

CEL Rater Reliability Research Overview: Recommendations for Achieving Increased Reliability in Classroom Observations for Teacher Evaluation

Challenge	Recommendation/Possible Solution
1. Aligning beliefs/vision among system stakeholders and evaluators can affect reliability (and, perhaps as importantly, perceptions of reliability)	<ul style="list-style-type: none"> Address vision and beliefs: all evaluators must be operating under the same vision of instruction in order to yield reliable results. Consider holding joint training for administrators and teachers: this lessens any tendency for suspicion of motives and provides an opportunity for evaluators and teachers to interact on the definition of teaching that underlies the performance standards before the actual evaluation process takes place.¹
2. Ensuring classroom observation framework or instrument is associated with reliable observation outcomes	<ul style="list-style-type: none"> Observation instrument should separate observation, interpretation and judgment; or distinguish between bias, interpretation and evidence.² Observation instrument should recognize teaching as complex, rather than simply providing a checklist of behaviors.³
3. Ensuring classroom observation framework or instrument is associated with reliable outcomes; using the same tool effectively for both summative and formative purposes	<ul style="list-style-type: none"> Consider whether the purpose of classroom observations is formative or summative evaluation. Conduct observations differently for different purposes.⁴ Differentiate feedback for different contexts.⁵ For formative feedback, conduct observations on two or more occasions within a time period; for example: on two consecutive days, have a professional development session, followed by two additional consecutive days of observations.⁶
4. Selecting qualified evaluators	<ul style="list-style-type: none"> Raters should have standards knowledge, content knowledge, and pedagogical knowledge, especially when conducting high-inference observations or when high-stakes decisions are involved. Consider using formal peer evaluators in these cases.⁷ Invest in systematic, high-quality training of observers.⁸ Include discussion and analysis of the instrument, including underlying beliefs, in training.⁹ Review and analyze multiple video examples of teaching for each competency at each performance level.¹⁰ Practice scoring videos, with explicit, concrete, personalized feedback from trainers.¹¹ Specifically address issues of rater bias (severity/leniency, halo effect, etc.).¹² Provide ample opportunities for interactions between facilitator and trainees during the training. Include question and answer sessions after each stage.¹³ Provide practice in making objective statements about teaching using observational framework.¹⁴ Use cooperative groups during training sessions to increase the level of objectivity among raters. Evaluators complete ratings independently, but discuss them within the group, listening to perspectives of other group members.¹⁵

<p>5. Minimizing rater bias; addressing interrater reliability</p>	<ul style="list-style-type: none"> • Use multiple raters (not just one observer) per teacher.¹⁶ • Rotate raters across teachers; randomly assign raters to classrooms.¹⁷ • Conduct multiple observations, especially when the stakes are high.¹⁸ • Ideal number for strong reliability is 4-6 observations per year.¹⁹ • The use of multiple evaluators increases reliability, especially if at least one of the evaluators is a professional peer rather than another administrator.²⁰ • Conduct system-level audits using impartial observers or oversight panels.²¹ • Consider a “hold harmless” period when participating schools begin making observations, practicing feedback and scoring teacher practice.²²
<p>6. Establishing evaluator accountability</p>	<ul style="list-style-type: none"> • Evaluators should formally demonstrate competency in using the evaluation tool accurately – using some kind of certification process and/or reliability standard – before conducting classroom observations. Competency should be judged in comparison with “expert raters.”²³ • If raters fail after multiple attempts, they should be declared ineligible to rate.²⁴ • Because the fail rate is somewhat predictable, train 10-15% more observers than you will need.²⁵ • Observers must pass a calibration exercise before they participate in rating on any particular day.²⁶ • Evaluators take periodic drift tests to ensure reliability.²⁷ • Require annual recertification of raters.²⁸
<p>7. Maintaining quality control and system oversight</p>	<ul style="list-style-type: none"> • Conduct ongoing training for evaluators. Include interrater reliability activities at least monthly.²⁹ • Use scoring leaders, or leadership teams, who manage a group of scorers and work more closely with discrepant scorers.³⁰ • Use double scoring for quality control (two raters scoring the same lesson).³¹ • Document everything to provide evidence as to extent of reliability. Expect to make changes after the first year of implementation.³² • Make use of a performance management system to track data and document feedback (for example, the eVAL tool provided by Washington state’s Office of Superintendent of Public Instruction).³³
<p>8. Addressing implementation issues</p>	<ul style="list-style-type: none"> • Base results on a single competency model in order to preserve alignment of the system.³⁴ • Stick to the system (fidelity of implementation).³⁵ • To make sure observations aren’t a function of time of year, observe on multiple consecutive days (for example, two days).³⁶ • For best results, schedule observations after the first 25 minutes of the school day.³⁷ • Evaluators should score away from distractions.³⁸ • Consider a combination of preannounced and unannounced visits (pre-scheduling can impede reliability).³⁹ • Average teachers’ scores over multiple observations (don’t use the results of a single observation).⁴⁰

Footnotes

1. Aspen Institute (2011); Curtis (2011); Danielson (2011); Donaldson (2009); Donaldson et al. (2010); Henry (2010); Kimball et al. (2001); Marzano (2011); Sartain et al. (2011)
2. Danielson (2011); Himmelein (2009); Humphrey et al. (2011); Marzano (2011); Measures of Effective Teaching (MET) Project (2012); Sartain et al. (2011)
3. Danielson (2011); Donaldson (2009); Sartain et al. (2011)
4. Donaldson et al. (2010); Fry et al. (2011); Goldstein (2007); Humphrey et al. (2011); Meyer et al. (2011); Muijs (2006)
5. Donaldson et al. (2010); Himmelein (2009)
6. Meyer et al. (2011)
7. Appeldoorn (2004); Donaldson (2009); Donaldson et al. (2010); Goe et al. (2011); Goldstein (2007); Hamre et al. (2009); Henry (2010); Kimball et al. (2001); Learning Sciences International (2011); Matsumura et al. (2008); Milanowski et al. (2011)
8. Appeldoorn (2004); Aspen Institute (2011); Bell et al. (2009); Curtis (2011); Curtis et al. (2012); Daley et al. (2010); Danielson (2011); Donaldson (2009); Donaldson et al. (2010); Fry et al. (2011); Goe et al. (2008); Goe et al. (2011); Goldstein (2007); Hamre et al. (2009); Henry (2010); Himmelein (2009); Humphrey et al. (2011); Kimball et al. (2001); Kimball et al. (2004); Kimball et al. (2009); Learning Sciences International (2011); Matsumura et al. (2008); McCaslin et al. (2006); Measures of Effective Teaching (MET) Project (2012); Milanowski (2011); Milanowski et al. (2011); Muijs (2006); National Council on Teacher Quality (2011); The New Teacher Project (2011); Nichols (2007); Odden (2004); Raudenbush et al. (2008); Sartain et al. (2011); Sawada et al. (2002); Schacter et al. (2005); Sterbinsky et al. (2003); Stuhlman et al. (2010); Stumbo et al. (2011); Weisberg et al. (2009)
9. Appeldoorn (2004); Curtis (2011); Danielson (2011); Fry et al. (2011); Goldstein (2007); Hamre et al. (2009); Henry (2010); Kimball et al. (2001); Learning Sciences International (2011); Measures of Effective Teaching (MET) Project (2012); Muijs (2006); Nichols (2007); Sartain et al. (2011); Sawada et al. (2002)
10. Curtis (2011); Curtis et al. (2012); Daley et al. (2010); Donaldson et al. (2010); Fry et al. (2011); Goe et al. (2011); Hamre et al. (2009); Kimball et al. (2001); Learning Sciences International (2011); McCaslin et al. (2006); Measures of Effective Teaching (MET) Project (2012); Sartain et al. (2011); Sawada et al. (2002)
11. Curtis (2011); Curtis et al. (2012); Daley et al. (2010); Danielson (2011); Donaldson et al. (2010); Fry et al. (2011); Hamre et al. (2009); Kimball et al. (2001); Learning Sciences International (2011); Matsumura et al. (2008); McCaslin et al. (2006); Measures of Effective Teaching (MET) Project (2012); Milanowski et al. (2011); Sartain et al. (2011); Sawada et al. (2002)
12. Bell et al. (2009); Curtis (2011); Donaldson (2009); Henry (2010); Measures of Effective Teaching (MET) Project (2012); Muijs (2006); Nichols (2007); Sartain et al. (2011)
13. Appeldoorn (2004); Curtis (2011); Danielson (2011); Fry et al. (2011); Kimball et al. (2001); Matsumura et al. (2008); Measures of Effective Teaching (MET) Project (2012); Nichols (2007); Sawada et al. (2002)
14. Curtis (2011); Humphrey et al. (2011); Measures of Effective Teaching (MET) Project (2012); Sartain et al. (2011); Weisberg et al. (2009)

15. Appeldoorn (2004); Curtis (2011); Danielson (2011); Fry et al. (2011); Goldstein (2007); Humphrey et al. (2011); McCaslin et al. (2006); Sawada et al. (2002)
16. Bell et al. (2009); Curtis (2011); Curtis et al. (2012); Donaldson (2009); Donaldson et al. (2010); Fry et al. (2011); Hamre et al. (2009); Himmelein (2009); Humphrey et al. (2011); McCaslin et al. (2006); Measures of Effective Teaching (MET) Project (2012); Milanowski (2011); Milanowski et al. (2011); National Council on Teacher Quality (2011); Raudenbush et al. (2008); Sartain et al. (2011); Sawada et al. (2002); Shin et al. (2011); Sterbinsky et al. (2003); Stuhlman et al. (2010); Tyler (2011)
17. McCaslin et al. (2006); Sterbinsky et al. (2003); Stuhlman et al. (2010)
18. Appeldoorn (2004); Aspen Institute (2011); Bell et al. (2009); Curby et al. (2011); Curtis (2011); Curtis et al. (2012); Daley et al. (2010); Donaldson (2009); Donaldson et al. (2010); Goe et al. (2008); Goldstein (2007); Goldstein et al. (2006); Himmelein (2009); Humphrey et al. (2011); Matsumura et al. (2008); Measures of Effective Teaching (MET) Project (2012); Meyer et al. (2011); Milanowski (2011); Milanowski et al. (2011); National Council on Teacher Quality (2011); Odden (2004); Raudenbush et al. (2008); Schacter et al. (2005); Sterbinsky et al. (2003); Stuhlman et al. (2010); Tyler (2011)
19. Bell et al. (2009); Curtis (2011); Curtis et al. (2012); Daley et al. (2010); Donaldson (2009); Goe et al. (2008); Measures of Effective Teaching (MET) Project (2012); Schacter et al. (2005); Tyler (2011)
20. Aspen Institute (2011); Curtis (2011); Curtis et al. (2012); Daley et al. (2010); Donaldson (2009); Goe et al. (2008); Goe et al. (2011); Goldstein (2007); Goldstein et al. (2006); Himmelein (2009); Humphrey et al. (2011); Measures of Effective Teaching (MET) Project (2012); Milanowski et al. (2011); Sartain et al. (2011); Schacter et al. (2005)
21. Curtis (2011); Curtis et al. (2012); Daley et al. (2010); Donaldson (2009); Donaldson et al. (2010); Fry et al. (2011); Goldstein (2007); Goldstein et al. (2006); Humphrey et al. (2011); Measures of Effective Teaching (MET) Project (2012); Milanowski et al. (2011); National Council on Teacher Quality (2011); Sartain et al. (2011); Sterbinsky et al. (2003); The New Teacher Project (2011); Weisberg et al. (2009)
22. Fry et al. (2011); Learning Sciences International (2011); National Council on Teacher Quality (2011)
23. Appeldoorn (2004); Bell et al. (2009); Curtis et al. (2012); Daley et al. (2010); Donaldson (2009); Fry et al. (2011); Goe et al. (2008); Hamre et al. (2007); Hamre et al. (2009); Henry (2010); Learning Sciences International (2011); Matsumura et al. (2008); McCaslin et al. (2006); Measures of Effective Teaching (MET) Project (2012); Milanowski et al. (2011); Muijs (2006); National Council on Teacher Quality (2011); Nichols (2007); Schacter et al. (2005); Stuhlman et al. (2010); The New Teacher Project (2011); Weisberg et al. (2009)
24. Bell et al. (2009); Hamre et al. (2009); Henry (2010); Matsumura et al. (2008); Measures of Effective Teaching (MET) Project (2012); Milanowski et al. (2011)
25. Hamre et al. (2009); Henry (2010)
26. Measures of Effective Teaching (MET) Project (2012)
27. Bell et al. (2009); Daley et al. (2010); Donaldson (2009); Fry et al. (2011); Goe et al. (2008); Hamre et al. (2009); Himmelein (2009); McCaslin et al. (2006); Milanowski et al. (2011); Nichols (2007); Schacter et al. (2005); Stuhlman et al. (2010); Weisberg et al. (2009)
28. Curtis et al. (2012); Daley et al. (2010); Hamre et al. (2009); Milanowski et al. (2011); Schacter et al. (2005)
29. Curtis et al. (2012); Daley et al. (2010); Donaldson (2009); Hamre et al. (2009); Goldstein (2007)

30. Daley et al. (2010); Goldstein (2007); Measures of Effective Teaching (MET) Project (2012); Milanowski et al. (2011)
31. Bell et al. (2009); Hamre et al. (2009); Measures of Effective Teaching (MET) Project (2012); Milanowski et al. (2011); Sartain et al. (2011); Sawada et al. (2002); Tyler (2011)
32. Aspen Institute (2011); Curtis (2011); Curtis et al. (2012); Humphrey et al. (2011); Measures of Effective Teaching (MET) Project (2012); Sartain et al. (2011); The New Teacher Project (2011)
33. Curtis (2011); Curtis et al. (2012); Learning Sciences International (2011); Sartain et al. (2011)
34. Aspen Institute (2011); Curtis (2011); Milanowski et al. (2011); Sartain et al. (2011)
35. Curtis (2011); Curtis et al. (2012); Goe et al. (2011); Kimball et al. (2001); Milanowski et al. (2011); Pianta et al. (2009); Schacter et al. (2005); The New Teacher Project (2011)
36. Curby et al. (2011); Matsumura et al. (2008); Sterbinsky et al. (2003)
37. Curby et al. (2011)
38. Milanowski et al. (2011)
39. Appeldoorn (2004); Curtis (2011); Curtis et al. (2012); Daley et al. (2010); Donaldson (2009); Goe et al. (2008); Goldstein (2007); Sartain et al. (2011); Schacter et al. (2005); Tyler (2011);
40. Aspen Institute (2011); Curtis (2011); Goe et al. (2008); Measures of Effective Teaching (MET) Project (2012); Meyer et al. (2011); Milanowski et al. (2011); Stuhlman et al. (2010)

References

- Appeldoorn, K. L. (2004). Developing and validating the collaboratives for excellence in teacher preparation (CETP) core evaluation classroom observation protocol (C.O.P.). (Doctoral dissertation). Retrieved from ProQuest (3154027).
- Aspen Institute. (2011). Building teacher evaluation systems: Learning from leading efforts. Retrieved from website: http://www.aspeninstitute.org/sites/default/files/content/docs/education/AI_Perf%20Mgmt_Synthesis.pdf
- Bell, C. A., Little, O. M., Croft, A. J., & Gitomer, D. H. (2009). *Measuring teaching practice: A conceptual review*. San Diego, CA: American Educational Research Association.
- Curby, T. W., Stuhlman, M., Grimm, K., Mashburn, A., Chomat-Mooney, L., Downer, J., Hamre, B., & Pianta, R. C. (2011). Within-day variability in the quality of classroom interactions during third and fifth grade: Implications for children's experiences and conducting classroom observations. *The Elementary School Journal*, 112(1), 16-37.
- Curtis, R. (2011). *District of Columbia public schools: Defining instructional expectations and aligning accountability and support*. Retrieved from Aspen Institute website: http://www.aspeninstitute.org/sites/default/files/content/docs/education/AI_DCPS_teacher%20evaluation.pdf
- Curtis, R., & Wiener, R. (2012). *Means to an end: A guide to developing teacher evaluation systems that support growth and development*. Retrieved from Aspen Institute website: <http://www.aspeninstitute.org/publications/means-end-guide-developing-teacher-evaluation-systems-support-growth-development>
- Daley, G., & Kim, L. (2010). *A teacher evaluation system that works* (Working paper). Retrieved from National Institute for Excellence in Teaching website: http://www.tapsystem.org/publications/wp_eval.pdf
- Danielson, C. (2011). Evaluations that help teachers learn. *Educational Leadership*, 68(4), 35-39.
- Donaldson, M. L. (2009). *So long, Lake Wobegon? Using teacher evaluation to raise teacher quality*. Retrieved from Center for American Progress website: www.americanprogress.org/issues/2009/06/pdf/teacher_evaluation.pdf
- Donaldson, M. L., & Peske, H. (2010). Supporting effective teaching through teacher evaluation: A study of teacher evaluation in five charter schools. Retrieved from Center for American Progress website: http://www.americanprogress.org/wp-content/uploads/issues/2010/03/pdf/teacher_evaluation.pdf
- Fry, R., & Ramsdell, R. (2011). *Enhancing teacher evaluation systems through effective classroom observation and tripod student surveys*. [PowerPoint slides]. Retrieved from Connecticut Association of Public School Superintendents website: <http://www.capss.org/page.cfm?p=439>
- Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from <http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>

Goe, L., Holdheide, L., & Miller, T. (2011). *A practical guide to designing comprehensive teacher evaluation systems*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from <http://www.tqsource.org/publications/practicalGuideEvalSystems.pdf>

Goldstein, J. (2007). Easy to dance to: Solving the problems of teacher evaluation with peer assistance and review. *American Journal of Education*, 113(3), 479–508.

Goldstein, J., & Noguera, P. A. (2006). A thoughtful approach to teacher evaluation. *Educational Leadership*, 63(6), 31-37.

Hamre, B. K., Goffin, S. G., & Kraft-Sayre, M. (2009). *Classroom assessment scoring system (CLASS) implementation guide*. Retrieved from Teachstone website: <http://www.teachstone.org/about-the-class/>

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007). Building a science of classrooms: Application of the CLASS framework in over 4,000 U.S. early childhood and elementary classrooms. Retrieved from Foundation for Child Development website: <http://fcd-us.org/sites/default/files/BuildingAScienceOfClassroomsPiantaHamre.pdf>

Henry, A. E. (2010). *Advantages to and challenges of using ratings of observed teacher-child interactions*. (Doctoral dissertation). Retrieved from ProQuest (3462227).

Himmelein, M.E. (2009). *An investigation of principals' attitudes toward teacher evaluation processes*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3399151)

Humphrey, D. C., Koppich, J. E., Bland, J. A., & Bosetti, K. R. (2011). *Peer review: Getting serious about teacher support and evaluation*. Menlo Park, CA: SRI International. Retrieved from Connecticut Association of Public School Superintendents website: http://www.capss.org/uploaded/Hard_Copy_Documents/Teacher_Evaluation_That_Works/PAR_PeerReviewReport_2011.pdf

Kimball, S., & Kelley, C.J. (2001). *Innovations in teacher evaluation: Case studies of two school districts with teacher evaluation systems based on the framework for teaching*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3033269)

Kimball, S. M., & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe county. *Peabody Journal of Education*, 79(4), 54-78.

Learning Sciences International (2011). Proposal for the Southeastern Wisconsin Teacher Evaluation Consortium: Technical assistance, Marzano Causal Teacher Evaluation Model and Evaluation Management System. Blairsville, PA. Retrieved from Teacher Evaluation Consortium Wiki website: <http://teachevalconsort.wikispaces.com/file/view/Learning+Sciences+International.pdf>

Marzano, R. G. (2011). The Marzano teacher evaluation model: Overview and resources. [PowerPoint slides]. Retrieved from Connecticut Association of Public School Superintendents website: <http://www.capss.org/page.cfm?p=439>

Matsumura, L. C., Garnier, H. E., Slater, S. C., & Boston, M. D. (2008). Toward measuring instructional interactions "at-scale." *Educational Assessment*, 13(4), 267-300.

McCaslin, M., Good, T. L., Nichols, S., Zhang, J., Wiley, C. R. H., Bozack, A. R., ... Cuizon-Garcia, R. (2006). Comprehensive school reform: An observational study of teaching in grades 3 through 5. *The Elementary School Journal*, 106(4), 313-331.

Measures of Effective Teaching (MET) Project (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation (2012). Retrieved from website: <http://www.metproject.org/reports.php>

Meyer, J. P., Cash, A. H., & Mashburn, A. (2011). Occasions and the reliability of classroom observations: Alternative conceptualizations and methods of analysis. *Educational Assessment*, 16(4), 227-243.

Milanowski, A. T. (2011, April 10). Validity research on teacher evaluation systems based on the framework for teaching. Paper presented at the American Educational Research Association Annual Meeting, New Orleans, LA. Retrieved from Education Resources Information Center (ERIC # ED520519)

Milanowski, A. T., Heneman, H. G., III, & Kimball, S. M. (2011). *Teaching assessment for teacher human capital management: Learning from the current state of the art* (WCER Working Paper No. 2011-2). Retrieved from Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2011_02.pdf

Muijs, D. (2006). Measuring teacher effectiveness: Some methodological reflections. *Educational Research and Evaluation*, 12(1), 53-74.

National Council on Teacher Quality (2011). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies*. Washington, DC: National Council on Teacher Quality. Retrieved from website: http://www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf

Nichols, A. R. (2007). *The evaluation of the technical adequacy of a direct observation code developed to assess teacher and student interactions in a preschool setting*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3264469)

Odden, A. (2004). Lessons learned about standards-based teacher evaluation systems. *Peabody Journal of Education*, 79(4), 126-137.

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

Raudenbush, S. W., Martinez, A., Bloom, H., Zhuc, P., & Lina, F. (2008). *An eight-step paradigm for studying the reliability of group-level measures*. Retrieved from William T. Grant Foundation website: http://www.wtgrantfdn.org/publications_and_reports/browse_reports/raudenbush_1

Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation. Retrieved from University of Chicago Consortium on Chicago School Research website: <http://ccsr.uchicago.edu/publications/rethinking-teacher-evaluation-chicago-lessons-learned-classroom-observations-principal>

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., & Benford, R. (2002). Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol. *School Science and Mathematics, 102*(6), 1–20.

Schacter, J., & Thum, Y. M. (2005). TAPping into high quality teachers: Preliminary results from the teacher advancement program comprehensive school reform. *School Effectiveness and School Improvement, 16*(3), 327-353.

Shin, Y., & Raudenbush, S. (2011). *Confidence bounds and power for the reliability of observational measures on the quality of a social setting*. New York, NY: William T. Grant Foundation. Retrieved from <http://www.wtgrantfdn.org/resources/measures-of-social-settings>

Sterbinsky, A., & Ross, S. M. (2003). *School observation measure reliability study*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.

Stuhlman, M. W., Hamre, B. K., Downer, J. T., & Pianta, R. W. (2010). A practitioner's guide to conducting classroom observations: What the research tells us about choosing and using observational systems to assess and improve teacher effectiveness. Charlottesville, VA: The Center for Advanced Study of Teaching and Learning, University of Virginia.

Stumbo, C., & McWalters, P. (2011). Measuring effectiveness: What will it take? *Educational Leadership, 68*(4), 10-15.

The New Teacher Project (2011). *Smart spending for better teacher evaluation systems*. Brooklyn, NY: The New Teacher Project. Retrieved from website: http://tntp.org/assets/documents/TNTP_Smart_Spending_2011.pdf

Tyler, J. T. (2011). *Designing high quality evaluation systems for high school teachers: Challenges and potential solutions*. Retrieved from Center for American Progress website: http://www.americanprogress.org/wp-content/uploads/issues/2011/11/pdf/high_school_evaluation.pdf

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project. Retrieved from website: <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>