec Ed Teachers	2H. Spec Ed Teachers' Views of Other Standards
3C. Effects on School (Gen Ed, Spec Ed)	3. Effects on Teachers and School
	3A. Effects on Gen Ed Teachers
	3B. Effects on Spec Ed Teachers
	3C. Effects on School (Gen Ed, Spec Ed, Students)
	3D. Proposed Changes

About the Authors

Alisha M. B. Braun
University of South Florida
alishabraun@usf.edu
ORCID https://orcid.org/0000-0001-7032-2572
Alisha M. B. Braun, Ph.D. is an Assistant Professor of Social Foundations in the Department of Educational and Psychological Studies at the University of South Florida. Her research explores educational policies and practices related to access to quality education for marginalized children, with a focus on children with disabilities in developing contexts.

Peter Youngs
University of Virginia
pay2n@virginia.edu
ORCID https://orcid.org/0000-0002-1711-1749
Peter Youngs, Ph.D. is a Professor of Education in the Department of Curriculum, Instruction and Special Education at the University of Virginia. His research explores the effects of educational policy and school social context on teaching and learning in the core academic subjects. His work has a special focus on the relationship between policy and practice in the areas of teacher education, induction, evaluation, and professional development.

About the Guest Editor

Audrey Amrein-Beardsley
Arizona State University
audrey.beardsley@asu.edu
Audrey Amrein-Beardsley, PhD., is a Professor in the Mary Lou Fulton Teachers College at Arizona State University. Her research focuses on the use of value-added models (VAMs) in and across states before and since the passage of the Every Student Succeeds Act (ESSA). More specifically, she is conducting validation studies on multiple system components, as well as serving as an expert witness in many legal cases surrounding the (mis)use of VAM-based output.