

Special Educator Evaluation: Cautions, Concerns and Considerations

Carrie Semmelroth, M.A.

Evelyn Johnson, Ed.D.

Keith Allred, Ph.D.

Boise State University

Abstract

There is currently little consensus on how special education teachers should be evaluated in a way that is effective, fair and responsive to their unique teaching responsibilities. In this paper, we explain several of the current approaches to teacher evaluation under consideration, and then provide an overview of the challenges associated with the use of these models for special education teachers. We describe a model currently under development that is designed to better meet the unique characteristics of special education teacher evaluation. Our alternative approach proposes to evaluate special education teacher effectiveness through two primary components: observations of the special educator's use of research-based instructional practices, and the resulting student outcomes reported through effect sizes on measures aligned with relevant student goals.

Special Educator Evaluation: Cautions, Concerns and Considerations

The purpose of special education is to provide individualized instruction to meet the needs of a heterogeneous group of students with disabilities. Students served through special education often have the most intense instructional needs, and require specially designed instruction; meeting the needs of this group of students is extremely challenging and requires teachers who are highly skilled. Unfortunately however, students with disabilities are more often served by a special education teaching force that is highly subject to attrition and turnover (Billingsley, 2004; Boe, Cook, & Sunderland, 2008; Connelly & Graham, 2009). Additionally, special education is consistently indicated as a high demand field, with positions filled by teachers who lack adequate preparation to meet the demands of the job (Boe et al., 2008). These factors impact student outcomes – nationally, as few as 30% of students with disabilities are able to meet performance standards (Cortiella, 2011) and post-school outcomes for students with disabilities are not encouraging (U.S. Department of Education, 2011).

To improve the outcomes for students with disabilities, the instructional practice of special education teachers must be improved. Fortunately, the past three decades of special education research have produced a foundational body of knowledge on the use and application of evidence-based instructional practices. However, while arguably no other content area in education has produced more instructional practice research than

special education, the profession itself has made little progress in putting these techniques into practice. Improving special education teacher practice requires a systems-level change that includes providing stronger teacher preparation, improved working conditions, and evaluation systems that focus on measuring instructional practice and supporting teachers in performance improvement (Johnson & Semmelroth, in press). The focus of this paper is on the last component, designing evaluation systems for special education teachers that reliably identify those teachers who are effective, and identifying ways to support the professional development of those who are not, in order to improve student outcomes (Danielson, 2010; Johnson & Semmelroth, in press).

Value-Added Models (VAM): The current approach to teacher evaluation

Within the past three years, 32 states have changed their policies regarding teacher evaluation, and approximately 20 states and the District of Columbia now focus heavily on using student achievement as a primary component of their systems (National Center for Teacher Quality, 2011). The Race to the Top (RTT) state applications, in tandem with *A Blueprint for Reform: The Reauthorization of the Elementary and Secondary Act* (U.S. Department of Education, 2010), were the primary catalysts prompting this policy focus on defining teacher effectiveness through student performance. The most well-known approach for incorporating student outcomes as a primary feature of teacher effectiveness is the value-added model (VAM). VAMs define a relationship between teacher effectiveness and student academic achievement through weighted statistical formulas that incorporate values from a variety of measurements including teacher observation scores, student achievement scores, student/parent surveys, and other factors (Kane & Staiger, 2012). VAMs attempt to account for the multiple factors that may impact student achievement (Scherrer, 2012), and are thought to help answer the question of how effective an individual teacher is at promoting student growth. However, critics argue that VAMs suffer from numerous methodological and philosophical flaws (Newton, Darling-Hammond, Haertel, & Thomas, 2010), and do little to ensure teacher quality or promote professional development, as an effective evaluation system should (Danielson, 2010). This is especially true for special education.

As VAMs become more prolific in teacher evaluation systems, many questions surrounding their application to special education teachers are surfacing (Armario, 2012). These questions relate to issues of effectiveness (i.e. is it a useful way to measure special education teacher effectiveness?) and fairness (i.e. does it capture the salient features of effective special education teaching and the individualized nature of services and outcomes?).

What are value-added models (VAMs)?

The value-added model is defined as “a collection of complex statistical techniques that use multiple years of students’ test score data to estimate the effects of individual schools or teachers” (McCaffrey, Koretz, Lockwood, Hamilton, 2003, p.11). Because of the complexity of this technique, value-added modeling can appear in different forms. Value-added modeling generally refers to a class of models, also referred to as value-added

assessment, that use a range of assumptions to measure an individual teacher's effects on an individual student's performance and growth on standardized measures over time. The assumptions made within a particular VAM include whether teacher effects can be measured at the individual, school or district level, and whether student outcomes include only students the teacher directly instructs, or a more broadly defined group of students (McCaffrey, et. al, 2003). A teacher's ranking in a VAM system is dependent on whether students meet, exceed or fail predicted achievement on state assessments, and a teacher is considered to be *effective* if his/her students perform better than predicted on state assessments, and less effective if most students fail to make predicted gains.

One of the most common value-added approaches relies on vertically equated, developmental scales that measure the same constructs across all grade levels (Martineau, 2006). A vertically equated scale assumes that the teacher has had a constant effect on all students relative to other teachers in the system, which results in a measured effect that is an approximation of a teacher's average effect on students in the population that are likely to be in the teacher's class (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004). Critics question the validity of this approach because it assumes a teacher's impact is immediate, and of a pre-determined and fixed duration (Martineau, 2006).

Value-added measurement issues within special education

The assumptions made in VAM systems are especially concerning when examined through the lens of special education services. Special education teachers typically serve students across a range of grades and settings for varying amounts of time and for various purposes, and in other instances, some students with disabilities receive instruction from the same team of special education teachers and paraprofessionals for multiple years. Other students receive direct instruction provided by a paraprofessional who is supervised by a special education teacher. Some students with disabilities are not directly taught by a special education teacher, however, that teacher may provide consultant services to the general education teacher that impact that student's performance. These distinctions immediately complicate decisions regarding what percentage of student growth should be allocated to identifying particular teacher effects. Additionally, students receiving special education services may have their assessment data excluded from accountability formulas, or may participate in a non-standardized alternate assessment. Thus, the two most important components of a VAM, teacher effect and student performance, cannot universally be quantified in special education.

Challenges of teacher evaluation within special education

In addition to the issues with VAM above, there are several constraints that further complicate the development of a special education teacher evaluation model. The primary challenges include the lack of prepared special education teachers entering the field, the heterogeneity of the contexts and settings under which special education teachers work, the heterogeneity of the population they serve, and the individualized nature of determining appropriate student goals and learning trajectories. We briefly review these challenges below.

Lack of prepared special education teachers. Holdheide, Goe, Croft, and Reschly (2010) identified systemic challenges uniquely associated with special education teachers and evaluation systems, including: a) special education is a high demand field, with many positions either vacant or filled with unqualified personnel (Billingsley, Fall & Williams, 2006; Boe & Cook, 2006; McLeskey, Tyler & Flippin, 2004); b) special education teachers are typically not highly qualified in the core content areas they teach (McLeskey & Billingsley, 2008); c) special education teacher preparation programs do not often integrate the use of evidence-based practices, thus leaving new special education teachers ill-prepared to meet the challenges of the special education classroom (Reschly, Holdheide, Smart & Oliver, 2007; Walsh, Glaser & Wilson, 2007). These types of issues speak to the need for an evaluation system that focuses on the use of effective instructional practices and provides feedback to special education teachers so that they can work to improve. In the words of Darling Hammond (2011), “we can’t fire our way to [effective teaching]”, and should therefore consider approaches to evaluation that emphasize continuous improvement and professional development.

Heterogeneity of special education teaching contexts. As noted earlier, special education teachers operate within a variety of contexts and assume a variety of roles. Parsing out the amount of impact on the performance of students served under collaborative, inclusive, resource or extended resource models is guesswork at best. While some argue that a percentage can be allocated based on the time a student is served in various settings (e.g. 80% in general education, 20% in special education), the validity of this approach is questionable. Unless the idiosyncratic nature of special education service delivery is adequately addressed, significant psychometric issues (i.e., reliability of student achievement scores) could undermine the use of VAM in special education.

Determining appropriate student outcomes, goals and trajectories. In addition to the context variability, the students served in special education reflect a very heterogeneous population. Even when students present with similar needs, they may function at vastly different performance levels. Depending on their baseline performance, their opportunities to learn, and the severity of their disability, students with disabilities will experience very different growth rates and consequently, meet very different outcome targets. Of the three factors related to outcomes for students in special education: a) baseline performance, b) opportunities to learn, and c) severity of disability; the only factor over which a special education teacher has control is opportunities to learn. The special education teacher’s role is to be knowledgeable about the appropriate practice to meet the needs of that particular student, and to be able to design and implement an instructional plan that will support the academic, social and emotional needs of that student.

Using student outcomes to define special education teacher effectiveness requires first being able to identify 1) what kind of student growth *measure* to use and 2) how much student growth to expect. Growth rates for students with disabilities are typically not consistent, and there is evidence that suggests that students with very low initial performances often experience the least growth even when exposed to evidence-based

instruction (Coyne, McCoach, Loftus, Zipoli, Ruby, Crevecoeur, & Kapp, 2010; Wei, Blackorby, & Schiller, 2011). This suggests that models of teacher evaluation that rely on student outcome measures or on a standard growth rate metric (i.e. VAM) may not be valid for special education.

There are clear measurement challenges to addressing both of these issues. The first challenge, defining what kind of student growth to use, is confounded because of the heterogeneous populations typically served in special education. Even small groups of students typically present a significant spectrum of academic, social, and behavioral needs. For example, an extended resource room might serve students representing a range of disabilities including cognitive impairment, autism, behavioral disorders, and other health impairments. Two students might be placed in the classroom with the same exceptionality, e.g. cognitive impairment, but might vary widely in their academic, functional, communicative, and social interaction skills. This variation in student needs makes it difficult to select one student outcome measure that best “fits” a particular exceptionality, student group, or even classroom.

Even if one student outcome could be identified as addressing the needs of all students in a special education classroom, the next perplexing step is to define how much academic growth is considered adequate. Assuming all targeted growth across students to be linear and consistent as represented by specific points on a vertical scale is naive. Differentiation in special education is based upon the notion that each student will achieve academic, social and behavioral growth at their particular pace depending upon factors typically beyond the control of the teacher.

Noting the twin challenges of determining 1) what student outcome measure is most appropriate, and 2) how much student growth is considered adequate for students with disabilities, illustrates the disparity between the necessary and sufficient conditions required by a teacher evaluation model such as VAM and some of the realities of special education. Given the unanswered measurement questions regarding how to define special education teacher effects and student performance, how can special educator teachers be evaluated fairly and effectively? Below are our considerations and suggestions for an alternative approach.

Examining effective instructional practices and student response to instruction

To summarize the discussion thus far, the two primary components of VAM: a) teacher effect and b) student outcomes, pose unique challenges within the field of special education that limit the validity of value-added models as a fair and effective special education teacher evaluation approach. Additionally, macro-level challenges in special education teacher training, recruitment and retention have resulted in a high percentage of underprepared special education teachers working in a challenging field that has been identified as a critical shortage area in many states.

These considerations require an alternative means of evaluating special education teacher effectiveness that focuses on increasing the use of evidence-based practices for students

with disabilities, and through the use of effective instruction, improving student outcomes. In their article discussing what hinders the effectiveness of special education, Heward & Ohio (2003) note the biggest reason we do not teach more children with disabilities better than we do is “not because we do not know enough but because we do not teach them as well as we know how” (p. 201). Considerations of fair and effective special education teacher evaluation systems must be based on the systematic measurement of the implementation of evidence-based practices to support the needs of students with disabilities (Johnson & Semmelroth, in press). Additionally, the measurement of student outcomes as related to the use of research-based practices must be included, and must also be flexible enough to capture the diverse needs of the heterogeneous special education population (Johnson & Semmelroth, in press). Improving outcomes for students with disabilities is the central purpose of such a system (Holdheide, 2012).

Therefore, we propose the following approach to evaluating special education teachers and offer some of the preliminary findings of our pilot work in its development. An effective special education teacher evaluation system that will lead to improved teaching practice and to improved outcomes for students with disabilities, is one that will: 1) reliably discriminate between effective and ineffective special education teachers, 2) measure and provide targeted, specific, corrective feedback for teacher instructional practice, 3) include the use of individualized student growth rates to define teacher effectiveness, and 4) be responsive to the variety of contexts in which special education teachers work. Over the last two years, we have worked on the development of a system grounded in these four principles, called the Recognizing Effective Special Education Teachers (RESET) observation tool.

The RESET observation tool is designed to evaluate instructional practice, provide feedback to special education teachers about the quality of their instruction and ultimately, improve the outcomes for students with disabilities (Johnson & Semmelroth, 2011). RESET is a computerized evaluation system that relies on the use of video capture of instruction which is then evaluated by a trained observer, using clearly specified criteria that align with the research-identified characteristics of effective instruction for students with disabilities (Johnson & Semmelroth, in press). Special education teachers evaluated under this system receive feedback on the specific dimensions of their teaching according to criteria derived from the research on effective instruction for students with disabilities. Additionally, individualized student growth measures are included as an indication of the special education teacher’s effectiveness. Much of the work is preliminary, and below we describe the current status of RESET’s development and validation studies.

Evaluating effective instruction. To evaluate instructional practice, we have created scoring criteria for several evidence-based instructional practices. The process of identifying evidence based practices began with current published reviews of effective practice, such as those published in the special issue on Evidence-Based Practices in Special Education (see *Exceptional Children*, 2009). Identification of other evidence-based practices was patterned on the review process described by Chard, Ketterlin-Geller,

Baker, Doabler, & Apichatabutra (2009). A significant portion of the review of evidence-based practices has been conducted to inform the pilot development of RESET (Johnson & Semmelroth, in press; Johnson, Semmelroth, & Beymer, 2012). From this review, the characteristics of effective instructional practice are specified to create the items used to evaluate a special education teacher's practice. Once these characteristics are defined by each evidence-based instructional practice, an evaluation rubric is created and used to assign scores to the observation of special education teachers delivering instruction.

Reliability. Our initial reliability studies examining the extent to which two independent raters can agree on evaluating a video capture of specific instructional practices are encouraging. We were able to achieve correlations in the moderate range across several of our criteria in pilot studies examining inter-rater reliability (Johnson & Semmelroth, in press), and in a more recent pilot study using revised rubrics, achieved correlations in the moderate to large range. Our next steps include further work to improve the reliability coefficients through more clearly defined criteria, and improved training for evaluators. Additionally, we are continuing to expand the range of evidence-based practices and related scoring criteria so that the RESET tool will be appropriate for use across more instructional contexts and settings.

Validity. Measures of student growth will be an integral component of RESET. For each of the evidence-based practices identified, a corresponding range of effect sizes reported in the research is noted. As we collect data on instructional practices, we are also collecting student growth data from participating special education teachers with the intent of determining whether special education teachers who implement evidence based practices with fidelity are able to report growth levels consistent with those reported in the research. We anticipate that high levels of fidelity of implementation of an instructional practice should correspond with high levels of student growth. Because we are using a measure of effect size, we are able to evaluate data across multiple measures, which addresses the need for a consistent yet flexible indicator of growth for students with disabilities.

Finally, we are also collecting data to examine change in teacher performance over time. To accomplish this, we are conducting a study in which special education teachers are randomly assigned to a treatment or control group. In the treatment group, teachers will have their instruction evaluated using the RESET tool, and will be provided the results of their evaluation. Teachers assigned to the control group will be evaluated, but results will not be shared. Then, improvement over time will be examined to determine the extent to which the feedback from the observations impacts teaching performance. This type of data will provide important information on the extent to which RESET acts as a means of improving instructional practice.

Conclusion

While there is general consensus that teacher evaluation systems are an important component of improving instructional practice, there is little consensus on how best to design a system that is fair and effective for special education teachers. Special education

poses unique challenges to teacher evaluation that current approaches, such as VAM, do not adequately address. In order to address the challenges of special education teacher evaluation, we must consider an evaluation tool that can 1) reliably discriminate between effective and ineffective special education teachers, 2) measure and provide targeted, specific, corrective feedback for teacher instructional practice, and 3) include the use of individualized student growth rates to define teacher effectiveness. We recognize that ongoing research is necessary to refine the RESET tool. Ultimately, though, tools such those developed through RESET and similar initiatives that focus on instructional practice, may be a primary means of helping students with disabilities maximize their potential because special education teachers are being supported in reaching their full potential.

References

- Armario, C. (2012, April 24). *Special education teacher evaluation methods a struggle*. *Huffington Post*. Retrieved from: http://www.huffingtonpost.com/2012/04/24/struggle-over-how-to-eval_0_n_1449418.html
- Billingsley, B. S. (2004). Special education teacher retention and attrition: A critical analysis of the research literature. *The Journal of Special Education, 38*, 39–55. doi:10.1177/00224669040380010401
- Boe, Erling, E., Cook, Lynne, H., & Sunderland, Robert, J. (2008). Teacher turnover: Examining exit attrition, teaching area transfer, and school migration. *Exceptional Children, 75*, 7–31.
- Connelly, V., & Graham, S. (2009). Student teaching and teacher attrition in special education. *Teacher Education and Special Education, 32*, 257–269. doi:10.1177/0888406409339472
- Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Ruby, M., Crevecoeur, Y. C., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness, 3*, 93–120. doi:10.1080/19345741003592410
- Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership, 68*, 35–39.
- Greenberg, J., Pomerance, L., & Walsh, K. (2011). *Student teaching in the United States*. Retrieved from http://www.nctq.org/edschoolreports/studentteaching/docs/nctq_str_full_report_final.pdf
- Heward, W. L. (2003). Ten faulty notions about teaching and learning that hinder the effectiveness of special education. *The Journal of Special Education, 36*, 186–205.
- Holdheide, L. (2012). *State considerations in designing and implementing evaluation systems that include teachers of students with disabilities*. Office of Special Education Programs Project Director's Conference. Washington DC.
- Holdheide, L., Goe, L., Croft, A., & Reschly, D. J. (2010). *Challenges in evaluating special education teachers and English language learner specialists* (pp. 1–40). Washington DC.

- Johnson, E., & Semmelroth, C. (in press). Examining interrater agreement analyses of a pilot special education observation tool. *Journal of Special Education Apprenticeship*.
- Johnson, E., & Semmelroth, C. (2011). *Special Education Teacher Evaluation: Issues, Obstacles, Potential Directions*, Boise, ID.
- Johnson, E., Semmelroth, C., & Beymer, L. (2012). *A review of special education evidence-based instructional practices*. Boise, ID.
- Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* (pp. 1–68). Retrieved from http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf
- Martineau, J. A. (2006). Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics*, 31, 35–62.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29, 67–101. doi:10.3102/10769986029001067
- 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, 293–305. doi:10.1177/0741932507312010
- McLeskey, J., Tyler, N. C., & Saunders Flippin, S. (2004). The supply of and demand for special education teachers : A review of research regarding the chronic shortage of special education teachers. *The Journal of Special Education*, 38, 5–21.
- National Council on Teacher Quality. (2011). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies*. Washington DC.
- Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, 18(23).
- Reschly, D. J., Holdheide, L. R., Smartt, S. M., & Oliver, R. M. (2008). *Evaluation of LBS-1 Teacher Preparation in Inclusive Practices, Reading, and Classroom Organization-Behavior Management*. Submitted to the Illinois State Board of Education, June 2008.
- Scherrer, J. (2012). What's the value of VAM (value-added modeling)? *Phi Delta Kappan*, 93, 58–61.
- U.S. Department of Education, Institute for Education Sciences, National Center for Special Education Research, National Longitudinal Transition Study-2. (2011). *The Post-High School Outcomes of Young Adults With Disabilities up to 8 Years After High School: A Report From the National Longitudinal Transition Study-2 (NLTS2)*. Retrieved from <http://ies.ed.gov/ncser/pubs/20113005/>
- UCLA/IDEA. (2011). "An Education Exchange with Linda Darling-Hammond," UCLA/IDEA Newsroom (April 1).
- Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in reading achievement of students with disabilities, ages 7 to 17. *Exceptional Children*, 78, 89–106.

About the Authors

Carrie L. Semmelroth, M.A. is a Project Coordinator in the Department of Special Education and Early Childhood Studies at Boise State University. Her research interests include special education teacher evaluation, and improving support for special education teachers in the classroom.

Evelyn S. Johnson, Ed.D. is a Professor of Special Education in the College of Education at Boise State University. Her research interests include identification and interventions for students with learning disabilities, special education teacher preparation and special education teacher evaluation.

Keith W. Allred, Ph.D. is an Associate Professor and Chair of the Department of Special Education and Early Childhood Studies at Boise State University. His research interests include how to effectively facilitate inclusion of students with disabilities and factors that contribute to effective parent-professional relations in education.