The Coverage of Classroom Management in Teacher Evaluation Rubrics

Allison F. Gilmour[1], Caitlyn E. Majeika[2], Amanda W. Sheaffer[2], and Joseph H. Wehby[2]

Abstract

Many teachers struggle with the classroom management skills necessary to support students with disabilities. Teacher evaluation is a potential method for supporting teachers' use of classroom management strategies. The authors evaluate the extent to which classroom management was included in state teacher evaluation rubrics and the types of classroom management topics that were included in the rubrics. They find that while one fifth of standards and indicators were related to classroom management, there was large variability across states. Rubrics that included classroom management focused on proactive strategies, but often did not include specific evidence-based practices that could help teachers improve their classroom management skills. Very few rubrics included any focus on how to address student misbehavior. In light of these findings, the authors provide a discussion of the importance of including high-quality standards and indicators about classroom management to support teachers of students with disabilities.

Keywords

classroom management, teacher evaluation, inclusion

Throughout the country, teachers report that they are unprepared to handle behavior problems in their classrooms (National Center for Education Statistics [NCES], 2015), that student misbehavior interferes with their teaching (NCES, 2015), and that they are not accepting of having students who exhibit problem behavior in their classrooms (Martin, Lloyd, Kauffman, & Coyne, 1995; Olson, Chalmers, & Hoover, 1997; Soodak, Podell, & Lehman, 1998; Westling, 2010). Teacher preparation programs rarely include training on classroom management (Freeman, Simonsen, Briere, & MacSuga-Gage, 2014; Greenberg, Putman, & Walsh, 2014; Oliver & Reschly, 2010; Reiff, Evans, & Cass, 1991), likely exacerbating or causing these challenges. The lack of teacher training in classroom management is particularly worrisome for teachers of students with disabilities (SWDs) as these students may be at greater risk for exhibit-

ing problem behavior than their peers without disabilities (Blackorby et al., 2005; U.S. Department of Education, 2006).

One proposed method for improving in-service teacher skills is through systematic teacher evaluation. Systematic teacher evaluation aims to improve teaching by identifying areas in need of improvement and providing teachers with additional supports (Papay, 2012). Recent federal initiatives, Race to the Top (U.S. Department of Education, 2009) and Elementary and Secondary Education Act

[1]Temple University, Philadelphia, PA, USA
[2]Vanderbilt University, Nashville, TN, USA

Corresponding Author:
Allison F. Gilmour, Temple University, 1301 Cecil B. Moore Ave., Ritter Hall 362, Philadelphia, PA 19122, USA.
Email: allison.gilmour@temple.edu

flexibility waivers (U.S. Department of Education, 2012), incentivized states to develop and implement teacher evaluation systems and evidence suggests that evaluation systems may lead to improvements in student outcomes (Dee & Wyckoff, 2015; Steinberg & Sartain, 2015; Taylor & Tyler, 2012). It is, however, not clear whether teacher evaluation tools incorporate the assessment of classroom management skills, skills integral for supporting SWDs in general and special education classrooms.

In this study, we evaluated the extent to which classroom management practices were included in state teacher evaluation rubrics and the classroom management topics addressed by evaluation rubrics. The exclusion of classroom management from teacher evaluation rubrics could be problematic for teachers who have SWDs in their classrooms and who may need more support and training in implementing effective classroom management programs. We begin the literature review by describing effective classroom management practices identified in research. We then review the existing research on teacher evaluation.

Literature Review

Effective Classroom Management

Though teachers report that they are unprepared to manage classroom behavior and discipline for difficult students (NCES, 2015), a large body of research has identified effective classroom management interventions (Cooper & Scott, 2017; Mitchell, Hirn, & Lewis, 2017; Simonsen, Fairbanks, Briesch, Myers, & Sugai, 2008). In one review of the literature on classroom management practices, Simonsen and colleagues (2008) noted five overarching classroom domains supported by research: (a) maximize structure and predictability; (b) post, teach, review, monitor, and reinforce expectations; (c) actively engage students in observable ways; (d) use a continuum of strategies to acknowledge appropriate behavior; and (e) use a continuum of strategies to respond to inappropriate behavior. Within these domains, strate-

gies such as active supervision (De Pry & Sugai, 2002), increased opportunities to respond (Moore-Partin, Robertson, Maggin, Oliver, & Wehby, 2010; Sutherland & Wehby, 2001b), and praise (Sutherland & Wehby, 2001a) are examples of classroom management practices that have demonstrated positive impact on student behavior. Teachers may report that they are underprepared in classroom management, but effective classroom management practices are well documented in research.

Additional evidence suggests that strong classroom management is associated with student academic and behavioral outcomes (Brownell et al., 2009; Garwood, Vernon-Feagans, & Family Life Project Key Investigators, 2017). For example, Brownell et al. (2009) found that beginning special education teachers with higher ratings on classroom management strategies had students who made greater gains in reading. The use of classroom management practices can positively impact student behavior by decreasing problem behaviors (Lambert, Cartledge, Heward, & Lo, 2006; Stormont, Smith, & Lewis, 2007), increasing on-task behavior (Williamson, Campbell-Whatley, & Lo, 2009), and increasing positive verbal interactions (Hansen & Lignugaris-Kraft, 2005). Considering the poor academic and social outcomes for students who display problem behaviors, strong classroom management practices are essential for successful student outcomes. Supporting teacher development in classroom management is especially important as a significant portion of SWDs receive their academic instruction in general education classrooms (U.S. Department of Education, 2016). General education teachers, however, may have less knowledge of behavior than special education teachers (King-Sears, Carran, Dammann, & Arter, 2012).

Teacher Evaluation as a Tool for Improvement

Providing feedback via formal evaluation procedures may be one method for improving teacher classroom management skills and supporting the behaviors of SWDs.

Evaluation can change teacher practices by helping teachers learn new skills or by increasing teacher awareness of desired instructional practices (Taylor & Tyler, 2012). Teacher evaluations typically include multiple components including (a) results of classroom observations, (b) quantitative measures of student outcomes such as value-added or student learning objectives, and/or (c) surveys of students and parents (Steinberg & Donaldson, 2016). In this article, we focused on the observation component of evaluation. Within these systems, teachers are evaluated and provided with feedback based on what is seen/measured by the individual conducting the observation (Steinberg & Donaldson, 2016). Feedback from the observation component of evaluation may apply to both academic instructional practices and classroom management skills.

Research has examined if the use of systematic teacher evaluation systems results in improved student academic outcomes. For example, Taylor and Tyler (2012) evaluated the impact of a teacher evaluation system with observations by peer evaluators and administrators using the *Framework for Teaching* (*FFT*; Danielson, 1996) on the value-added scores of midcareer teachers. Teachers in this study had an average growth in math value-added of 0.12 standard deviations after the year they were evaluated. The results of this study and others (Dee & Wyckoff, 2015; Steinberg & Sartain, 2015) suggest that systematic evaluation that includes observations may be a mechanism for improving teacher practice; teachers may change their instruction based on the evaluation process.

However, not all teacher evaluation systems will result in improved teacher performance. Successful teacher evaluation partially depends on the quality of the observation rubrics used during the evaluation process. High-quality rubrics can inform teaching and professional development (Pianta & Hamre, 2009) and may correlate with teacher value-added scores (Grossman, Cohen, Ronfeldt, & Brown, 2014; Grossman, Loeb, Cohen, & Wyckoff, 2013; Hill, Kapitula, & Umland, 2011). In addition, observation rubrics may capture teaching practices valued by schools that are not entirely captured in value-added scores (Grossman et al., 2014) and provide teachers with information about expected and encouraged teacher practices that are reinforced by administrators during the evaluation process (Pianta & Hamre, 2009).

Teacher Evaluation and Special Education

As compared with general education practices, research related to evaluation as a mechanism for improving the outcomes of SWDs or supporting special education teachers is scarce. Primarily, studies have focused on the inclusion of SWDs in value-added measures (Buzick & Jones, 2015; Jones, Buzick, & Turkan, 2013; Steinbrecher, Selig, Cosbey, & Thorstensen, 2014) or the appropriateness of evaluation systems in capturing the role of special educators (Jones & Brownell, 2014; Sledge & Pazey, 2013). In addition, most states do not provide separate evaluation rubrics for special education teachers, and the information that states do provide to districts regarding special education teacher evaluation primarily focuses on how to incorporate student growth into evaluation scores for special education teachers (Jones & Gilmour, in press).

Some researchers have recommended the development of a separate tool for evaluating special education teachers (e.g., Holdheide, Goe, Croft, & Reschly, 2010; Jones & Gilmour, in press; Semmelroth & Johnson, 2014; Sledge & Pazey, 2013). Many SWDs, however, receive the majority of their instruction in general education classrooms (U.S. Department of Education, 2016). As such, an evaluation rubric explicitly addressing components of special education teaching effectiveness in special education settings may not provide enough information or guidance to teachers struggling to support the behavior of these students in a general education setting. Teachers may benefit from a comprehensive evaluation rubric that incorporates classroom management strategies to support all learners, especially those with disabilities.

Study Purpose

Teacher evaluation is one proposed tool for developing and sustaining effective teaching practices. The evaluation process, however, may not support teacher growth in classroom management if rubrics do not include these skills as measured components. The exclusion of classroom management on teacher evaluation rubrics could be particularly problematic for teachers who have SWDs in their classrooms. The objectives of this study were (a) to examine the extent to which classroom management items are included as standards and indicators in each state's teacher evaluation and (b) to examine the content of the classroom management items.

Method

Sample

Identifying state rubrics for inclusion was a multistep process. First, we looked at the data on each state's teacher evaluation system available from the American Institutes for Research (2015) database and the National Council on Teacher Quality's State of the States report (Doherty & Jacobs, 2015). We included the District of Columbia (DC) and referred to DC as a state throughout this article for ease of discussion. Through these databases, we identified states that had required or model rubrics. Second, we compared these data to information available on state websites about specific requirements for evaluation rubrics.

Instead of evaluating all possible rubrics at the state and district level, we coded only state-developed required and model rubrics. We included state rubrics based on published evaluation tools if they had been altered by the state. We included states that created rubrics based on professional standards in the state. We downloaded the most current rubric as of January 15, 2017, that we could access from state department of education websites and used these publicly available documents in our analyses. Several states used commercially published evaluation rubrics, such as

Danielson's *FFT*, or widely used standards, such as *InTASC* standards, verbatim. We coded the *FFT* and *InTASC* instruments separately, rather than evaluate the same tool multiple times by state. We did not separately code state rubrics that were taken verbatim from published instruments. Because many states use the *InTASC* standards or the *FFT* as their main evaluation system, we also coded the *FFT* and *InTASC* standards and include them in our sample. Our final sample consisted of 30 rubrics; 28 states, the *InTASC* standards, and the *FFT* (listed in Table 1). A list of states and reasons for exclusion are included in the appendix.

Rubric Coding

We coded evaluation rubrics for (a) basic descriptive information, (b) the inclusion of classroom management components, and (c) the content of the classroom management components. We recorded if it was stated somewhere in the rubric or accompanying documentation that the evaluation was based on an existing published rubric (e.g., *FFT*) or set of teaching standards, but did not use the instrument or standards verbatim. Observation rubrics typically include two parts that we referred to as *standards* and *indicators*. Standards address a larger component of teaching, for example, the section titled *Planning and Preparation* on the *FFT*. Indicators provide more information about the actions that make up the larger component. Indicators are focused on specificity while standards are general constructs. For example, an indicator under *Planning and Preparation* on the *FFT* is *Setting Instructional Outcomes*. We counted the number of standards, the largest level of evaluation, and then counted the number of indicators, the smaller level of evaluation nested under the standards.

Next, we recorded if the standard or indicator included any component of classroom management. We used an iterative process to develop our final list of codes based on the categories of classroom management reported by Simonsen et al. (2008) and key words identified in the documents (Miles & Huberman,

Table 1. Summary of Standards and Indicators by State.

State	Source	No. of standards	No. of standards about CM	% of standards about CM	No. of indicators	No. of indicators about CM	% of indicators about CM
AL	FFT	5	1	20	39	5	12.82
AK	NR	8	1	12.5	30	1	3.33
CA	NR	6	2	33.33	38	9	23.68
CO	NR	5	2	40	27	4	14.82
CT	NR	4	1	25	12	3	25.00
DE	FFT	4	1	25	18	6	33.33
DC	NR	9	1	1.11	30	4	13.33
GA	NR	10	2	20	71	9	12.68
IN	NR	3	1	33.33	19	4	21.05
IA	NR	8	1	12.5	42	8	19.05
KS	InTASC	4	1	25	10	2	20.00
LA	FFT	3	1	33.33	5	3	60.00
MA	NR	4	1	25	16	5	31.25
MN	NR	4	1	25	34	7	20.59
MS	NR	5	1	20	20	7	35.00
MO	FFT	9	1	11.11	36	6	16.67
MT	FFT	4	1	25	19	5	26.32
NE	FFT, InTASC	7	1	14.29	44	4	9.09
NV	NR	10	0	0	34	1	2.94
NM	NR	4	1	25	22	8	36.36
NC	NR	5	0	0	25	3	12.00
OH	FFT, other	3	0	0	10	2	20.00
RI	FFT	2	1	50	8	4	50.00
SC	NR	10	2	20	34	7	20.59
TN	NR	3	1	33.33	19	5	26.32
TX	NR	4	1	25	16	5	31.25
UT	NR	10	1	10	21	4	19.05
WV	NR	7	1	14.29	16	3	18.75
InTASC		10	1	10	75	1	1.33
FFT		4	1	25	22	6	27.27

Note. CM = classroom management; FFT = Framework for Teaching; NR = not reported.

1994). The final list of topics and key words are listed in Table 2. We coded the standard or indicator as related to classroom management if it included a key word or synonym.

Most rubrics included descriptions of teacher practice at each performance level (e.g., Proficient, Exceptional) for the standard or indicator. If the description of a performance level included any terms we considered to be related to classroom management, the indicator or standard was coded as related to classroom management. All standards and indicators were coded separately. If a standard was about classroom management, it did not necessarily mean that all the indicators under that standard were related to classroom management. Conversely, if a standard was not about classroom management, the indicators nested under the standard could be coded as assessing classroom management.

Finally, we coded the classroom management topic of each indicator that was about classroom management. We did not code the content of standards because they addressed general skills. In contrast, the indicators provided specific information about measured

Table 2. Terms Coded as Related to Classroom Management Based on Simonsen, Fairbanks, Briesch, Myers, and Sugai (2008).

Structure and predictability	Expectations	Engagement	Acknowledgment of behavior	Response to behavior	Other
Classroom management	Student behavior	Supervises students	Positive behavior supports	System to address misbehavior	Learning environment
Routines	Rules	Actively engages students	Reinforcement system	Differential reinforcement	Classroom climate
Transitions	Positively stated expectations	Opportunities to respond	Specific and contingent praise	Planned ignoring	Classroom environment
Classroom structure	Posted expectations		Group contingencies	Response cost	
Classroom arrangement	Prompts or pre-corrects expectations		Behavior contracts		
			Token economies		

Note. We also coded the standard or indicator as related to classroom management if it included a synonym for any of these terms.

teacher behavior. First, we divided topics into four categories based on Simonsen et al. (2008): structure and predictability, expectations, engagement, acknowledgment of behavior, and response to behavior. Then, we added two categories, general and learning environment. The general category captured indicators that simply mentioned classroom management or student behavior but none of the four topics based on Simonsen et al. (2008). We coded indicators' topics as learning environment if they included general statements about the learning environment or classroom climate but none of the four topics based on Simonsen et al. (2008). Table 3 includes the general topics and specific content included in each topic.

Training Procedures

The first author trained the other authors in the coding procedures. During a 1-hour training meeting, the first author presented the components categorized as classroom management (Table 2) and rubric content (Table 3) before modeling coding a rubric. Then, the other authors coded a training rubric until each met a criterion of 90% exact agreement with the first author before independently coding the remaining evaluation rubrics. Each rubric was independently coded by two coders. The team met as a group to resolve any coding discrepancies. The interrater agreement for standards, using exact agreement, averaged 97.50 (SD = 7.39, range = 75-100).

The interrater agreement for indicators averaged 92.34% (SD = 6.26, range = 75-100). The interrater agreement for content coding was 89.34%.

Analysis

Because of the exploratory nature of our research, we used basic descriptive statistics to address our primary research question. For each state, we calculated the percentage of standards related to classroom management. Across the entire sample, we calculated the average percentage of standards related to classroom management and examined the range. We followed the same approach for indicators. We then examined the percentage of classroom management standards and indicators that included information about each classroom management topic. Examining the average percentage of standards and indicators related to classroom management provided us with information about the extent to which classroom management was included in evaluation rubrics and the classroom management content addressed in these rubrics.

Results

Classroom Management Standards and Indicators

Standards. We evaluated the extent to which classroom management was included in state

Table 3. Content Coding.

Topic	Components included in topic
General	1. Classroom management
	2. Student behavior
Maximize structure and predictability	1. Routines
	2. Transitions
	3. Physical arrangement
	4. Classroom structure
Post, teach, review, monitor, and reinforce expectations	1. Positive behavior support
	2. Rules
	3. Reinforcement system
	4. System to address misbehavior
	5. Positively stated expectations
	6. Supervises students
	7. Prompting or pre-corrects expectations
	8. Specific and contingent praise
	9. Acknowledge appropriate behavior
	10. Group contingencies
	11. Behavior contracts
	12. Token economies
Actively engage students	1. Actively engages students and/or engages students in learning
	2. Opportunities to respond
Addresses inappropriate behavior	1. Strategies for responding to inappropriate behavior
	2. Differential reinforcement
	3. Planned ignoring
	4. Response cost
	5. Time out
Learning environment	1. Learning environment
	2. Classroom environment/climate
	3. Environment that facilitates learning

teacher evaluation rubrics at the standard and indicator levels. In Table 1, we present the number of standards, number of standards about classroom management, number of indicators, and number of indicators about classroom management for each state rubric we reviewed. The average number of standards on the rubrics was 5.8 (SD = 2.62), with a range of two to 10. The average number of standards about classroom management was 1.03 (SD = 0.49), with a range of zero to two. Four states did not have any standards that addressed classroom management, 21 states had one standard addressing classroom management, and five states had two standards addressing classroom management. The *FFT* included four standards about classroom management, 25% of the total *FFT* standards. On

the *InTASC* rubric, one out of the 10 standards involved classroom management. The percentage of standards about classroom management ranged from 0% to 50%. The average percentage of standards about classroom management on the rubrics we reviewed was 20.47 (SD = 12.10).

Indicators. There was more variability in the number of indicators across rubrics (see Table 3). The average number of indicators was 27.07 (SD = 16.28), with a range of five to 75. The average number of indicators about classroom management was 4.70 (SD = 2.28), ranging from one indicator to nine indicators about classroom management. The percentage of indicators about classroom management ranged from 1.33% to 60%. Louisiana

Table 4. Topics Covered by Indicators Related to Classroom Management.

	General classroom management	Maximize structure and predictability	Post, teach, review, monitor, and reinforce expectations	Actively engage students	Addresses inappropriate behavior	Learning environment
Indicators (*n* = 141)	0.71%	34.04%	28.37%	40.43%	13.48%	37.59%

Note. The percentage of indicators addressing each content area does not sum to 100% because indicators could receive multiple content codes.

had three indicators about classroom management and a small total number of indicators (five), resulting in the highest percentage (60%) of indicators about classroom management. The *FFT* included 22 total indicators, with six, 27.27%, related to classroom management. Out of 75 total indicators on the *InTASC* standards, there was one related to classroom management. The average percentage of indicators about classroom management was 22.13 (*SD* = 12.80).

Classroom Management Content

After we identified the standards and indicators on each rubric related to classroom management, we coded each indicator by topic. These results are reported in Table 4. Indicators most often addressed a teachers' ability to actively engage their students (40.3%). Many indicators, 37.59%, dealt with the classroom environment or climate. Thirty-four percent and 28% percent of indicators dealt with maximizing structure and predictability or establishing rules, respectively. Only one indicator addressed classroom management generally without addressing one of the more specific content areas (0.71%). Only 13.79% of indicators made any mention of how teachers addressed inappropriate student behavior.

Discussion

The purpose of this study was to examine the extent to which classroom management is included on observation rubrics used for teacher evaluation. Classroom management is an important skill for all teachers, especially for teachers working with SWDs. States have

adopted new teacher evaluation systems in an effort to improve teaching effectiveness, but rubrics that exclude classroom management may overlook an important teacher skill related to student outcomes.

We found that about 20% of standards and indicators on the evaluation rubrics we reviewed addressed classroom management. The classroom management topics addressed by the indicators varied, with most focused on proactive classroom management strategies such as actively engaging students, posting and reinforcing expectations, and maximizing the structure and predictability of the classrooms. Very few indicators, 13.48%, addressed how a teacher responded to inappropriate behavior. In the following sections, we discuss how classroom management is currently integrated into evaluation rubrics, the focus on proactive strategies, and recommendations for improving and augmenting evaluation rubrics.

Use of Rubrics as a Tool

Evaluation rubrics include indicators that outline the specific practices on which teachers are rated by evaluators. These ratings translate into high-stakes outcomes for teachers: a teacher may be retained, given a leadership opportunity, or, in some districts, paid more based in some part on these ratings (Steinberg & Donaldson, 2016). More importantly, evaluations are a method for helping teachers to improve their practice. Given the importance of rubrics as development tools, the language used to describe desired practices, specifically for indicators, should be clear, specific, and actionable.

We found that both the number and, more importantly, the explicitness of indicators related to classroom management varied across rubrics. This can be problematic for both teachers and evaluators. Other researchers (Hill & Grossman, 2013; Kraft & Gilmour, 2016) point out that the challenges of implementing evaluation could be partially due to vague language. For example, language such as "creates a safe and purposeful learning environment" does not provide observable information about what safe looks like. On the contrary, language such as "the teacher creates a safe and purposeful learning environment by setting clear expectations to guide student classroom behavior" provides more utility to teachers by indicating a practice they can use that can facilitate student learning. Rubrics are unlikely to assist teachers in changing their practice, or provide guidance to evaluators on how to assist teachers in changing their practice, if the language used on the rubric does not precisely describe desired teacher practices. This seems particularly important in regard to classroom management, as many states focused on less well-defined topics such as the learning environment or classroom climate.

Focus on Prevention

On the rubrics we reviewed, most of the indicators that addressed classroom management focused on proactive, preventive strategies, but the indicators did not include guidelines for responding to negative behavior. For example, "establishes clear expectations for classroom rules, routines, and procedures" is a strategy that can be implemented class-wide to promote positive behavior and lessen the likelihood of misbehavior (Georgia Department of Education, 2011, 7.2), while "redirects student off-task behavior to make the most of instructional time" is a method to counteract problem behavior that has already occurred (California Standards for the Teaching Profession, 2009, 2.7). California's focus on what to do when misbehavior occurs was rare; only 15 other rubrics mentioned responding to inappropriate behavior.

Clearly, helping teachers take a proactive approach to classroom management is important, but the nearly total emphasis on universal behavior support may be problematic when teachers have students with significant behavioral needs, such as those with or at risk for emotional and behavioral disorder (EBD) in their classrooms. Prevention is important and a large component of School-Wide Positive Behavioral Interventions and Supports (SWP-BIS) and other multitiered systems for supporting student behavior, but within a tiered system, a small proportion of students are likely to exhibit problem behavior despite generally effective, proactive classroom management strategies (Sugai & Horner, 2006). The absence of expectations for how to respond when students do misbehavior may leave teachers without tools for appropriately responding to misbehavior. This omission is particularly alarming for teachers who have students with or at risk for EBD, students who may exhibit frequent problem behavior, in their classrooms. For these students, a de-emphasis on effective teacher response to negative behavior could result in punitive disciplinary action, loss of instructional time, and more restrictive placements (U.S. Department of Education, 2016). Focusing on proactive classroom management strategies is important, but teachers may need more guidance on how to address misbehavior when it arises than the guidance and strategies included in evaluation rubrics.

Limitations

The results of this review should be interpreted with limitations in mind. We examined state-developed model or required rubrics, the *InTASC* standards, and *FFT*. Many states allow districts to use locally designed rubrics; therefore, rubrics used by districts may vary from the rubrics we reviewed. In addition, we evaluated the content of the rubrics but not the actual implementation. The utility of a rubric depends in part on the feedback teachers receive from their evaluators. Despite these limitations, the descriptive information in this review is useful for policymakers and practitioners.

Recommendations

Based on the results of this review, we offer four recommendations. First, many state rubrics in this review included classroom management standards, but not all. As most rubrics contained an overwhelming number of indicators, implementers may attend more to the overall standards when providing feedback and development opportunities to teachers (Kraft & Gilmour, 2016). States that have not included a standard related to classroom management should include such a standard on subsequent versions of observation rubrics. This is especially pertinent given that many teachers struggle to manage problem behavior in the classroom (NCES, 2015).

Our second recommendation is related to the utility of the indicators. Most standards across the states did not provide explicit descriptions of desired teacher behaviors, but the corresponding indicators did provide more specific information, particularly around academic instruction. These indicators clarified teacher behaviors that earned a proficient or higher rating on the standard, thus providing useful information for feedback and coaching. For these indicators to be useful to teachers and evaluators, however, explicit descriptions of teacher behaviors related to classroom management should be included. For example, an indicator stating "the teacher will create a positive classroom environment" is less useful than an indicator that states "the teacher will teach, post, and reinforce positive student behavior expectations." Rubrics with vague language should be revised to include explicit descriptions of expected classroom management behavior.

Third, most evaluation rubrics do not address teacher responses to student misbehavior. While we applaud the focus on positive, proactive strategies that align with SWPBIS and other multitiered approaches to behavior, we question the exclusion of what a teacher should do when problem behavior occurs. Prevention might be enough for many students, but teachers likely need more support for what to do when they have students who are more likely to exhibit problem behavior, such as SWDs, in their classes. Perhaps including strategies for addressing misbehavior does not belong on evaluation rubrics, but most teachers will likely need access to additional guidance on how to respond to student behavior in the likely situation that it occurs.

Fourth, evaluation users, at the teacher preparation level or in schools, may need to supplement evaluation systems with other measures that address classroom management depending on the extent to which their local rubrics address classroom management. As teacher education programs aim to align their evaluation and training of preservice teachers with the skills and competencies that these teachers will be evaluated on once in the field, the lack of clarity, or inclusion, of classroom management skills may decrease the amount of time focused on these important skills. Teacher preparation programs and in-service teacher evaluators could supplement their observation protocols with direct observation measures such as the *Brief Student-Teacher Classroom Interaction Observation* (Reinke, Herman, & Newcomer, 2016). We recognize that it might be infeasible to add additional measures to already complicated evaluation systems.

In recent years, states have heavily invested in teacher evaluation systems as one method to improve teaching quality. In this article, we reviewed state evaluation rubrics to examine the extent to which classroom management was included in observation rubrics. We found that in most states one fifth of the standards and indicators were related to classroom management, but the topics included in the rubrics, and the usefulness of language, varied widely. Some states, however, did not include any standards related to classroom management. This is worrisome in light of prior research suggesting that teachers enter the workforce unprepared in classroom management and in need of more in-service support with these skills. For rubrics to support teacher development in classroom management, they must provide detailed information about classroom management that is useful to teachers and evaluators. This is especially important for

supporting teachers, in both general education and special education settings, who instruct SWDs. Classroom management skills are integral for the successful inclusion of SWDs in general education classrooms and for improving SWDs' academic outcomes.

Appendix

States That Were Not Included and Reasons for Exclusion.

State	Exclusion reason
AZ	District choice/State does not provide a model or suggested evaluation rubric.
AR	Uses FFT rubric.
FL	Uses Marzano rubric.
HI	Uses FFT rubric.
DE	Uses FFT rubric.
ID	Uses FFT rubric.
KT	Uses FFT rubric.
MD	Uses FFT rubric.
MI	District choice/State does not provide a model or suggested evaluation rubric.
NH	District choice/State does not provide a model or suggested evaluation rubric.
NJ	District choice/State does not provide a model or suggested evaluation rubric.
NY	District choice/State does not provide a model or suggested evaluation rubric.
ND	District choice/State does not provide a model or suggested evaluation rubric.
OK	District choice/State does not provide a model or suggested evaluation rubric.
OR	District choice/State does not provide a model or suggested evaluation rubric.
PA	Uses FFT rubric.
SD	Uses FFT rubric.
VT	District choice/State does not provide a model or suggested evaluation rubric.
WA	District choice/State does not provide a model or suggested evaluation rubric.
WI	Uses FFT rubric.
WY	District choice/State does not provide a model or suggested evaluation rubric.

Note. FFT = Framework for Teaching.

Authors' Note

The second and third authors contributed equally to this article.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

American Institutes for Research. (2015). *Databases on state teacher and principal evaluation policies*. Washington, DC: Center on Great Teachers & Leaders at American Institutes for Research. Retrieved from http://resource.tqsource.org/stateevaldb/UseResource.aspx

Blackorby, J., Wagner, M., Cameto, R., Davies, E., Levine, P., Newman, L., . . . Sumi, C. (2005). *Engagement, academics, social adjustment, and independence: The achievements of elementary and middle school students with disabilities*. Washington, DC: SRI International.

Brownell, M. T., Bishop, A. G., Gersten, R., Klingner, J. K., Penfield, R. D., Dimino, J., . . . Sindelar, P. T. (2009). The role of domain expertise in beginning special education teacher

quality. *Exceptional Children*, *75*, 391-411. doi:10.1177/001440290907500401

Buzick, H. M., & Jones, N. D. (2015). Using test scores from students with disabilities in teacher evaluation. *Educational Measurement: Issues and Practice*, *34*(3), 28-38. doi:10.1111/emip.12076

California Standards for the Teaching Profession. (2009). Retrieved from http://www.ctc.ca.gov/educator-prep/standards/CSTP-2009.pdf

Cooper, J. T., & Scott, T. M. (2017). The keys to managing instruction and behavior: Considering high probability practices. *Teacher Education and Special Education*, *40*, 102-113. doi:10.1177/0888406417700825

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

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

De Pry, R. L., & Sugai, G. (2002). The effect of active supervision and pre-correction on minor behavioral incidents in a sixth grade general education classroom. *Journal of Behavioral Education*, *11*, 255-267. doi:10.1023/A:1021162906622

Doherty, K. M., & Jacobs, S. (2015). *State of the states: Evaluation teacher, leading, and learning*. Washington, DC: National Council on Teacher Quality. Retrieved from http://www.nctq.org/dmsView/StateofStates2015

Freeman, J., Simonsen, B., Briere, D. E., & MacSuga-Gage, A. S. (2014). Pre-service teacher training in classroom management: A review of state accreditation policy and teacher preparation programs. *Teacher Education and Special Education*, *37*, 106-120. doi:10.1177/0888406413507002

Garwood, J. D., Vernon-Feagans, L., & Family Life Project Key Investigators. (2017). Classroom management affects literacy development of students with emotional and behavioral disorders. *Exceptional Children*, *83*, 123-142. doi:10.1177/0014402916651846

Georgia Department of Education. (2011). *Teacher keys effectiveness system*. Retrieved from http://www.nctq.org/docs/TKES_Handbook_-_FINAL_5-30-14_(1).pdf

Greenberg, J., Putman, H., & Walsh, K. (2014). *Training our future teachers: Classroom management*. Washington, DC: National Council on Teacher Quality.

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment. *Educational Researcher*, *43*, 293-303. doi:10.3102/0013189X14544542

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The Relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education*, *119*, 445-470. doi:10.1086/669901

Hansen, S. D., & Lignugaris-Kraft, B. (2005). Effects of a dependent group contingency on the verbal interactions of middle school students with emotional disturbance. *Behavioral Disorders*, *30*, 170-184.

Hill, H. C., & 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

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, *48*, 794-831. doi:10.3102/0002831210387916

Holdheide, L. R., Goe, L., Croft, A., & Reschly, D. J. (2010). *Challenges in evaluation special education teachers and English language learners specialist* (TQ research and policy brief). Washington, DC: National Comprehensive Center for Teacher Quality.

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

Jones, N., & Gilmour, A. (in press). Special education teacher evaluation: Examining current practices and research. In J. B. Crockett, B. Billingsley, & M. L. Boscardin, (Eds.), *Handbook of leadership and administration for special education* 2nd ed. (pp. 458–477). New York: Taylor & Francis.

King-Sears, M., Carran, D. T., Dammann, S. N., & Arter, P. S. (2012). Multi-site analyses of special education and general education student teachers' skill ratings for working with

students with disabilities. *Teacher Education Quarterly*, *39*, 131-149.

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*, 711-753. doi:10.1177/00131 61X16653445

Lambert, M. C., Cartledge, G., Heward, W. L., & Lo, Y. (2006). Effects of response cards on disruptive behavior and academic responding during math lessons by fourth-grade urban students. *Journal of Positive Behavior Interventions*, *8*, 88-99. doi:10.1177/10983007060080020701

Martin, K. F., Lloyd, J. W., Kauffman, J. M., & Coyne, M. (1995). Teachers' perceptions of educational placement decision for pupils with emotional or behavioral disorders. *Behavioral Disorders*, *20*, 106-117.

Miles, M., & Huberman, M. (1994). *Qualitative data analysis: A expanded sourcebook* (2nd ed.). Thousand Oaks, CA: SAGE.

Mitchell, B. S., Hirn, R. G., & Lewis, T. J. (2017). Enhancing effective classroom management in schools: Structures for changing teacher behavior. *Teacher Education and Special Education*, *40*, 140-153. doi:10.1177/0888406417700961

Moore-Partin, T. C., Robertson, R., Maggin, D. M., Oliver, R. M., & Wehby, J. H. (2010). Using teacher praise and opportunities to respond to promote appropriate student behavior. *Preventing School Failure*, *54*, 172-178. doi:10.1080/10459880903493179

National Center for Education Statistics. (2015). *Schools and Staffing Survey*. Washington, DC: Institute of Education Sciences. Retrieved from https://nces.ed.gov/surveys/sass/

Oliver, R. M., & Reschly, D. J. (2010). Special education teacher preparation in classroom management: Implications for students with emotional and behavioral disorders. *Behavioral Disorders*, *35*, 188-199.

Olson, M. R., Chalmers, L., & Hoover, J. H. (1997). Attitudes and attributes of general education teachers identified as effective inclusionists. *Remedial and Special Education*, *18*, 28-35. doi:10.1177/074193259701800106

Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review*, *82*, 123-141.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational*

Researcher, *38*, 109-119. doi:10.3102/0013189X09332374

Reiff, H. B., Evans, E. D., & Cass, M. (1991). Special education requirements for general education certification: A national survey of current practices. *Remedial and Special Education*, *12*, 56-60. doi:10.1177/074193259101200508

Reinke, W. M., Herman, K. C., & Newcomer, L. (2016). The brief student-teacher classroom interaction observation: Using dynamic indicators of behaviors in the classroom to predict outcomes and inform practice. *Assessment for Effective Intervention*, *42*, 32-42. doi:10.1177/1534508416641605

Semmelroth, C. L., & Johnson, E. (2014). Measuring rater reliability on a special education observation tool. *Assessment for Effective Intervention*, *39*, 131-145. doi:10.1177/1534508413511488

Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education & Treatment of Children*, *31*, 351-380. doi:10.1353/etc.0.0007

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. doi:10.1177/0888406413489839

Soodak, L. C., Podell, D. M., & Lehman, L. R. (1998). Teacher, student, and school attributes as predictors of teachers' responses to inclusion. *Journal of Special Education*, *31*, 480-497. doi:10.1177/002246699803100405

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

Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's excellence in teaching project. *Education Finance and Policy*, *10*, 535-572. doi:10.1162/EDFP_a_00173

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

Stormont, M. A., Smith, S. C., & Lewis, T. J. (2007). Teacher implementation of precorrection and praise statements in Head Start classrooms as a component of a program-

wide system of positive behavior support. *Journal of Behavioral Education*, *16*, 280-290. doi:10.1007/s10864-007-9040-3

Sugai, G., & Horner, R. R. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review*, *35*, 245-259.

Sutherland, K. S., & Wehby, J. H. (2001a). The effects of self-evaluation on teaching behaviors in classrooms for students with emotional and behavioral disorders. *Journal of Special Education*, *35*, 161-171. doi:10.1177/002246690103500306

Sutherland, K. S., & Wehby, J. H. (2001b). Exploring the relation between increased opportunities to respond to academic requests and the academic and behavioral outcomes of students with EBD: A review. *Remedial and Special Education*, *22*, 113-121. doi:10.1177/074193250102200205

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, *102*, 3628-3651. doi:10.1257/aer.102.7.3628

U.S. Department of Education. (2006). *Facts from the NLTS2: School behavior and disciplinary experiences of youth with disabilities* [Fact sheet]. Retrieved from http://ies.ed.gov/ncser/pdf/NLTS2_discipline_03_21_06.pdf

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

U.S. Department of Education. (2016). *The digest of education statistics, 2015* (NCES 2016-014). Washington, DC: National Center for Education Statistics.

Westling, D. L. (2010). Teachers and challenging behavior: Knowledge, views, and practices. *Remedial and Special Education*, *31*, 48-63. doi:10.1177/0741932508327466

Williamson, B. D., Campbell-Whatley, G. D., & Lo, Y. (2009). Using a random dependent group contingency to increase on-task behaviors of high school students with high incidence disabilities. *Psychology in the Schools*, *46*, 1074-1083. doi:10.1002/pits.20445

Author Biographies

Allison F. Gilmour is an assistant professor of special education in the Department of Teaching and Learning at Temple University. She studies policies and interventions to improve teaching quality for students with disabilities.

Caitlyn E. Majeika is a PhD student in the Department of Special Education at Peabody College of Vanderbilt University. Her research interests include improving ways to more effectively implement Tier 2 behavioral interventions.

Amanda W. Sheaffer is a PhD student in the Department of Special Education at Peabody College of Vanderbilt University. Her research interests include girls with or at risk for emotional or behavioral disorders, school-based problem behavior, and classroom behavioral interventions.

Joseph H. Wehby is an associate professor in the Department of Special Education at the Peabody College of Vanderbilt University. His research interests are in classroom management, adaptive behavioral interventions for students with emotional/behavioral disorders, and observational methods.