ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/342010366

Policies That Define Instruction: A Systematic Review of States' and Districts' Recommendations for Evaluating Special Educators

Article *in* Educational Researcher · June 2020

DOI: 10.3102/0013189X20935039

CITATIONS
0

READS
39

2 authors:

Allison Gilmour
Temple University
18 PUBLICATIONS **197** CITATIONS

SEE PROFILE

Nathan D. Jones
Boston University
26 PUBLICATIONS **315** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project

The Day Reconstruction Method: A New Tool for Measuring Teachers' Work, Work Contexts and Affect View project

All content following this page was uploaded by Allison Gilmour on 08 June 2020.

The user has requested enhancement of the downloaded file.

Policies that Define Instruction: A Systematic Review of States' and Districts' Recommendations for Evaluating Special Educators

Allison F. Gilmour[1]

Temple University

Nathan D. Jones[2]

Boston University

This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.3102/0013189X20935039

Citation:

Gilmour, A. F., & Jones, N. D. (2020). Policies that define instruction: A systematic review of states' and districts' recommendations for evaluating special educators. *Accepted in Educational Researcher.* https://doi.org/10.3102/0013189X20935039

Author Note

We would like to thank Cathy Moyer for her assistance with coding evaluation rubrics.

[1] Corresponding author. 1301 Cecil B. Moore Ave., Ritter Hall 362, Philadelphia, PA 19122. allison.gilmour@temple.edu. 215-204-5413.
[2] 2 Silber Way, Boston, MA 02215. ndjones@bu.edu. 617-353-5233.

Abstract

Educational policies addressing instruction may fail to acknowledge that effective instruction is not the same for all learners. We reviewed teacher evaluation systems across all states and the 25 largest districts to determine how states and districts approach the evaluation of special education teachers, a policy aimed at improving teaching effectiveness. We found that most states and districts did not provide guidance to schools for adapting evaluation systems for these teachers. Some states provided guidance on technical aspects of special education teacher evaluation, such as incorporating student achievement into special education teachers' scores. Districts were more likely to focus on instructional considerations. We discuss the implications of these findings for policies that aim to promote the use of effective instructional practices.

Policies that Define Instruction: A Systematic Review of States' and Districts' Recommendations for Evaluating Special Educators

At the heart of teacher evaluation reform lies the goal of changing teachers' instruction to improve teaching quality for all students. Evaluation systems achieve this goal by (1) identifying effective and ineffective teachers for human capital decisions, and (2) promoting teacher development by changing instruction (Donaldson & Papay, 2014; Kraft & Gilmour, 2017; Papay, 2012; Steinberg & Donaldson, 2016). To assess if and how evaluation improves teaching quality, researchers have studied the validity and reliability of different measures of teaching effectiveness (e.g., Kane et al., 2012; Kane et al., 2013), the impact of evaluation systems on student outcomes, (e.g., Dee & Wyckoff, 2015; Taylor & Tyler, 2012), and system implementation by educational leaders (e.g., Bell et al., 2018; Donaldson & Woulfin, 2018; Grissom & Bartanen, 2018; Kraft & Gilmour, 2016; Kraft & Gilmour, 2017). Collectively, these studies suggest the promise of using teacher evaluations for driving educational change through improving teaching quality and illustrate implementation challenges.

What these studies have not addressed is if evaluation systems are appropriate for all teachers. Teachers within and across schools work with students with varying needs and take on roles that may not align with traditional conceptions of classroom teachers (i.e., one teacher working with a group of students), particularly when they provide critical academic support to students who have specific or significant academic or social-emotional needs (e.g., reading specialists, special educators, emotional support teachers, and teachers of English learners). The existence of these specialized roles highlights the reality that a common set of tools for evaluating and improving instruction may not meet the needs of all educators. In their seminal paper on teacher observations, Hill and Grossman (2013) raised concern that general observation

tools may not capture the relevant content for a given teacher's assignment nor be sufficiently targeted to provide actionable feedback for improving teachers' instruction. If evaluation systems promote models of teaching effectiveness that do not map onto the instruction of key subgroups of teachers, it could result in inaccurate information about which teachers are effective, be of limited utility for teachers in non-traditional roles, and, worse, promote development in instructional practices that are ineffective for addressing students' specific needs.

These questions are particularly salient for teachers of students with disabilities (SWDs). Special education teachers (SETs) have unique roles and responsibilities, incorporating student academic growth into SETs' scores is challenging and sometimes impossible, and the observation rubrics commonly adopted by districts may prioritize instruction that conflicts with evidence-based instructional practices in special education (e.g., Brownell & Jones, 2015; Jones et al., 2013; Jones & Brownell, 2014). SETs are also a critical subgroup for states and districts. First, the continued gap in academic performance between SWDs and their peers warrants urgency in improving the effectiveness of their teachers (Gilmour et al., 2019). Second, the continued shortages in the SET workforce (Dee & Goldhaber, 2017), exacerbated by heightened levels of attrition (Billinglsey & Bettini, 2019), demand the identification and provision of supports to highly-effective SETs to ensure they remain in the classroom. More generally, SETs provide a case for exploring how states and districts are implementing policies that focus on instruction in ways that are relevant and useful to all teachers.

If states and districts are to build evaluation systems to support SET instructional improvement, they will need to both acknowledge challenges of evaluating SETs and adopt strategies to ensure that the evaluation process is accurate and useable. The guidance currently provided by policymakers at the state and district level provides insight into the extent to which a

problem is viewed as important. We used this framework to guide our evaluation of *whether* each state and the 25 largest districts, in addition to the District of Columbia, provided guidance regarding SET evaluation and *what* this guidance addressed.

The Challenges of Evaluating Special Education Teachers

Across the US, most teacher evaluation systems include at least two sources of information, observation scores and some measure of student academic progress (Steinberg & Donaldson, 2016). Both of these components present challenges when evaluating SETs due to SETs' roles within schools, the instruction SETs should provide, and the students that SETs instruct (Buzick & Jones, 2015; Jones et al., 2013; Steinbrecher et al., 2014).

Challenges Due to Roles and Responsibilities

A first challenge to evaluating SETs is that their *roles* in schools vary widely. Some instruct classrooms of students similar to general education teachers; others share instructional responsibilities with a general education teacher. Others work with individuals or small groups of students. Classroom observation systems that aim to measure traditional classroom teachers (i.e., one teacher with multiple students) may not apply (Jones & Brownell, 2014; Steinbrecher et al., 2014). These multiple roles and and models for delivering instruction also present challenges to incorporating student growth into SETs' evaluation scores (Hock & Isenberg, 2017; Steinbrecher et al., 2014).

Comprehensive evaluation systems also need to attend to SETs' abilities to handle non-instructional *responsibilities*. SETs often spend a smaller proportion of their time directly providing instruction than other teachers (Vannest et al., 2011), complicating the use of classroom observation systems to evaluate their effectiveness. SETs collaborate and consult with other teachers and service providers, coordinate students' services, such as speech therapy, and

write and implement individualized education plans ([IEPs]; Sledge & Pazey, 2013). A rubric that only addresses the instruction a SET provides to students fails to support development in these non-instructional requirements of an effective SET (Brownell & Jones, 2015).

Effective Instruction in Special Education

Another challenge in observing SETs using systems developed in general education is that the two fields conceptualize effective *instruction* in different ways (Cohen, 2018; Jones & Brownell, 2014; Mathews et al., 2020). As Jones and Brownell (2014) argue, special education prioritizes instruction that is teacher-led, individualized, and focused in explicit skills. In contrast, general education teachers are expected to guide students in constructing knowledge. This definition of "good teaching" from a constructivist approach is reflected in commonly used observation rubrics (e.g., the Framework for Teaching [FFT]; Johnson & Semmelroth, 2014; Jones & Brownell, 2014; Mathews et al., 2020). Special education administrators, individuals at the state and district level in charge of special education services, recognize this tension in how instructional effectiveness is defined. One national survey of special education administrators found that 49.9% agreed that SET and general education teacher effectiveness should not be assessed using the same evaluation, and the majority (84%) agreed that SETs need different skills than general education teachers (Holdheide et al., 2010).

Measuring Student Learning

A final challenge is how evaluation systems address measures of *student achievement* for SETs. Assessment conditions and the low scores of many SWDs influence the reliability and validity of SETs' scores from value-added measures (VAMs; measures that estimate the average contribution of a teacher to students' academic growth; Buzick & Jones, 2015; Jones et al., 2013; Steinbrecher et al., 2014). A large proportion of SETs cannot receive VAM scores because they

teach untested areas or students who take an alternate assessment. Some have recommended that, for these teachers, students' IEPs could serve as a suitable measure of academic growth (Council for Exceptional Children, 2012), but others have raised concerns about linking teacher accountability to students' IEP goals (Brownell & Jones, 2015; Holdheide et al., 2012). Evaluating SETs using their students' IEP goals could contaminate the intended purpose of the goals. Rather than focusing on setting ambitious goals that drive students' educational services and outcomes, using IEP goals to evaluate teachers could incentivize setting easily attainable goals. Though policymakers and educational leaders have widely recognized that one aspect of evaluating teachers should include the learning of their students, questions remain regarding how to incorporate student outcomes in SETs' evaluation scores.

Purpose

Special educators are challenging to evaluate due to their unique roles and responsibilities, their use of instructional approaches not prioritized in general education, and the concerns associated with including student outcomes in effectiveness scores. In light of these challenges, we examined if states and districts provided guidance for evaluating SETs and the content of the guidance. The results of this study help to identify the specific challenges that states and districts recognized regarding evaluating teachers who work with students with unique needs and fulfill non-standard roles in schools. More generally, this case illustrates the challenges and concerns of policies that aim to address single-faceted views of effective teaching.

Methods

Sample

Our sample included all 50 states and the 25 largest school districts in the country and the District of Columbia (DC; see Table 1). We focused on the largest districts and DC because many large districts were early adopters of evaluation reforms and to parallel other analyses of teacher evaluation policies (e.g., Steinberg & Donaldson, 2016). We used a multistep process to identify evaluation materials. We visited state and district websites and downloaded all publicly available teacher evaluation information, including: evaluation guides (i.e., materials providing directions for teachers and evaluators), observation rubrics (i.e., descriptions of practice with associated rating scales), training materials (i.e., presentation slides for training teachers or evaluators), frequently asked questions (FAQs), memorandums, crosswalks (i.e., documents that link two sets of information on teachers), and scenarios (i.e., vignettes that describe teacher practice or context, and link the descritions to the evaluation system). We took a broad approach in order to capture any and all information available to school-personnel regarding teacher evaluation. When we identified duplicate documents, we included the most recently dated document. We contacted state and district personnel by e-mail and phone to identify evaluation materials that were not publicly available and confirm that the materials we collected were current and relevant. We identified who to contact through state and district websites regarding teacher evaluation, teaching quality, or personnel. When we could not identify a state or district employee to contact directly, we submitted general requests through contact forms on websites. Six districts (23.1%) and 13 states (26.0%) did not respond to requests to confirm the accuracy of the documents we identified for inclusion. Of the states that responded, all indicated that the materials we identified through our search were relevant and six provided additional information. Five districts shared documents that we had not identified in our search.

Coding

After identifying all available teacher evaluation documents, we developed a coding manual to use in our document analysis. The codes addressed (1) whether the state provided any evaluation guidance that addressed SETs and (2) the content of this guidance. We coded a state or district as having guidance regarding SET evaluation if the materials we identified included any mention of SETs or related synonyms (e.g., teachers of SWDs). We then categorized the documents that states provided into non-mutually exclusive types: new or modified observation system or rubric that had been created for SETs, general informational documents mentioning how to evaluate SETs, or training materials for school or district administrators. We used the first set of codes to contextualize where the later content findings were addressed and to delineate more general guidance from complete observation systems focused on SETs. In line with the challenges outlined earlier in the paper, we coded the content of materials for if they addressed (1) SETs' roles, (2) SETs' procedural responsibilities, (3) instructional considerations, and/or (4) incorporating student achievement into SETs' ratings.

We coded documents as addressing SETs' roles if the document addressed SETs acting as consultants or without their own classrooms, co-teaching, pull-out instruction, or small group instruction. For example, some states and districts offered different guidance materials depending on whether SETs worked in resource rooms or as co-teachers. We defined procedural responsibilities as activities related to writing IEPs, tracking IEP progress, organizing IEP meetings, or completing student evaluations. For example, some districts considered on-time IEP completion when measuring SET professionalism.

We coded a document as addressing instruction if it included any language or examples discussing academic or behavioral strategies that SETs should use. This code captured specific interventions (e.g., the Picture Exchange Communication System), intervention frameworks

(e.g., Applied Behavior Analysis, explicit instruction), and broader instructional strategies (e.g., use of accommodations, individualized intervention). We coded a SET evaluation document as addressing student achievement in SETs' evaluation scores when the document included specific information about, for example, weighting test scores in the overall rating, how to link specific students to SETs, or calculating value-added scores or other measures of student learning. Appendix A includes the complete codebook. As we coded, we highlighted the text in each document that corresponded to the content codes. We then organized the text into spreadsheets by code, read all text corresponding to the code, and selected quotes illustrative of how states and districts conceptualized SET evaluation.

Training Procedures and Inter-Scorer Agreement

The first author and a research assistant (RA) coded all documents following a two-part training. First, the first author presented each component of the coding guide (Appendix A) alongside an example and non-example. Second, they each independently coded two sets of materials, one from a state and one from a district, and then met to discuss each code. The initial training lasted six hours. Following training, the first author and RA independently coded all materials for each state and district (i.e., all materials were double-coded), meeting weekly to conduct point-by-point inter-scorer agreement and to discuss discrepancies. Content coding exact agreement averaged 86.19% (SD=13.98%, range= 54.55-100%) and we reached consensus on discrepancies.

Results

As shown in Tables 1 and 2, the majority of states and districts did not provide guidance on SET evaluation in their teacher evaluation systems. We identified documents regarding SET evaluation in 21 states (42%; Table 2) and 8 districts (30.8%; Table 1). In Figure 1, we display

the percentage of states and districts with SET evaluation guidance that addressed the roles of SETs, procedural responsibilities, instruction, and incorporating student achievement into evaluation scores.

Roles

Many states and districts offering guidance provided information related to SETs' roles, 66.7% of states and 62.5% of districts respectively (note that the denominators in these calculations are the states and districts *with* SET guidance; Figure 1). These roles were often defined by where a SET provided instruction (e.g., resource room, life skills classroom) or how the SET was expected to instruct students (e.g., as a co-teacher or through one-on-one or small group instruction).

Many documents emphasized the importance of evaluating SETs in the setting where they most commonly provided instruction and asked evaluators to consider the influence of service delivery model on instruction. Guidance from Chicago Public Schools, for example, recommended that evaluators consider "common planning time, pedagogy and expertise of each teacher, division of roles and responsibilities, parity and compatibility of the teachers, overall workload of each teacher" when evaluating SETs in co-teaching settings. This quote highlights the many factors potentially influencing what an evaluator observed in relation to a SET's role.

State-level guidance also urged evaluators to consider the roles of SETs and adjust evaluation procedures when needed. In acknowledging the many roles that SETs fill, guidance from Colorado stated that "evaluators and special education teachers should consider flexibility in changing or adjusting the weighting of the standards in determining effectiveness, in consideration of that teacher's unique role and population being served." This example illustrates

how guidance acknowledged that the roles of SETs may change how an evaluator should engage with SETs during the evaluation process.

Procedural Responsibilities

Procedural responsibilities such as writing IEPs, leading meetings, coordinating related services, and conducting student evaluations are requirements of SETs' jobs. As shown in Figure 1, about half of states (52.4%) and districts (50%) signaled the importance of these activities within guidance documents. For example, supplemental guidance materials from Cobb County included expectations for developing IEPs and leading meetings. States and districts often focused on timelines during the IEP and re/evaluation or eligibility process. This focus is demonstrated by the requirement from the DC Public Schools that 10% of SETs' final evaluation scores are tied to completing IEPs and eligibility paperwork on-time. Their stated reason for this was that "timely renewal of IEPs is critical to ensuring that our students receive all the services they need. Furthermore, it is required by federal law." These examples highlight the frequency with which states and districts acknowledge procedural responsibilities as an important component of SET quality.

Instructional Considerations

In reviewing guidance documents, we identified an important difference between states and districts. Districts were far more likely to directly address instruction than states (87.5% of districts vs. 52.4% of states; Figure 1). A common form of instructional guidance was to provide evaluators with documents mapping specific instructional actions onto the domains or standards of existing evaluations designed for general education teachers. Broward County, for example, provided three "crosswalks" (i.e., documents that link an observation rubric to another set of information about instruction) that matched the Marzano observation rubric, a commercially

available observation system, when teachers taught in self-contained settings. Similarly, Pennsylvania and Washington relied on FFT's published scenarios for linking FFT domains to different special education settings. These crosswalks and scenarios described teacher behaviors expected in each of these settings. In some cases, guidance documents even identifying specific evidence-based practices, demonstrating an understanding that effective SET instruction might require modifications to the general education observation system.

The crosswalks offered by some states and districts tended to list expected instructional activities, but they did not formally define different levels of proficiency. Three districts, Chicago Public Schools, DC Public Schools, and Hillsborough, provided complete observation rubrics that defined levels of proficiency specific to SETs. These rubrics were described as supplements to the general education teacher rubric but provided significantly more detail on expected instructional practices than the crosswalks provided by other districts. For example, DC provided SET-specific observation rubrics, with defined ratings of specific practices, for SETs who provided specialized instruction generally and SETs who worked with students in specific programs (e.g., Applied Behavior Analysis program).

More common than specific SET practices were general recommendations about instruction. States and districts focused on instruction that was "developmentally appropriate," "focused on students individual needs," or "research-based." This guidance included statements that SETs should use evidence-based practices or instructional approaches, but offered few details about those practices. The vagueness of this guidance did not provide information to SETs about the instruction that they should provide to students and instead left teachers to define appropriate, research-based, or evidence-based practices on their own.

In other cases, states only went so far as to suggest *possible* alterations, providing even less prescribed instructional guidance. Guidance from Illinois and Minnesota suggested that evaluators receive special training in evaluating SETs, that evaluators use pre-observation conferences to discuss the type of instruction that SETs may use, and that districts should consider relying on evaluators with expertise in special education. Louisiana, Pennsylvania, and Tennessee also highlighted the roles of evaluators by providing evaluators with questions to ask SETs working with SWDs before they conduct classroom observations. These examples illustrate the varied methods that districts, and sometimes states, used to address SETs' use of instruction that differed from that of general educators.

Incorporating Student Achievement

The most common guidance offered by states (81%) surrounded recommendations or rules for incorporating student outcomes into SETs' evaluation scores; illustrated in the top panel of Figure 1. However, shown in the bottom panel of Figure 1, only three districts (37.5%) included guidance on incorporating student achievement into SETs' evaluation scores. The district guidance varied across the three districts. The School District of Philadelphia, for example, provided guidance on the number of students who should share a learning objective, while guidance from the DC focused on the rigor of methods for evaluating student achievement.

States that mentioned the use of IEP goals for measuring student achievement offered contradictory guidance. Four states explicitly stated that IEP goals could be used as student learning objectives for measuring teaching effectiveness. In contrast, five states explicitly prohibited the use of IEP goal progress as a measure of student achievement within SETs' evaluation scores. For example, guidance from Maine stated: "Using student progress on IEP goals may compromise the integrity of the IEP, shifting its focus […] to the performance of the

teacher." Taken together, the guidance on integrating student learning into SETs' evaluation did not present a clear picture about best practices.

Discussion

Teacher evaluation is a mechanism for improving teaching quality. States and districts adopt standard systems with single faceted views of teaching for important reasons, including scale, comparability, and cost. But, as the case of SETs illustrates, one type of instruction is not effective for all learners. Additionally, SETs fulfill a wide variety of roles, have non-instructional responsibilities, provide instruction that may not be captured or addressed in evaluation systems, and teach students for whom it may be difficult to measure growth, thus limiting the utility of these systems for promoting SET quality (Jones & Brownell, 2014). This investigation highlighted how states and districts conceptualized deviations from traditional teachers by examining the guidance that states and districts provided to implementers on how to evaluate SETs.

Overall, our analysis suggests two major findings. First, with less than half of states and less than half of districts providing guidance regarding SET evaluation, states and districts may not have recognized the limitations of a single definition of effective teaching. Second, substantive differences existed across state and district guidance highlighting the varied priorities of these stakeholders. We discuss the implications of these findings for developing policies that can support teaching quality for all educators.

Standardization vs. Flexibility

There is a tension between standardization and flexibility within evaluation systems, and policies that address instruction more generally. This tension may underlie the finding that most states and districts did not provide guidance for evaluating SETs. Standardization in teacher

evaluation measures helps to ensure that scores are comparable across teachers and that, at face value, systems are fair. However, standardization ignores the varied roles teachers fill, their specific responsibilities, and that effective instruction may look different for students with different needs. A truly fair system must acknowledge these differences across teachers. Flexibility within evaluation systems allows for adjustments to match specific contexts. This could increase the likelihood that evaluations are able to support teacher improvement, while increasing cost and sacrificing comparability.

Notably, the tension between flexibility and standardization in evaluation systems, particularly the observation component, is not unique to SETs or school personnel who fulfill specialized roles in schools. States and districts have elected to adopt evaluation systems that include different observation protocols that likely define effective instruction in different ways (Steinberg & Donaldson, 2016). Others have noted that subject agnostic observation systems may not result in improvements to teachers' subject specific pedagogy, or pedagogical content knowledge for teaching, required to promote student achievement (Hill & Grossman, 2013). The need for flexible systems that are relevant and useful to all teachers extends beyond the case of SETs.

Two issues addressed by the guidance we identified illustrated the tension between flexibility and standardization: guidance regarding the use of IEP goals to assess teacher effectiveness and guidance on who evaluates SETs. Incorporating student learning to assess SET effectiveness presents a multitude of challenges (Jones et al., 2013), yet most evaluation systems require that teachers are partially evaluated on the learning of their students (Steinberg & Donaldson, 2016). How student learning is measured has some flexibility because test scores are unavailable for many teachers, not just SETs. In this study, four states recommended that schools

should use IEP goals to measure student learning and five states prohibited the use of IEP goals to measure student learning. While these varied approaches highlight flexibility, in the case of using IEP goals to assess teachers, standardization across states might be appropriate. IEP goals are intended to direct the educational services that SWDs receive and assess if students are receiving the appropriate instruction, accommodations, or modifications that they need to make educational progress. Linking IEP goals to teachers' outcomes risks making these IEP goals less useful for students (Goodhart, 1975). We caution against the use of IEP goals to measure teacher quality as the risk of contaminating these goals that establish the backbone of educational services provided to students with disabilities is too high.

Altering standardized evaluation systems to flexibly address the needs of specific teachers requires expertise. Some states and districts included recommendations for the use of pre-observation conferences between the evaluator and the SET and choosing evaluators familiar with special education. Some also provided suggested questions and topics for SETs and evaluators to review in order to define the instructional role of the SET and to identify the specific needs of SETs' students. The strategic use of pre-observation conferencing presents an opportunity to SETs and principals to make observations more responsive (i.e., flexible) to SETs' positions, but creates implementation challenges and could decrease standardization of the protocol.

Pre-observation conferences are promoted or required in many evaluation systems (Steinberg & Donaldson, 2016), but there is evidence that they may not take place due to time constraints (Kraft & Gilmour, 2016). Additionally, even if school leaders use pre-observation conferences to customize evaluation systems, evaluators without special education experience struggle with evaluating SETs, primarily SETs' use of instructional practices aligned with

effective instruction for SWDs (Lawson & Cruz, 2018a; 2018b). Flexibility resulting in effective and appropriate modifications require a knowledgeable evaluator, a knowledgeable SET, and a cooperative relationship between the evaluator and the SET.

Outcomes vs. Instruction

We identified substantial variation across the guidance provided by states and districts regarding student outcomes and instruction. All but four of the states providing SET evaluation guidance included recommendations or requirements for incorporating student achievement into SETs' evaluation scores. The importance placed on this component of teacher evaluation likely reflects the political nature of these measures and the methodological concerns regarding measuring SWDs' academic growth (Buzick & Jones, 2015). However, the focus on the technical aspect of measuring SETs' quality using student outcomes at the state-level over providing guidance regarding instruction may also implicitly suggest that a single evaluation observation rubric can capture effective teaching, that is, good teaching looks the same for all learners.

In contrast, districts that addressed SETs focused on modifying existing observation instruments, or in some cases creating entirely new rubrics, to capture what they defined as "good teaching" for SWDs. Districts may be aware of the nuances of effective instruction and understand that high-quality instruction may look different for students with varying needs. Or, it could be that districts are providing guidance to supplement state policies and recommendations, recognizing that attending to student achievement alone is insufficient for addressing the instructional needs of SETs.

Limitations

The results of this investigation should be considered in light of several limitations. We were unable to confirm the accuracy of published materials in 23% of districts and 13% of states. We may have not located all available guidance in these states and districts, but if this information is not readily available online it may limit the extent to which stakeholders access the information. We did not specifically investigate if districts were more likely to provide guidance on topics when their state did not offer any guidance. This compensatory approach could also apply to how evaluators at individual schools address the challenges of evaluating SETs when districts do not provide guidance. We did not evaluate if the specific guidance aligned with preferred practices in special education, but instead examined the extent to which states and districts in our sample considered the elements of being a SET that could influence the use of evaluation systems for improving teaching quality. This limits the specific recommendations that we can make for adapting or adopting evaluation systems that address specific groups of teachers. Finally, we only examined the country's largest districts. Large districts may have more resources, or greater need, for developing SET evaluation materials; thus this sample may not represent the information provided to evaluators regarding SETs in smaller districts.

Recommendations and Conclusions

The results suggest that although some state and district personnel are attending to the challenges of evaluating SETs, the majority are not. The lack of guidance offered by many is problematic for defining and supporting effective instruction for SWDs and holding SETs accountable for delivering effective instruction and fulfilling their noninstructional responsibilities. More broadly, it also suggests that policymakers may not understand that effective instruction is not the same for all learners and that improving teaching quality cannot

focus on one definition of "good teaching." Widely-used observation systems written with general education teachers in mind might penalize SETs and other school personnel for using instructional practices uncaptured in these instruments (Jones & Brownell, 2014), raising questions of fairness and whether scores lead to inaccurate personnel decisions. Worse, the use of instruments that privilege a single view of "good teaching" might discourage SETs from adopting practices known to help SWDs, but are not included in the rubric, resulting in poor student outcomes.

The larger issue related to defining effective instruction warrants the field's attention. Teacher evaluation policies are not the first in the history of education to attempt to shape teaching and learning (see the incentivizing of adoption of common standards under Race to the Top [U.S. Department of Education, 2009] and as a required component of ESEA flexibility [U.S. Department of Education, 2012] as two recent examples). Reforms related to teaching and learning reveal our assumptions about how classrooms are structured and what constitutes high-quality instruction. Our lingering concern is that when policies fail to account for critical subgroups of teachers and their students, it only reinforces longstanding divides between general and special education and prevents the adoption of effective instruction that may not look the same for all learners.

References

Bell, C., Jones, N., Qi, Y., & Lewis, J. (2018) Strategies for assessing classroom teaching: Examining administrator thinking as validity evidence, *Educational Assessment, 23,* 229-249, doi:10.1080/10627197.2018.1513788

Billingsley, B., & Bettini, E. (2020). Special education teacher attrition and retention: A review of the literature." *Review of Educational Research*. Advance online publication. doi:10.3102/0034654319862495.

Brownell, M. T., & Jones, N. D. (2015). Teacher evaluation in special education: Approaches, supporting research, and challenges confronted. *Journal of Special Education Leadership, 28*, 63–73.

Buzick, H. M., & Jones, N. D. (2015). Using test scores from students with disabilities in teacher evaluation. *Educational Measurement: Issues and Practice*, *34*, 28–38. doi:10.1111/emip.12076

Cohen, J. (2018). Practices that cross disciplines?: Revisiting explicit instruction in elementary mathematics and English language arts. *Teaching and Teacher Education: An International Journal of Research and Studies*, *69*(1), 324-335.

Council for Exceptional Children. (2012). *The Council for Exceptional Children's position on special education teacher evaluation.* Arlington, VA: Council for Exceptional Children

Dee, T. S., & Goldhaber, D. (2017). *Understanding and addressing teacher shortages in the United States.* Washington, DC: Brookings. Retrieved from https://www.brookings.edu/research/understanding-and-addressing-teacher-shortages-in-the-united-states/

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from

IMPACT. *Journal of Policy Analysis and Management*, *34*, 267-297. doi:10.1002/pam.21818

Donaldson, M. L., & Papay, J. P. (2014). Teacher evaluation for accountability and development. In H. F. Ladd & M. E. Goertz (Eds.), *Handbook of research in education finance and policy* (2nd ed., pp. 190-209). Routledge.

Donaldson, M. L., & Woulfin, S. (2018). From tinkering to going "rogue": How principals use agency when enacting new teacher evaluation systems. *Educational Evaluation and Policy Analysis, 40*, 531-556. doi:10.3102/0162373718784205

Gilmour, A. F., Fuchs, D., & Wehby, J. H. (2019). Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities. *Exceptional Children, 85*, 329-346. doi:10.1177/0014402918795830

Goodhart, C. (1975) Problems of monetary management: The UK experience. In in *Papers in Monetary Economics, Volume I*, Reserve Bank of Australia, 1975.

Grissom, J. A., & Bartanen, B. (2019). Principal effectiveness and principal turnover. *Education Finance and Policy, 14*, 355-382. doi:10.1162/edfp_a_00256

Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*, 371-384. doi:10.17763/haer.83.2.d11511403715u376

Hock, H., & Isenberg, E. (2012). Methods for accounting for co-teaching in value-added models. *Statistics and Public Policy, 4*, 1-11. doi:10.1080/2330443X.2016.1265473

Holdheide, L., Browder, D., Warren, S., Buzick, H., & Jones, N. (2012). *Using student growth to evaluate educators of students with disabilities: Issues, challenges, and next steps*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved

from https://gtlcenter.org/sites/default/files/38_1_25_2013_IL%20SWD%20growth_%20TeacherEvaluation_Illinois_0.pdf

Holdheide, L. R., Goe, L., Croft, A., & Reschly, D. J. (2010). *Challenges in evaluating special education teachers and English language learner specialists*. National Comprehensive Center for Teacher Quality. Retrieved from https://eric.ed.gov/?id=ED520726

Johnson, E. S., & 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. doi:10.1177/1534508413513315

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*, 112–124. doi:10.1177/1534508413514103

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

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment* (MET Project). Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from https://eric.ed.gov/?id=ED539299

Kane, T. J., Staiger, D. O., & McCaffrey, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Bill and Melinda Gates Foundation. Retrieved from https://eric.ed.gov/?id=ED540960

Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development as

evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly, 52(5)*, 711–753. doi:10.1177/ 0013161X16653445

Kraft, M. A., & Gilmour, A. F. (2017). Revisiting The Widget Effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, *46*, 234–249. doi:10.3102/0013189X17718797

Lawson, J. E., & Cruz, R. A. (2018a). Evaluating special education teachers' classroom performance: Rater reliability at the rubric item level. *Teacher Education and Special Education, 41*, 263-276. doi:10.1177/0888406417718250

Lawson, J. E., & Cruz, R. A. (2018b). Evaluating special educators' classroom performance: Does rater "type" matter? *Assessment for Effective Intervention*, *43*, 227-240. doi:10.1177/1534508417736260

Mathews, H., Strark, K., Jones, N. D., Brownell, M. T., & Bell, C. B. (2020). Danielson's Framework for Teaching: Convergence and divergence with conceptions of effectiveness in special education. Working paper.

Papay, J. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review, 82*(1), 123-141. doi:10.17763/haer.82.1.v40p0833345w6384

Sledge, A., & Pazey, B. L. (2013). Measuring teacher effectiveness through meaningful evaluation: Can reform models apply to general education and special education teachers?. *Teacher Education and Special Education, 36*, 231-246. doiL10.1177/0888406413489839

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, 340-359. doi:10.1162/EDFP_a_00186

Steinbrecher, T. D., Selig, J. P., Cosbey, J., & Thorstensen, B. I. (2014). Evaluating special educator effectiveness: Addressing issues inherent to value-added modeling. *Exceptional Children*, *80*, 323–336. doi:10.1177/0014402914522425

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628-51. doi:10.1257/aer.102.7.3628

U.S. Department of Education. (2009). *Race to the Top program executive summary*. Washington, DC: Author. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education. (2012). *ESEA flexibility*. Washington, DC: Author. Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

Vannest, K. J., Hagan-Burke, S., Parker, R. I., Roares, D. A. (2011). Special education teacher time use in four types of programs. *The Journal of Educational Research, 104*(4), 219-230.

Table 1

Districts included and topics covered documents related to SET evaluation

	Materials	Role of SET	Procedural	Instructional	Outcomes
Baltimore County	None				
Broward	Guidance	X	X	X	
Charlotte-Mecklenburg	None				
Chicago	Rubric	X	X	X	
Clark County+	None				
Cobb County	Guidance	X	X	X	
Cypress-Fairbanks	None				
Dallas+	None				
Dade	None				
District of Columbia	Rubric	X	X	X	X
Duval	None				
Fairfax County	None				
Gwinnett County	None				
Hillsborough	Rubric			X	X
Houston	None				
Los Angeles Unified	None				
Montgomery County+	None				
New York City+	None				
Orange	None				
Palm Beach	None				
Philadelphia+	Guidance				X
Pinellas	Guidance			X	
Prince George's County+	None				
San Diego Unified+	None				
Shelby County	Guidance	X		X	
Wake County	None				

Note. Districts marked with a plus (+) did not respond to requests from the researchers to confirm the identification of materials related to evaluating SETs. SET= Special education teachers;

Table 2
States included and topics covered documents related to SET evaluation

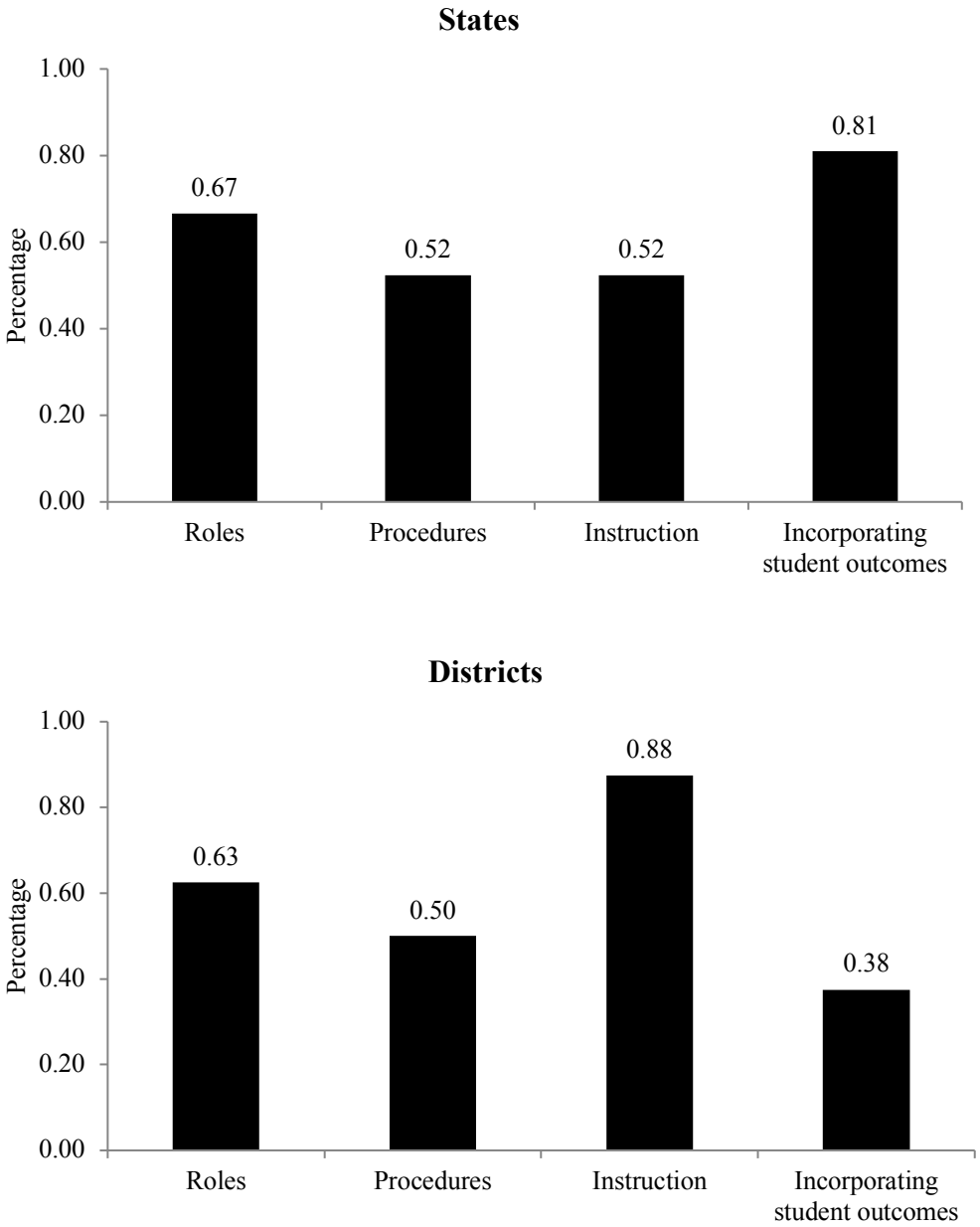
	Materials	Role of SET	Procedural	Instructional	Outcomes
Alabama[+]	None				
Alaska[+]	Guidance				X
Arizona	Guidance				X
Arkansas	Guidance		X	X	
California[+]	None				
Colorado	Guidance	X	X	X	X
Connecticut	None				
Delaware	None				
Florida	None				
Georgia[+]	None				
Hawaii	Guidance				X
Idaho	None				
Illinois	Guidance	X	X	X	X
Indiana	Guidance				X
Iowa	None				
Kansas	None				
Kentucky	Guidance	X	X	X	
Louisiana	Guidance		X	X	X
Maine[+]	Guidance	X			X
Maryland	None				
Massachusetts	Guidance	X			X
Michigan	Guidance				X
Minnesota	Guidance	X	X	X	X
Mississippi	None				
Missouri	None				
Montana	None[+]				
Nebraska	None				
Nevada[+]	None				
New Hampshire	None				

New Jersey+	Guidance	X		X	X
New Mexico+	Guidance	X	X	X	
New York	Guidance	X			X
North Carolina+	None				
North Dakota	None				
Ohio	Guidance	X			X
Oklahoma	None				
Oregon+	None				
Pennsylvania	Guidance	X	X	X	X
Rhode Island	None				
South Carolina	None				
South Dakota	Guidance	X	X		X
Tennessee	Guidance	X	X	X	X
Texas	None				
Utah+	None				
Vermont	None				
Virginia	None				
Washington	Guidance	X	X	X	
West Virginia	None				
Wisconsin+	None				
Wyoming+	None				

Note. States marked with a plus (+) did not respond to requests from the researchers to confirm the identification of materials related to evaluating SETs. SET= Special education teachers

Figure 1

Content of Special Education Teacher Evaluation Guidance Across States with Guidance (42%) and Districts with Guidance (31%)



Note. The bars represent the proportion of states or districts with guidance that included guidance on the specific topic.

Appendix A
Codebook

Question	Coding	Additional Notes
1. Does the district/state provide guidance regarding SET evaluation?	**1** = District/state provides any of the following: SET rubric Guidance document Presentation or training regarding SET evaluation SET "crosswalks" (adaptations to the evaluation system specific to SETs or SWDs) SET scenarios SET brief SET related FAQs *Note* that information related to general education teachers including SWDs but does not address evaluating SETs is coded as "0" **0** = District/state does not offer additional guidance on SET evaluation. District/state uses the same evaluation system for all teachers.	If Q1=0, do not code remaining variables for the district/state.
2. Does the district/state provide a separate observation rubric for evaluating SETs?	**1** = District/state provides a separate observation rubric for SETs **0** = All teachers are evaluated using the same rubric	A district/state might have multiple documents and be coded as "1" for multiple items (Q2-Q5)
3. Does the district/state provide a guidance document for evaluating SETs?	**1** = The district/state has released a memo, brief, FAQ form, or other written document that provides information about evaluating SETs. **0** = Code "0" if the information provided to evaluators is a separate evaluation rubric, training (including a presentation), or a document that details special education scenarios or adaptations to the generally used evaluation rubric	
4. Does the district/state provide additional training to	**1**= The district/state stated that they provide separate training to evaluators regarding SET evaluation OR the district/state provides a presentation (slides) regarding SET evaluation for evaluators	

evaluators for evaluating SETs?	**0**= The district/state does not provide additional training or training documents related to SET evaluation	
5. Does the district/state provide descriptive information about adapting the general teacher evaluation rubric for SETs?	**1**= The district/state provides guidance such as documents that link the teacher evaluation system to SET teaching standards from state, district, or professional groups (often called "crosswalks") OR the district/state provides examples of how the rubric could be adapted to address situations specific to SETs (these might be scenarios or stories) If districts have created a rubric specific to SETs, Q5 is coded as "0" and Q2 is coded as "1" **0**= The district/state does not provide additional documents with descriptive information about adapting the existing evaluation system	
6. Do the documents address the specific roles of SETs?	**1**= The district/state documents include a discussion of: SETs in a consultant role (do not have their own classroom) Co-teaching Pull-out or small group instruction **0**= Documents do not address roles of SETs	A district/state might have multiple documents and be coded as "1" for multiple items (Q6-Q9). Code the content of each document.
7. Do the documents address procedural activities?	**1**= The district/state documents include a discussion of: Writing IEPs Tracking IEP progress Organizing/attending IEP meetings Completing evaluations/re-evaluations **0**= Documents do not address SETs' procedural activities	
8. Do the documents address instructional considerations?	**1**= The district/state documents include a discussion of specific academic/behavioral strategies or interventions used by SETs **0**= The district/state does not address SETs' instruction	

9. Do the documents address measuring student achievement or incorporating student achievement into teachers' evaluations?	**1**= The district/state documents include information related to measuring student achievement including: Writing student learning objectives Including student growth in evaluation score (weighting of test scores, which students count, whether disability-specific covariates are included in student growth measures, provides guidance around shared instructional responsibilities for SWDs) Calculating value-added or other quantitative measures of student learning **0**= The district/state does not address incorporating student achievement into SETs' scores	
10. Do the documents address teaching students with specific disabilities?	**1**= The district/state addresses specific disability categories (e.g., intellectual disabilities, behavior disorders) **0**= The district/state addresses SWDs as one group	
11. Which disabilities do the documents address?	**0**= The district/state addresses SWDs as one group. Code "0" if Q10=0 1= Learning disability 2= Other health impairment 3= Autism spectrum disorder 4= Emotional disturbance/behavior disorder 5= Sensory impairment (deafness, hearing impairment, visual impairment) 6=Speech/language impairment 7= Intellectual disability 8= Other (Orthopedic impairment, traumatic brain injury, multiple disabilities)	Code all disability categories addressed For example: 1, 3, 5

View publication stats