Current State of High Stakes Teacher Evaluation for Special Education Teachers

Robin A. Snyder[1] and Lisa A. Pufpaff[2]
[1]Berrien RESA, Berrien Springs, MI
[2]Department of Special Education, Ball State University

Education reform legislation has led to an upwelling of mandatory teacher evaluation for all elementary, middle school, and high school educators, including those who teach special education. While this reform effort aimed to improve overall teacher effectiveness and student learning outcomes, the teacher evaluation for special education is a retrofitted version of the general education teacher evaluation model and yields little meaningful information and increased stress levels for both the special education teacher and their evaluator. Before this problem can be rectified, the barriers standing in the way to a meaningful special education teacher evaluation need to be identified. This article examines existing literature related to special education teacher evaluation to identify those barriers, so progress can be made to improve the delivery of special education instruction and outcomes for students with disabilities.

Keywords: teacher evaluation, special education teacher, barriers, special education teacher roles, student growth, evaluator knowledge of special education

The fundamental purpose of teacher evaluations is to assist in providing effective services for students (Joint Committee on Standards for Educational Evaluation, 2009). Teacher evaluation is a formal process designed to measure teacher quality and encourage professional growth through feedback (Darling-Hammond et al., 2012; Council for Exceptional Children, 2012; Marzano, 2012a; Marzano, 2012b). In some states, teacher evaluation ratings are used to inform decisions relative to the promotion and retention of teachers, tenure, certification, termination, and salary (Holdheide et al., 2010; Semmelroth & Johnson, 2014). Information from evaluations can also be used by administrators for the planning of professional development activities.

The purpose of evaluating teachers is to improve overall teacher effectiveness and student learning outcomes. However, for special education teachers, that goal is unlikely met due to the poor fit between the evaluation systems being used and the roles of special education teachers.

Evaluation practices used with special education teachers are retrofitted versions of those developed for use among general education teachers.

Measures of Teacher Effectiveness in Relationship to Current Legislation

Teacher supervision systems have existed since the 1700s when the quality of instruction provided by teachers was monitored by clergy and members of local government. Supervision systems have evolved along with shifts in views and theories about education and related research. Since the early 2000s, the emphasis has shifted from the supervision of teachers and their behaviors that influence teaching to the evaluation of teachers relative to student achievement (Marzano et al., 2011). The timeline of the shift from teacher supervision to teacher evaluation practices in the United States coincides with the passing of several pieces of federal legislation designed to improve the learning outcomes of students in public schools. The legislation includes No Child Left Behind (2001), Race to the Top (Department of Education, 2009), the Elementary and Special Education Act (2012), and Every Student Succeeds Act (2015).

No Child Left Behind (NCLB) Act of 2001

The 1983 report, *A Nation at Risk*, described how the education system in the United States was failing to educate students. NCLB was a response to *A Nation at Risk*. NCLB addressed: all students attaining grade-level proficiency in reading and math, all students graduating from high school, and the evaluation of all teachers. Teacher evaluation, including special education teachers, was based on student growth to determine teacher success in order to receive federal waivers. NCLB significantly increased the federal government's role in education (Mills, 2008; NCLB, 2002).

There were several notable features of NCLB that impacted the evolution of the teacher evaluation system in the United States. First, one of the primary requirements of NCLB was accountability. All students, including students with disabilities, were required to participate in high-stakes, statewide standardized tests. The results from those tests were used to determine if schools made adequate yearly progress (AYP). AYP was the amount of yearly progress students in a school were expected to make on high-stakes, statewide assessments in order to ensure students who were lower achieving could meet the high-performance criteria NCLB established for all students. Failure of a school to make AYP had implications for funding and enrollment (Mills, 2008; NCLB, 2002).

Second, NCLB required the use of scientifically-based research standards to ensure teaching strategies and methods based on research (i.e., evidence-based practices). Additionally, NCLB required that teachers must be highly qualified. Attaining highly-qualified status required a teacher to have full state certification, a bachelor's degree, and pass a subject matter knowledge test. Additionally, tracking of student progress over time was used to measure a teacher's contribution to student learning (Mills, 2008; NCLB, 2002).

Race to the Top (RttT)

Race to the Top (RttT) was established by the United States Department of Education (DOE) as part of the American Recovery and Reinvestment Act (2009). RttT was a competitive incentive federal grant aimed at implementing education reforms in the areas of implementing standards, improving teacher effectiveness, improving data collection, use

of data for educational decision-making, and supporting struggling schools (DOE, 2009).

RttT incentivized states to adopt annual teacher evaluations that included measures of student growth and multiple performance rating categories. It also allowed for teacher pay incentives and promotions based on evaluation results (Croft et al., 2018; DOE, 2009; Education Commission of the States, 2018; Holdheide et al., 2010). The emphasis on teacher effectiveness in RttT triggered states to examine and revise their teacher evaluation policies (Glowacki & Hackman, 2016; Howell, 2015). As a result of RttT, the majority of states adopted legislation that mandated the use of measures of student growth into teacher evaluations; these included value-added methods, school-wide growth, and student learning objectives (Croft et al., 2018; Holdeheide et al., 2010).

Elementary and Secondary Education Act (ESEA) Flexibility 2012

ESEA flexibility plans (U.S. DOE, 2012) for states were granted in 2012 to provide relief from some of the provisions of NCLB that continually identified schools as failing even though there was evidence to demonstrate that gains were made relative to improved student outcomes. ESEA 2012 contained provisions of flexibility and alternative measures for determining student growth in grades and subject areas, such as special education, that were not included in high-stakes, statewide assessments. Alternative measures of student growth included student results on pre-/post-tests, end-of-course tests, objective performance-based assessments, student learning objectives, and performance on English language proficiency assessments. ESEA 2012 encouraged states to use teacher evaluation

systems that provided timely and useful feedback that identified a teacher's areas of strength and areas that needed improvement. Information from the teacher evaluation was to be used to guide individualized professional development plans designed to help each teacher increase pedagogical skills and grow professionally (Croft et al., 2018; Education Commission of the States, 2018).

Every Student Succeeds Act (ESSA) of 2015

ESSA returned back to states and local districts much of the decision-making power relative to education that was given to the federal government under NCLB. ESSA allowed for state flexibility in developing teacher evaluation systems. It also discontinued the focus on student growth as a means of determining teacher effectiveness. ESSA provided funds to develop teacher evaluation systems that better differentiated between effective and ineffective teachers using rigorous, fair, transparent methods that included evidence of student achievement, but not necessarily measures of student growth (Croft et al., 2018; Education Commission of the States, 2018).

Legislation Since ESSA

ESSA (2015) had an impact on state legislation relative to teacher evaluations. In 2015, 43 states had teacher evaluation systems that required measures of student growth as a component of determining teacher effectiveness. By the end of 2017, 39 states required student growth in teacher evaluations. Additionally, as of 2017, ten states passed legislation or adopted resolutions that lessened the impact of student growth in teacher evaluations (Education Commission of the States, 2018). For example, Indiana enacted legislation (H.B. 1003) that provided districts with flexibility in the use of student

growth in teacher evaluations, Michigan (S.B. 133) removed the student growth requirement and provided district flexibility, and Tennessee (H.B. 309) temporarily reduced the student growth impact in teacher evaluations.

Evaluation Models Used Among Special Education Teachers

The teacher evaluation process is complex and challenging for general education teachers but even more so for special education teachers. The effectiveness of general education teachers is measured against grade-level norms, standardized assessments, and common learning standards. Although there is variability among systems used to evaluate teachers, the majority of evaluation systems include formal and informal observations by an administrator. Growth measures and rubrics are also used to rate the performance of teachers. Growth measures or student growth scores are typically calculated using student results from standardized state assessments (Sawchuk, 2015).

There are three main types of growth measures. First, vertical scales are a growth measure where student growth is calculated by the use of equivalent tests to calculate the student's difference in scores (i.e., growth or gain score) at two different periods of time (Marzano, 2014; Popham, 2013). Second, student growth percentiles use a statistical calculation to compare a student's growth to their academic peers with similar prior test scores. A student growth percentile represents how much a particular student grew in comparison to their academic peers (Marzano, 2014). Third, value-added measures (VAMs) are calculated using predictor variables that could include student demographic information, prior achievement, and time spent in a specific teacher's class (Betebenner, 2009; Marzano, 2014).

Rubric-based evaluation instruments are used to measure pedagogical skills during observations. Rubric-based evaluation instruments include The Marzano Teacher Evaluation Model (Learning Sciences Marzano Center, 2013), Charlotte Danielson's Framework for Teaching (FFT; Danielson, 2014) and the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008). Each of these rubric-based evaluation instruments breaks pedagogical skills (e.g., student interactions, instruction, planning, assessment, and professionalism) into domains or categories. The observer/evaluator then uses a rating scale to rate a teacher's effectiveness in each category based on the criteria outlined in the rubric (Jones & Brownell, 2013; Learning Sciences Marzano Center, 2013).

The intent of the teacher evaluation process is to measure quality teaching and encourage professional growth to assist in providing effective services for students (CEC, 2012; Darling-Hammond et al., 2012; Joint Committee on Standards for Educational Evaluation, 2009; Marzano, 2012a; Marzano, 2012b). Current evaluation instruments and practices may be adequate for general education teachers; however, research indicates they are not sufficient for special education teachers (CEC, 2012). The goals of measuring quality teaching and encouraging professional growth are unlikely to be met because of the poor fit between the current evaluation instruments and the jobs of special education teachers.

At present, the majority of teacher evaluation instruments and practices do not align with the jobs and duties of special education teachers. They are retro-fitted versions of the evaluation instruments and processes specifically designed for use with

general education teachers. The misalignment of the instruments and processes compromises the validity and reliability of the outcomes and brings in to question how much meaningful information they can yield relative to special education teachers.

Impact of Evaluation Results

Special education teachers often feel overworked, underappreciated, and discouraged in comparison to their general education counter-parts (Berry, 2012; McLeskey & Billingsley, 2008). Feelings of frustration, coupled with evaluation results that inadequately measure their effectiveness, has resulted in qualified and experienced special education teachers leaving the profession (Ryan et al., 2017; The IRIS Center, 2013; Thornton et al., 2007). The field of special education is already faced with a critical shortage of qualified special education teachers (Berry, 2012; Irinaga-Bistolas et al., 2007; Williams & Dikes, 2015) and this shortage is leading districts to hire individuals who do not have the credentials or training needed to teach students with disabilities (Billingsley & Bettini, 2019). These teachers are often less effective than teachers who have been formally trained to teach special education, ultimately leading to reduced learning outcomes for students with disabilities (Williams & Dikes, 2015). This practice can exacerbate the challenge of attracting and retaining high-quality teachers in special education positions in school districts (Johnson & Semmelroth, 2014: Irinaga-Bistolas et al., 2007).

The fallout from inadequate evaluation of special education teachers can also have significant financial implications for a school district. Although students with disabilities make up a small proportion of the total student population, their progress counts towards a school's and district's AYP. If students in the special education subgroup, or any subgroup, do not meet performance expectations, there can be significant consequences for the administrator, school, and the district (Wakeman et al., 2006). Evaluators must be able to provide meaningful feedback to special education teachers to improve their instructional practices and, ultimately, student outcomes to help ensure AYP in the special education subgroup to receive special education funds.

Barriers to Providing Valid and Reliable Evaluations for Special Education Teachers

The process of evaluating special education teachers is extremely challenging. Research in the area of special education teacher evaluations is limited. However, common barriers identified include measuring student growth, differing roles of general and special education teachers, the consideration of the special education teacher in the development of teacher evaluation systems, and the evaluator's limited knowledge of special education.

Student Growth

Measures of student growth are included in teacher evaluations as a means of demonstrating teacher effectiveness. For general education teachers, student growth is frequently determined based on student performance on standardized state assessments aligned with state academic standards. For special education teachers, student growth can also be determined based on student performance on standardized state assessments or alternative state assessments.

However, the use of standardized test scores for teacher evaluations to demonstrate teacher effectiveness for both general education and special education teachers is inherently flawed. Standardized

test scores are designed, tested, and normed for evidence of reliability and validity in the measurement of student achievement. Standardized assessments are not designed nor intended to measure teacher effectiveness. They cannot be considered a reliable or valid measure of teacher effectiveness until the tests have undergone the same level of testing that was conducted to determine their reliability and validity in determining student achievement. Until that level of testing occurs, student results on standardized assessments only demonstrate an indirect measurement of teacher effectiveness (Betebenner, 2009; Steinbrecher et al., 2014).

Additionally, the use of scores from standardized assessments to show growth for a student with a disability can be misleading and can demonstrate little, if any, progress. Standardized assessments will yield skewed results for students who are performing below chronological grade level, which can adversely influence measures of student growth and consequently perceived teacher effectiveness. Skewed standardized assessment results can be attributed to floor effects (Steinbrecher et al., 2014).

Floor effects occur when a student is performing at a grade level lower than the grade level of the assessment the student is required to take. When the student's level of achievement is below the threshold for the grade level of the assessment, the entire assessment is too difficult for the student. Thus, little variation in the student's performance between test items will occur. The further a student with a disability is behind their chronological grade level, the less reliable the results from a standardized assessment become, and the more of an impact floor effects will have.

Floor effects are consistent with results that show an absence of student progress and insinuate low levels of teacher effectiveness (Steinbrecher et al., 2014).

Student learning objectives (SLOs) are used in some states to determine student growth for grade-levels and subjects where standardized test scores are unavailable, including special education. As part of the SLO process, a teacher identifies expected learning outcomes, aligned with state standards, for a group of students for a given school year based on the students' present levels. Throughout the school year, student progress on SLOs is measured using formative and summative assessments. Teacher effectiveness is determined based on the students' progress toward SLOs (Holdheide et al., 2012).

While SLOs could be an option for calculating student growth for students with disabilities, there are some challenges. First, student progress falling below the learning target for the SLO may go undocumented and unnoticed. Falling below the learning target of an SLO is a common occurrence for students with disabilities, as their progress is often slower than that of students without disabilities. Also, SLOs necessitate that teachers have the capacity to interpret data and understand that student growth may be limited against rigorous standards-aligned goals (Holdheide et al., 2012).

Alternative assessments are also means for states to obtain student growth data for students with more significant disabilities whose IEP team determines that even with accommodations, the state standardized assessment is not appropriate. The alternative assessments themselves can be problematic. Many alternative assessments are portfolio based and lack comprehensive evidence to support the validity and reliability of the instrument

(Cameto et al., 2009; Holdheide, et al., 2012; Steinbrecher et al., 2014; Towles-Reeves et al., 2009). The alternative assessments are also poorly aligned with grade-level content standards. Additionally, the alignment of grade-level content standards varies wildly from state to state. Furthermore, the heterogeneity of students with more significant disabilities makes it difficult to develop standardized measures that account for the variance in learning trajectories (Holdheide et al., 2012). Thus, the resulting student growth data from alternative assessments are subjective, variable, and provide weak evidence of teacher effectiveness (Holdheide et al., 2012; Steinbrecher et al., 2014).

Differing Roles

The role of the special education teacher is more varied and complex than that of a general education teacher. The main roles of the general education teacher are to deliver instruction, assess student competencies of skills and knowledge identified in state grade-level content standards, and manage the classroom environment. The roles of the special education teacher vary widely based on the nature and severity of the disabilities of the students with whom the special education teacher works.

In addition to providing instruction for IEP goals and differentiated academic content at multiple grade levels, the special education teachers' roles frequently include: providing instruction on adaptive, communication, social, mobility and functional skills; managing challenging behaviors and teaching replacement behaviors; completing standardized and criterion-referenced assessments; consulting and collaborating with other teachers and other school personnel (e.g., paraprofessionals, school psychologist, speech-language pathologist, occupational therapist, behavior consultant, school nurse); and completing required paperwork (CEC, 2012; Ruppar et al., 2015). Special education teachers collaborate with general education teachers to develop and implement strategies for working with students with IEPs so they can make progress in the general education setting (Williams & Dikes, 2015). They work with, train, and supervise paraprofessionals to provide support for each student with an IEP in general education and special education settings (Crowe et al., 2017; Ruppar et al., 2015). Special education teachers collaborate with other professionals, including speech-language therapists, physical therapists, behavior specialists, occupational therapists, and medical professionals to integrate and implement their recommendations into the instructional program for students with disabilities (Ruppar et al., 2015; Williams & Dikes, 2015). They also complete extensive educational and legally required paperwork (CEC, 2012; Elliot et al., 2014; Ruppar et al., 2015).

While providing academic instruction is the most significant component of a special education teacher's day, research has shown it only accounts for 15.6% of the school day (Glowacki & Hackman, 2016; Vannest & Hagan-Burke, 2010). Providing instructional support, including curriculum and material modification, behavior management, assistive technology, and augmentative and alternative communication (AAC) comprises 14.6% of the school day. Paperwork duties account for 12.1% of the school day including: referrals, IEPs, data collection, progress monitoring, evaluations/re-evaluations, functional behavioral assessments, behavior intervention plans,

and general documentation of events, incidents, and conversations. The remainder of the average school day for a special education teacher is comprised of personal time (9.5%), consultation and collaboration (8.6%), other responsibilities (7.9%), supervision (7.2%), planning (5.4%), non-academic instruction (4.4%), assessment (4.4%), and IEP meetings (2.9%) (Vannest & Hagan-Burke, 2010).

The teacher's ability to deliver academic instruction and manage the classroom environment is evaluated during observations. The observation component of many teacher evaluation systems is often weighted most heavily, sometimes accounting for up to 75% of a teacher's total evaluation score (Michigan Department of Education, 2016). When combined, the categories of academic instruction and instructional support only account for 30% of a special education teacher's school day. The remaining 70% of the roles and duties required of a special education teacher are rarely acknowledged in current evaluation systems (Vannest & Hagen-Burke, 2010).

Consideration of Special Education in Teacher Evaluation Instruments

Teacher evaluation instruments were designed to evaluate teachers in the general education setting who are providing instruction to students with no identified disabilities. Self-contained special education classrooms were intentionally excluded in the development and validation of the Classroom Assessment Scoring System (Pianta et al., 2008) and the Framework for Teaching (Danielson, 2014) because the developers felt measures of effectiveness being used in the tools might not apply to the self-contained special education setting. Critical differences between the special education self-contained setting and the general education setting cited as reasons

for exclusion included atypical student behaviors, varying learning needs, physical demands, number of transitions, and multiple service providers. Exclusion from the development and validation process based on the unique attributes of special education classrooms and students suggested teacher evaluation instruments were not explicitly designed to evaluate special education teachers, are not reliable for use among special education teachers, and fail to provide a meaningful evaluation of special education teachers (Crowe et al., 2017).

The Classroom Assessment Scoring System (Pianta et al., 2008) and the Framework for Teaching (Danielson, 2014) have been mentioned in the literature relative to special education teacher evaluation (Johnson & Semmelroth, 2014; Jones & Brownell, 2013; Noell et al., 2014; Sawchuk, 2015; Semmelroth & Johnson, 2014; Semmelroth et al., 2013). Further examination of each of these instruments was done to see how the designers viewed the validity and reliability of their use for special education teachers.

Classroom Assessment Scoring System (CLASS)

The CLASS evaluation instrument was designed and normed for observations of teachers who teach typically developing students in general education settings. CLASS uses four cycles of fifteen-minute observations, conducted by a trained and certified CLASS observer, to measure teacher-student interactions in the classroom. The results from the observations are used to identify areas of professional growth for teachers, particularly as they pertain to improving student interactions (Hadden & Mountz, n.d.) given that "CLASS is the only validated

tool that specifically addresses teacher-child interactions" (Hadden & Mountz, p. 3).

Initial studies of the CLASS evaluation instrument were conducted in general education classrooms. CLASS developers cited concerns that the instrument's effectiveness may not apply to special education settings. Although CLASS is used in classrooms containing students with disabilities, the publishing company explicitly notes using the instrument to evaluate these teachers could be problematic, particularly if the classroom is a self-contained special education setting. Hadden and Mountz (n.d.) cautioned that a teacher's scores can be adversely impacted by the behavior of a child who disrupts classroom activities and upsets other children. Behavior Management, Instructional Learning Formats, and Productivity scores can also be impacted when a teacher is required to spend significant time dealing with problem behaviors (Hadden & Mountz, n.d.). While interfering and problem behaviors are not exclusive to self-contained special education classrooms, they are commonplace in that setting and, even if handled appropriately, would, by design, adversely impact the results of a special education teacher being evaluated using the CLASS.

There are additional hindrances for the use of the CLASS among special education teachers. First, when using the CLASS, the observer must use the version of the instrument that corresponds to the chronological age of the majority of the students, not the developmental age. Second, the CLASS protocol does not allow for accommodations, modifications, or exemptions to address unique circumstances that may be observed in a self-contained special education classroom. Also, the publishing company acknowledges that the CLASS may not be appropriate for the evaluation of teachers in classrooms where high percentages of students have significant needs that impact cognition and communication (Hadden & Mountz, n.d.).

Framework for Teaching (FFT)

FFT (Danielson, 2014) was developed from the Praxis III teacher certification examination. The FFT instrument is used to evaluate teachers on 22 components that are grouped in four domains: (1) Planning and Preparation, (2) The Classroom Environment, (3) Instruction, and (4) Professional Responsibilities. Each component defines a specific aspect of a domain (e.g., a component of the domain The Classroom Environment is "Creating an environment of respect and rapport"). Then each of the components is further divided into elements that describe a specific feature of teaching (e.g., a feature of "Creating an environment of respect and rapport" is "Teaching interactions with students"). As with CLASS, observations are conducted by a trained and certified FFT rater. The developers of the FFT feel that all teachers should be evaluated using the same set of standards. Consequently, there are no explicit directions, accommodations, or considerations for evaluating special education teachers using the FFT tool (Jones & Brownell, 2013).

FFT has been adopted for use in evaluation systems in many school districts and states. However, because the scope of the evaluation instrument is so vast, some districts are choosing modified versions of FFT that only use Domains 2 (Classroom Environment) and 3 (Instruction) to evaluate teachers. The modified FFT becomes problematic for the evaluation of special education teachers because the omitted domains are most relevant to teaching students with disabilities. For

example, Domain 1 (Planning and Preparation) contains features related to ensuring the use of developmentally appropriate materials for students who are functioning below grade level. In districts using the modified evaluation tool, these components of effective special education instruction would not be evaluated and would likely adversely impact the results for special education teachers (Jones & Brownell, 2013).

FFT is built on the constructivist approach to learning where students develop an understanding of new concepts through an interpretation of their existing cognitive structures and experiences. In this approach, teachers need to design instructional activities that enable students to construct their own knowledge rather than provide explicit instruction. However, research has shown that students with disabilities benefit from and need teachers to provide explicit instruction through modeling, clear explanations, and opportunities for practice (e.g., Bowman et al., 2019; Jones & Brownell, 2013) and, in fact, explicit instruction was recently declared by CEC to be one of the 22 high leverage practices for students with disabilities (McLeskey et al., 2017). The conflict between the type of instruction that is required for students with disabilities and the theoretical underpinnings of the FFT tool may result in a distorted representation of a special education teacher's effectiveness as teachers are rewarded for student-led instruction on the FFT and receive less effective ratings for teacher-led instruction (Jones & Brownell, 2013).

Evaluator's Knowledge of Special Education

Regardless of the evaluation instrument used, the accuracy of the results of a special education teacher's evaluation is heavily reliant upon the knowledge of the evaluator relative to special education (Lawson & Cruz, 2018). The majority of administrators/evaluators lack sufficient expertise or knowledge of special education to accurately evaluate the performance of special education teachers (Sledge & Pazey, 2013). Available research indicates 92% of building principals are not certified or licensed to teach special education; 45.9% of principals did not complete any special education classes as part of their administrator training program; 27.8% had competed one course in special education; and 59.4% of principals had attended two or fewer professional development trainings focused on special education in the past two years (Sledge & Pazey, 2013; Wakeman et al., 2006).

It is common for a special education teacher to have more knowledge relative to the learning and behavior characteristics of students with disabilities than the evaluator. The disparity in knowledge between special education teachers and their evaluators includes an understanding of evidence-based instructional practices for students with disabilities. The knowledge gap weakens the credibility of the evaluator to provide an accurate evaluation of the special education teacher. It also trivializes the quality of the evaluator's feedback and recommendations regarding professional growth (Glowacki & Hackman, 2016; Johnson & Semmelroth, 2014; Sledge & Pazey, 2013; Semmelroth & Johnson, 2014; Wakeman et al., 2006). A weak evaluation instrument in the hands of a poorly qualified evaluator can yield devastating impacts on both the school system and individual students.

Teachers of Students with Severe Disabilities

At present, very little literature explicitly focuses on teachers of students with severe disabilities. Most literature relative to special education teacher evaluation and effectiveness is focused on special education teachers in general, including those who teach students with high incidence disabilities who may access a curriculum more similar to that of their non-disabled peers with the assistance of accommodations, modifications, and specialized targeted instruction (Ruppar, et al., 2015). The lack of research in this area is a challenge for measuring the effectiveness of teachers of students with moderate to severe disabilities.

Some of the duties required of special education teachers of students with severe disabilities are required by all special education teachers. Those common duties include providing targeted, systematic instruction; adapting and modifying curriculum; providing support in the general education classroom; collaborating with other school professionals; supervising paraprofessionals; and completing case-management duties/paperwork (CEC, 2012; Crowe et al., 2017; Elliot et al., 2014 ; Ruppar et al., 2015; Williams & Dikes, 2015). However, teachers of students with severe disabilities often have roles/duties very dissimilar from other teachers. These roles/duties can include: providing instruction on activities of daily living; incorporating the instruction and use of AAC into daily classroom routines; managing and teaching replacement behaviors for persistent, challenging behavior; and managing students' significant health needs (Ruppar et al., 2017). These additional responsibilities and duties are not reflected and accounted for in current evaluation instruments and practices (Ruppar et al., 2015).

Proposed Solutions

Student Growth

At present, not much is known about the quality of academic growth measures for students with disabilities, particularly for students with severe disabilities. Additional research is needed to determine effective, accurate, and fair methods of measuring student growth for students with disabilities. Specific to students with severe disabilities, recommended research (e.g., Holdheide et al., 2012) includes: (1) an analysis of alternate assessment and student results to determine if the assessments are designed with sufficient sensitivity to measure student growth; (2) an analysis of alternate assessments to determine if they have the capacity to report growth scores; (3) research to determine if progress on IEP goals can reliably and validly be used to document student growth; (4) an analysis of portfolio reviews to determine if results can reliably and validly be used to document student growth; and (5) research to determine if student learning objectives can be used to evaluate special education teacher effectiveness. Holdheide et al. (2012) cautioned against the use of student growth in special education teacher evaluations when using the results for high-stakes decisions until further research can be completed to support the validity of claims from the measures.

Accommodate Differing Roles

The teacher evaluation process should take into account the specific roles of teachers and their teaching contexts (Holdheide et al., 2010). CEC (2012) and others (Holdheide et al., 2010; Johnson & Semmelroth, 2014; Ruppar et al., 2015; Semmelroth & Johnson, 2014) encouraged

the use of instruments and processes that have unique components that are more reflective of the roles of special education teachers. The unique components should allow for the evaluation of (a) the implementation of evidence-based practices, (2) student growth or achievement, and (c) flexibility to accommodate the variety of roles and contexts of the special education teacher.

Semmelroth and Johnson (2014) proposed the Recognizing Effective Special Education Teachers (RESET) observation tool to evaluate special education teacher effectiveness based on the teacher's use of evidence-based instructional practices for students with high-incidence learning disabilities. The tool contains descriptions and critical components of evidence-based practices that observers use to evaluate a special education teacher. However, there is difficulty establishing inter-rater reliability between observers, which diminishes the reliability and effectiveness of the tool in its current state.

The use of supplementary checklists or rubrics that can be used in conjunction with current evaluation instruments has been proposed (Sledge & Pazey, 2013). These checklists or rubrics would attempt to reflect the duties and practices of special education teachers. Elements of effective teachers of students with severe disabilities that have been identified in research (Ruppar et al., 2015; 2017) could be used in the creation of the supplementary rubrics or checklists.

Build Evaluator Knowledge of Special Education

The lack of evaluator knowledge of special education and evidence-based practices for students with disabilities has been consistently identified as a factor that heavily influences the results of a special education teacher's evaluation (Glowacki & Hackman, 2016; Johnson & Semmelroth, 2014; Joint Committee on Standards for Educational Evaluation, 2009; Semmelroth & Johnson, 2014; Sledge & Pazey, 2013; Wakeman et al., 2006). The majority of administrators/evaluators lack sufficient expertise or knowledge of special education to accurately evaluate the performance of special education teachers (Sledge & Pazey, 2013; Wakeman et al., 2006). Professional development for those who will administer teacher evaluations of special education teachers is urgently needed.

Options are currently available to increase an evaluator's/administrator's knowledge of special education. Online learning resources such as Autism Focused Intervention Resources & Modules (AFIRM; https://afirm.fpg.unc.edu/), Autism Internet Modules (AIM; https://autisminternetmodules.org/), Modules Addressing Special Education and Teacher Education (MAST; http://mast.ecu.edu/), and the IRIS Center (https://iris.peabody.vanderbilt.edu/) can be promoted by local district superintendents and special education directors, and professional organizations as a source of professional development. All are free resources that have learning modules that explain and provide examples of topics pertinent to special education and evidence-based practices that have been validated by research to be effective for students with disabilities. These learning modules are not a substitute for specific training necessary to implement these practices with fidelity. However, they can quickly help to build an administrator's knowledge about special education.

The addition of professional development relative to special education and evidence-based practices for students

with disabilities as part of training, workshops, and conferences provided by professional organizations for administrators could assist with increasing administrators' knowledge of special education. Institutes of higher education could consider adding special education coursework or a special education internship as part of administrator preparation programs. Additionally, states could add hours of training in special education or evidence-based practices for students with disabilities as a requirement for administrative license renewal.

Teachers of Students with Severe Disabilities

As stated previously, there is limited literature related to the evaluation of special education teachers. There is even less literature available that specifically examines teacher evaluations for special education teachers who teach students with severe disabilities (Ruppar et al., 2015; Holdheide et al., 2010). The literature available on special education teacher evaluation and effectiveness is focused on special education teachers in general, including those who teach students with mild disabilities who may access a curriculum more similar to that of their non-disabled peers with the assistance of accommodations, modifications, and specialized targeted instruction. The literature available for teachers of students with severe disabilities focuses primarily on teacher quality (experience, credentials, and self-efficacy), but those factors do not necessarily translate to improved teaching practices and outcomes for students with severe disabilities (Holdheide et al., 2010). A single examination of the appropriateness of using student results from alternate assessments as a factor in determining teacher effectiveness has been conducted

(Kearns et al., 2015). The general conclusion was that we do not yet have a way to validly and reliably use outcomes from alternate assessments to determine teacher effectiveness. Kearns et al. suggested a closer examination of the feasibility of two multi-state initiatives that have developed alternate assessments closely aligned to the Common Core State Standards (i.e., *National Center and State Collaborative* [NCSC] and *Dynamic Learning Maps* [DLM^{TM}]) as having the potential to provide one indicator of teacher effectiveness. In general, though, the current consensus is that no singular student outcome should be used as a defining factor in teacher effectiveness among those who serve students with severe disabilities. It is crucial that as a field, we determine what effective teaching looks like (Ruppar et al., 2015) and under what circumstances (Kearns et al., 2015; Gee & Gonsier-Gerdin, 2018), as well as how to measure "expert" skills among teachers of students with severe disabilities (Kearns et al., 2015; Ruppar et al., 2017).

Conclusion

The teacher evaluation process has existed since the 1700s. The intent of the evaluation process is to measure teacher quality and encourage professional growth through feedback (CEC, 2012; Darling-Hammond et al., 2012; Marzano, 2012a; Marzano, 2012b). However, as legislation has increased school accountability for student learning the teacher evaluation system has become more complex and challenging for all teachers, but even more so for special education teachers. The high-stakes decisions (e.g., promotion, retention, tenure, certification, termination, and salary) tied to teacher evaluation causes increased stress for special education teachers. Given the nature of their students' disabilities, they know they are

unlikely to meet the conventional expectations of the teacher evaluation designed for general education teachers (Holdheide et al., 2010; Semmelroth & Johnson, 2014). These high levels of stress and burnout from a difficult and demanding job increase the attrition rates of special education teachers. In a field that is already facing significant teacher shortages, changes in the evaluation system need to be made so we can retain special education teachers and increase their skills rather than further deter them from the profession (Irinaga-Bistolas et al., 2007).

There are significant barriers in the evaluation process for special education teachers in the areas of student growth, differing teacher roles, the lack of special education teacher consideration in the development of teacher evaluation instruments, and evaluators' limited knowledge of special education. Proposed solutions focus on continuing research on teacher evaluation models for use among teachers of students with disabilities (e.g.,

methods for showing student growth, particularly among students with severe disabilities; development and testing of evaluation tools specifically for special education teachers; and factors that improve teaching practices and outcomes for students with severe disabilities). In addition to research, work needs to be done to improve administrator/evaluator knowledge of special education and teaching practices that improve outcomes for students with disabilities.

With expanded research and improved teacher evaluation practices for special education teachers, the field can better support and retain special education teachers. By retaining special education teachers and helping them to grow professionally, we can ultimately improve the learning and quality-of-life outcomes for students with disabilities.

References

Education Commission of the States. (2018). *Policy snapshot teacher evaluations: What is the issue and why does it matter?* https://www.ecs.org/wp-content/uploads/Teacher_Evaluations.pdf

American Recovery and Reinvestment Act. (2009). Congress.gov https://www.congress.gov/bill/111th-congress/house-bill/1/text

Berry, A. (2012). The relationship of perceived support to satisfaction and commitment for special education teachers in rural areas. *Rural Special Education Quarterly, 31*(1), 3-14. https://doi.org/10.1177/875687051203100102

Betebenner, D. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice, 28*(4), 42-51. https://doi.org/10.1111/j.1745-3992.2009.00161.x

Billingsley, B. & Bettini, E. (2019). Special education teacher attrition and retention: A review of the literature. *Review of Educational Research, 89*(5), 697-744. https://doi.org/10.3102.0034654319862495

Bowman, J. A., McDonnell, J., Ryan, J. H., & Fudge-Coleman, O. (2019). Effective mathematics instruction for students with moderate and severe disabilities: A review of the literature. *Focus on Autism and Other Developmental Disabilities, 34*(4), 195-204. https://doi.org/10.1177/108835761982932

Cameto, R., Knokey, A., Nagle, K., Sanford, C., Blackorboy, J., Sinclair, B., Riley, D., & Lauer, K. (2009). *National profile on alternate assessments based on alternate achievement standards: A report from the national study on alternate assessments.* https://ies.ed.gov/ncser/pdf/20093014.pdf

Council for Exceptional Children. (2012). *The Council for Exceptional Children's position on special education teacher evaluation.* https://www.cec.sped.org/~/media/Files/Policy/CEC%20Professional%20Policies%20and%20Positions/Position_on_Special_Education_Teacher_Evaluation_Background.pdf

Croft, M., Guffy, G., & Vitale, D. (2018, July). The shrinking use of growth: Teacher evaluation legislation since ESSA. *ACT Research & Policy*, 1-15. https://www.act.org/content/dam/act/unsecured/documents/teacher-evaluation-legislation-since-essa.pdf

Crowe, C. C., Rivers, S. E., & Bertoli, M. C. (2017). Mind the gap: Accountability, observation and special education. *Assessment in Education: Principles, Policy & Practice, 24*(1), 21-43. https://doi.org/10.1080/0969594X.2015.1114913

Danielson, C. (2014). *The framework for teaching evaluation instrument*. Danielson Group https://www.danielsongroup.org/download/?download=448

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93*(6), 8-15. https://doi.org/10.1177/003172171209300603

Department of Education (DOE). (2009). Race to the top fund: Notice of proposed priorities, requirements, definitions, and selection criteria. Federal Register. https://www.govinfo.gov/content/pkg/FR-2009-07-29/pdf/E9-17909.pdf

Elliott, S. N., Roach, A. T., & Kurz, A. (2014). Evaluation and advancing the effective teaching of special educators with a dynamic instructional practices portfolio. *Assessment for Effective Intervention, 39(2)*, 83-98. https://doi.org/10.1177/1534508413511491

Every Student Succeeds Act, Public Law No. 114-95 § 114 Stat. 1177 (2015).

Glowacki, H., & Hackman, D. G. (2016). The effectiveness of special education teacher evaluation processes: Perspectives from elementary principals. *Planning and Changing, 47*(3/4), 191-209.

Hadden, S. D., & Mountz, A. E. (n.d.). *Recommendations for using the CLASS in inclusive early childhood programs, birth-age 8*. Teachstone http://info.teachstone.com/hubfs/recommendations%20for%20using%20the%20class.pdf?t=1523901786609

Holdheide, L., Browder, D., Warren, S., Budzick, H., & Jones, N. (2012). *Summary of "using student growth to evaluate educators of students with disabilities: Issues, challenges, and next steps"* (Teacher Quality Research & Policy Brief). National Comprehensive Center for Teacher Quality. https://www.iaase.org/Documents/Ctrl_Hyperlink/Session_13_D_Using_Student_Growth_Summary0112_uid9152013354352.pdf

Holdheide, L., Goe, L., Croft, A., & Reschly, D. (2010). *Challenges in evaluating special education teachers and English language learners specialists* (Teacher Quality Research & Policy Brief). National Comprehensive Center for Teacher Quality. ERIC https://files.eric.ed.gov/fulltext/ED520726.pdf

Howell, W. G. (2015). Results of President Obama's race to the top. *Education Next, 15*(4), 58-66.

Individuals with Disabilities Education Improvement Act, Pub. L. No. 108-446. (2004).

Irinaga-Bistolas, C., Schalock, M., Marvin, R., & Beck, L. (2007). Bridges to success: A developmental induction model for rural early career special educators. *Rural Special Education Quarterly, 26(1),* 13-22. https://doi.org/10.1177/875687050702600103

Johnson, E., & Semmelroth, C. L. (2014). Special education teacher evaluation: Why it matters, what makes it challenging, and how to address these challenges. *Assessment for Effective Instruction, 39*(2), 71-82. https://doi.org/10.117/1534508413513315

Joint Committee on Standards for Educational Evaluation. (2009). *The personnel evaluation standards: How to assess systems for evaluating educators* (2nd ed.). Corwin Press.

Jones, N. D., & Brownell, M. T. (2013). Examining the use of classroom observations in the evaluation of special education teachers. *Assessment for Effective Intervention, 39*(2), 112-124. https://doi.org/10.1177/1534508413514103

Kearns, J. F., Kleinert, H. L., Thurlow, M. L., Gong, B., & Quenemoen, R. (2015). Alternate assessments as one measure of teacher effectiveness: Implications for our field. *Research and Practice for Persons with Severe Disabilities, 40*(1), 20-35. https://doi.org/10.1177/1540796915558510.5

Lawson, J. E., & Cruz, R. A. (2018). Evaluating special educators' classroom performance: Does rater "type" matter? *Assessment for Effective Intervention, 43*(4), 227-240.

Learning Sciences Marzano Center. (2013, April). *Developing a passion for professional teaching: The Marzano teacher evaluation model*. https://www.learningsciences.com/wp/wp-content/uploads/2018/05/The-Marzano-Teacher-Evaluation-Model.pdf

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching.* ASCD.

Marzano, R. J. (2012a). Art and science of teaching/reducing error in teacher observation scores. *Educational Leadership, 70*(3), 82-83.

Marzano, R. J. (2012b). The two purposes of teacher evaluation. *Educational Leadership, 70*(3), 14-19.

Marzano, R. J. (2014). Proximal versus distal validity coefficients for teacher observational instruments. *Teacher Educator, 49*(2), 89-96. https://doi.org/10.1080/08878730.2014.885783

McLeskey, J., Barringer, M. D., Billingsley, B., Brownell, M., Jackson, D., Kennedy, M., & Ziegler, D. (2017). *High leverage practices in special education*. Council for Exceptional Children & CEEDAR Center.

McLeskey, J., & Billingsley, B. S. (2008). How does the quality and stability of the teaching force influence the research-to-practice gap? A perspective on the teacher shortage in special education. *Remedial and Special Education, 29*(5), 293-305. https://doi.org/10.1177/0741932507312010

Michigan Department of Education. (2016). Public Act 170, H.B. No. 4493 (2016 June 14, 2016). https://www.michigan.gov/mde/0,4615,7-140-5683_75438_78526---,00.html

Mills, J. I. (2008). A legislative overview of No Child Left Behind. *New Directions for Evaluation, 117*, 9–20. https://doi.org/10.1002/ev.248

No Child Left Behind Act, Public Law No. 107-110, 20 U.S.C. § 6319 (2001).

Noell, G. H., Brownell, M. T., Buzick, H. M., & Jones, N. D. (2014). *Using educator effectiveness measures to improve educator preparation programs and student outcomes* (Document No. LS-1). University of Florida, Collaboration for Effective Educator, Development, Accountability, and Reform Center https://ceedar.education.ufl.edu/wp-content/uploads/2014/09/LS-1_FINAL_08-27-14.pdf

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™: Manual K-3*. Paul H. Brookes.

Popham, W. J. (2013). Can classroom assessments of student growth be credibly used to evaluate teachers? *The English Journal, 103*(1), 34-39.

Ruppar, A., Roberts, C., & Olson, A. (2015). Faculty perceptions of expertise among teachers of students with severe disabilities. *Teacher Education and Special Education, 38*(3), 240-253. https://doi.org/10.1177/0888406414552331

Ruppar, A. L., Roberts, C. A., & Olson, A. J. (2017). Perceptions about expert teaching for students with severe disabilities among teachers identified as experts. *Research and Practice for Persons with Severe Disabilities, 42*(2), 121-135. https://doi.org/10.1177/1540796917697311

Ryan, S. V., von der Embse, N. P., Pendergast, L. L., Saeki, E., Segool, N., & Schwing, S. (2017). Leaving the teaching profession: The role of teacher stress and educational accountability policies on turnover intent. *Teaching and Teacher Education, 66*, 1-11. https://doi.org/10.1016/j.tate.2017.03.016

Sawchuk, S. (2015, September 3). *Teacher evaluation: An issue overview*. Education Week http://www.edweek.org/ew/section/multimedia/teacher-performance-evaluationissue-overview.html

Semmelroth, C. L., & Johnson, E. (2014). Measuring rater reliability on a special education observation tool. *Assessment for Effective Instruction, 39*(3), 131-145. https://doi.org/10.1177/1534508413511488

Semmelroth, C. L., Johnson, E., & Allred, K. (2013). Special education teacher evaluation: Cautions, concerns, and considerations.

Journal of American Academy of Special Education Professionals, 109-118. National Association of Special Education Teachers https://www.naset.org/index.php?id=3852

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*(3), 231-246. https://doi.org/10.1177/088840641348983 9

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*(3), 323-336. https://doi.org/10.1177/001440291452242 5

The IRIS Center. (2013). *Teacher retention: Reducing the attrition of special educators*. https://iris.peabody.vanderbilt.edu/module/ tchr-ret/

Thornton, B., Peltier, G., & Medina, R. (2007). Reducing the special education teacher shortage. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 80*(5), 233-238. https://doi.org/10.3200/TCHS.80.5.233-238

Towles-Reeves, E., Kleinert, H., & Muhomba, M. (2009) Alternate assessment: Have we learned anything new? *Exceptional Children, 75(2)*, 233-252. https://doi.org/10.1177/001440290907500 206

U.S. Department of Education. (2012). *ESEA flexibility: Review guidance*. Office of Elementary and Secondary Education. www.ed.gov/sites/default/files/review-guidance.doc

Vannest, K. J., & Hagan-Burke, S. (2010). Teacher time use in special education. *Remedial and Special Education, 31*(2), 126-142. https://doi.org/10.1177/074193250832745 9

Wakeman, S. Y., Browder, D. M., Flowers, C., & Ahlgrim-Delzell, L. (2006). Principals' knowledge of fundamental and current issues in special education. *NASSP Bulletin, 90*(2), 153-174. https://doi.org/10.1177/019263650628885 8

Williams, J., & Dikes, C. (2015). The implications of demographic variables as related to burnout among a sample of special education teachers. *Education, 135*(3), 337-345.

_____

Author Note

Correspondence concerning this article should be addressed to Robin A. Snyder, Supervisor of Autism Programs, Berrien RESA, 711 St. Joseph Ave., Berrien Springs, MI 49103. Email: Robin.snyder@berrienresa.org