References

Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65–75.

Askew, M., Brown, M., Rhodes, V., Wiliam, D., & Johnson, D. (1997). Effective Teachers of Numeracy: Report of a study carried out for the Teacher Training Agency. London: King's College, University of London

Ball D.L and Hill H.C. (2009). Measuring teacher quality in practice In D.H. Gitomer (Ed.), *Measurement issues and assessment for teaching quality*. Washington, DC: Sage Publications.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M.A. Gernsbacher, et al (Ed) *Psychology and the real world: Essays illustrating fundamental contributions to society*, New York: Worth Publishers (p56-64). [Avaliable at http://bjorklab.psych.ucla.edu/pubs/EBjork_RBjork_2011.pdf]

Boaler, J. (2008). When politics took the place of inquiry: A response to the National Mathematics Advisory Panel's review of instructional practices. *Educational Researcher*, 37(9), 588-594.

Bouchamma, Y. 2005. Evaluating teaching personnel. Which model of supervision do Canadian teachers prefer? *Journal of Personnel Evaluation in Education, 18*(4), pp. 289–308.

Bramschreiber, T. (2012). Taking Peer Feedback to Heart. *Educational Leadership, 70*(3), retrieved from http://www.ascd.org/publications/educational-leadership/nov12/vol70/num03/Taking-Peer-Feedback-to-Heart.aspx

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*: Expanded edition. Washington, DC: National Academy Press. http://www.colorado.edu/MCDB/LearningBiology/readings/How-people-learn.pdf

Brophy, J., Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of Research on Teaching* (pp. 328 – 375). New York: Macmillian.

Brown, P. C., Roediger III, H. L., & McDaniel, M. A. (2014). *Make it Stick: The Science of Successful Learning*. Harvard University Press.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2011). The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood (NBER Working Paper No. w17699). National Bureau of Economic Research.

Coe, R. (2002) 'Evidence on the Role and Impact of Performance Feedback in Schools' in A.J.Visscher and R. Coe (eds.) *School Improvement Through Performance Feedback*. Rotterdam: Swets and Zeitlinger

Creemers, B. P. M., & Kyriakides, L. (2006). Critical analysis of the current approaches to modelling educational effectiveness: The importance of establishing a dynamic model. *School Effectiveness and School Improvement*, 17, 347–366. http://www.rug.nl/staff/b.p.m.creemers/testing_the_dynamic_model_of_educational_effectiveness.pdf

Creemers, B. P., & Kyriakides, L. (2011). *Improving Quality in Education: Dynamic Approaches to School Improvement*. Routledge, Taylor & Francis Group.

Danielson, C. (2007). Enhancing professional practice: A framework for teaching. ASCD.

Danielson, C., & McGreal, T. L. (2000). Teacher evaluation to enhance professional practice. ASCD.

Stipek, D. (2010) *How Do Teachers' Expectations Affect Student Learning* http://www.education.com/reference/article/teachers-expectations-affect-learning/

Dumay, X., Coe, R., and Anumendem, D. (2013) 'Stability over time of different methods of estimating school performance'. *School Effectiveness and School Improvement*, vol 25, no 1, pp 65-82.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4-58.

Dweck, C.S. (1999) 'Caution – Praise Can Be Dangerous' *American Educator*, Spring 1999, p4-9. https://www.aft.org/pdfs/americaneducator/spring1999/PraiseSpring99.pdf

Faubert, V. (2009). School evaluation: current practices in OECD countries and a literature review. *OECD Education Working Papers*, No. 42, OECD Publishing. doi:10.1787/218816547156.

Geake, J. G. (2008) Neuromythologies in education. Educational Research 50, 123–133.

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches*. (REL 2014–010). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast.

Glickman, C.D. 1990. *Supervision of instruction: A developmental approach*. 2nd ed. Needham Heights, MA: Allyn and Bacon.

Goe, L., Bell, C., & Little, O. (2008). Approaches to Evaluating Teacher Effectiveness: A Research Synthesis. National Comprehensive Center for Teacher Quality. http://files.eric.ed.gov/fulltext/ED521228.pdf

Goldhammer, R. 1969. Clinical supervision. Special methods for the supervision of teachers. New York, NY: Holt, Reinhart, and Winston

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.

Gorard, S., Hordosy, R., & Siddiqui, N. (2012). How Unstable are 'School Effects' Assessed by a Value-Added Technique?. *International Education Studies*, 6(1), p1.

Gorard, S., See, B. H., & Davies, P. (2012). The impact of attitudes and aspirations on educational attainment and participation. *York: Joseph Rowntree Foundation. Available at:* http://www.jrf.org.uk/sites/files/jrf/education-young-people-parents-full.pdf.

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The Test Matters: The Relationship Between Classroom Observation Scores and Teacher Value Added on Multiple Types of Assessment. *Educational Researcher, 43*(6). 293–303.

Hamre, B.K., Goffin, S.G. & Kraft-Sayre, M. (2009) Classroom Assessment Scoring System Implementation Guide: Measuring and Improving Classroom Interactions in Early Classroom Settings. http://www.teachstone.org/wp-content/uploads/2010/06/CLASSImplementationGuide.pdf

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.

Higgins, S., & Moseley, D. (2001). Teachers' thinking about information and communications technology and learning: Beliefs and outcomes. *Teacher Development*, 5(2), 191-210.

Higgins, S., Katsipataki, M., Kokotsaki, D., Coleman, R., Major, L.E., & Coe, R. (2013). The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit. London: Education Endowment Foundation. [Available at http://www.educationendowmentfoundation.org.uk/toolkit]

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831.

Hill, H. C., Rowan, B., and Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, 42(2), 371-406.

Hinchey, P. H. (2010). Getting Teacher Assessment Right: What Policymakers Can Learn from Research. National Education Policy Center. http://nepc.colorado.edu/publication/getting-teacher-assessment-right

Hobson, A. J., & McIntyre, J. (2013). Teacher fabrication as an impediment to professional learning and development: the external mentor antidote. *Oxford Review of Education*, *39*(3), 345-365.

Howard-Jones, P.A. (2014) Neuroscience and education: myths and messages. *Nature Reviews Neuroscience*. Advanced Online Publication, published online 15 October 2014; doi:10.1038/nrn3817. http://www.nature.com/nrn/journal/vaop/ncurrent/pdf/nrn3817.pdf

Husbands, C., & Pearce, J. (2012). What is great pedagogy: A literature review for the National College of School Leadership. National College for School Leadership.

Isoré, M. (2009). Teacher evaluation: current practices in OECD countries and a literature review. OECD Education Working Papers, No. 23 http://hdl.handle.net/123456789/2541

Jacob, B. A., & Lefgren, L. (2008). Can principals identify effective teachers? Evidence on subjective performance evaluation in education. *Journal of Labor Economics*, *26*(1), 101-136.

James, M. and Pollard, A. (2011) TLRP's ten principles for effective pedagogy: rationale, development, evidence, argument and impact. *Research Papers in Education*, 26 (3). pp.

Johnson S. M., Fiarman, S. E, Munger, M. S., Papay, J. P., & Qazilbash, E. K. (2009). *A user's guide to Peer Assistance and Review*. Retrieved from http://web.archive.org/web/20100812220001/http://www.gse.harvard.edu/~ngt/par/resources/users_guide_to_par.pdf

Kahlenberg, R. D. (2007). Peer Assistance and Review. Retrieved from http://www.aft.org/newspubs/periodicals/ae/fall2007/kahlenbergsb1.cfm

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment (MET Project Research Paper). Seattle, WA: Bill & Melinda Gates Foundation.

Kennedy, M. M. (2010). Attribution error and the quest for teacher quality. *Educational Researcher*, 39(8), 591-598.

Kirschner, P. A., Sweller, J., and Clark, R. E. (2006) Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist 41 (2) 75-86.

Ko J., Sammons P. and Bakkum, L. (2013) Effective Teaching: a review of research and evidence. CfBT Education Trust. http://cdn.cfbt.com/~/media/cfbtcorporate/files/research/2013/r-effective-teaching-2013.pdf

Kohut, G. F., Burnap, C., & Yon, M. G. (2007). Peer observation of teaching: Perceptions of the observer and the observed. *College Teaching, 55*(1), 19-25.

Kraft, M. A., & Papay, J. P. (2014). Can Professional Environments in Schools Promote Teacher Development? Explaining Heterogeneity in Returns to Teaching Experience. Educational Evaluation and Policy Analysis, 0162373713519496.

Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. Educational Evaluation and Policy Analysis, 25, 287–298.

Kyriakides L., Creemers B.P.M. & Antoniou P. (2009) Teacher behaviour and student outcomes: Suggestions for research on teacher training and professional development. *Teaching and Teacher Education*, 25, 12-23.

Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: A meta-analysis of studies exploring factors of effective teaching. Teaching and Teacher Education, 36, 143–152.

Kyriakides, L., Creemers, B., Antoniou, P., & Demetriou, D. (2010). A synthesis of studies searching for school factors: Implications for theory and research. British Educational Research Journal, 36, 807–830.

Little, O., Goe, L., & Bell, C. (2009). A Practical Guide to Evaluating Teacher Effectiveness. National Comprehensive Center for Teacher Quality.

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, 44(1), 47–67.

Lord, B., Cress, K., & Miller, B. (2008). Teacher leadership in support of large-scale mathematics and science education reform. In M.M. Mangin & S.R. Stoelinga (Eds.), *Effective teacher leadership: Using research to inform and reform.* New York: Teachers College Press.

Marsh, H. W. (1982) SEEQ: a reliable, valid and useful instrument for collecting students' evaluations of university teaching, *British Journal of Educational Psychology, 52*, 77–95.

Marsh, H.W. (2007), Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases and usefulness. In Perry, R.P. and Smart, J.C. (Eds), *The Scholarship of Teaching and Learning in Higher Education: An Evidence-Based Perspective*. Berlin: Springer, pp. 319-83.

Marshall, J. C., Horton, B., & White, C. (2009). EQUIPping teachers: A protocol to guide and improve inquiry-based instruction. *The Science Teacher, 76*(4), 46–53.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). Evaluating Value-Added Models for Teacher Accountability. Monograph. RAND Corporation. PO Box 2138, Santa Monica, CA 90407-2138. http://www.rand.org/pubs/monographs/MG158.html

McCaffrey, D., Sass, T., Lockwood, J., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates, *Education Finance and Policy*, 4(4), 572-606. http://dx.doi.org/10.1162/edfp.2009.4.4.572

Mihaly, K., McCaffrey D. F., Staiger D. O., and Lockwood J. R. (2013) A Composite Estimator of Effective Teaching: Report for the Measures of Effective Teaching Project, Bill and Melinda Gates Foundation. Available at http://www.metproject.org/downloads/MET_Composite_Estimator_of_Effective_Teaching_Research_Paper.pdf

Muijs, D, Kyriakides L, van der Werf G, Creemers B, Timperley H & Earl L (2014) State of the art – teacher effectiveness and professional learning, *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 25:2, 231-256

National Mathematics Advisory Panel (2008) Foundations for Success: The Final Report of the National Mathematics Advisory Panel. March 2008, U.S. Department of Education, Washington, DC. http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf

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 PolicyAnalysis Archives*, 18(23), n23.

Papay, J. P. (2011). Different Tests, Different Answers: The Stability of Teacher Value-Added Estimates Across Outcome Measures. *American Educational Research Journal*, 48(1), 163-193.

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: concepts and evidence. *Psychological Science in the Public Interest*, 9(3), 105-119. http://www.psychologicalscience.org/journals/pspi/PSPI_9_3.pdf

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system*. Baltimore: Paul H. Brookes.

Polikoff, M. S. (2014). Does the Test Matter? Evaluating teachers when tests differ in their sensitivity to instruction. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.). Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project (pp. 278-302). San Francisco, CA: Jossey-Bass.

Popham, W.J. (1998). The dysfunctional marriage of formative and summative teacher evaluation. Journal of Personnel Evaluation in Education, 1: 269-273.

Popham, WJ and Ryan JM (2012) Determining a high-stakes test's instructional sensitivity. Paper presented at the annual meeting of the National Council on Educational Measurement, April 12-16, Vancouver, B.C., Canada.

Raudenbush, S.W. (2004), 'What Are Value-added Models Estimating and What Does This Imply for Statistical Practice?', *Journal of Educational and Behavioral Statistics* 29(1):121–129.

Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. Education, 4(4), 492-519.

Riener, C., & Willingham, D. (2010). The myth of learning styles. Change: The magazine of higher learning, 42(5), 32-35. http://new.peoplepeople.org/wp-content/uploads/2012/07/The-Myth-of-Learning-Styles.pdf

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. The American Economic Review, 94(2), 247-252.

Rockoff, J. E., Jacob, B. A., Kane, T. J., & Staiger, D. O. (2011). Can you recognize an effective teacher when you recruit one? *Education*, 6(1), 43-74.

Rosenshine, B. (2010) Principles of Instruction. International Academy of Education, UNESCO. Geneva: International Bureau of Education. http://www.ibe.unesco.org/fileadmin/user_upload/Publications/Educational_Practices/EdPractices_21.pdf

Rosenshine, B. (2012) Principles of Instruction: Research based principles that all teachers should know. *American Educator*, Spring 2012. http://www.aft.org/pdfs/americaneducator/spring2012/Rosenshine.pdf

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020-1049.

Shulman, L. S. (1988). A Union of Insufficiencies: Strategies for Teacher Assessment in a Period of Educational Reform. *Educational Leadership*, *46*(3), 36-41.

Shulman, L. (2009). Assessment of teaching or assessment for teaching? In D.H. Gitomer (Ed.), Measurement issues and assessment for teaching quality. Washington, DC: Sage Publications.

Strong, M., Gargani, J., & Hacifazlioglu, O. (2011). Do we know a successful teacher when we see one? Experiments in the identification of effective teachers. Journal of Teacher Education, 62(4), 367–382. [Abstract at: http://jte.sagepub.com/content/62/4/367.abstract]

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *The American Economic Review*, *102*(7), 3628-3651.

Timperley, H., Wilson, A., Barrar, H. & Fung, I. (2007) Teacher professional learning and development: Best evidence synthesis iteration. Wellington, New Zealand: Ministry of Education. http://www.educationcounts.govt.nz/publications/series/2515/15341

Timperley, H. (2008). Teacher professional learning and development. International Academy of Education, International Bureau of Education, UNESCO. http://www.ibe.unesco.org/fileadmin/user_upload/Publications/Educational_Practices/Ed Practices_18.pdf

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, 73(1), 89–122.

Willingham, D. T. (2008). What Will Improve a Student's Memory?. *American Educator*, *32*(4), 17-25.

Appendix

A: Original research questions

In more detail, the study set out to review the evidence from existing research to address the following questions:

1. What are the elements of teaching effectiveness and classroom/teaching quality?

   a. What does the educational effectiveness[7] literature say about the factors/characteristics/behaviours of teachers/teaching that are associated with high student attainment/progress?

   b. What is the evidence from intervention studies (eg RCTs) about the classroom strategies that produce increased attainment/progress?

   c. What evidence from psychology (eg on learning, memory, neuropsychology) indicates pedagogical practices that are most likely to lead to deep understanding and retained knowledge?

2. What frameworks/protocols exist for measuring classroom/teaching quality (including use of video and student surveys)?

   a. What frameworks/protocols have been used in research studies? What evidence is there of how effectively these frameworks capture real quality? To what extent are they aligned with the evidence reviewed in 1?

   b. What frameworks/protocols have been used in schools (by practitioners) around the world for measuring teacher effectiveness/quality? What evidence is there of how effective and reliable these frameworks are? To what extent are they aligned with the evidence reviewed in 1?

   c. What requirements for training, accreditation and quality assurance do these frameworks have?

3. In what ways have these frameworks been used in practice to improve practice?

   a. What kinds of outputs/reporting have been developed for these frameworks/protocols?

   b. What models of observer/observed have been tried (eg peer-to-peer, self-evaluation, principal/line manager, external evaluator), and how have professionals collaborated on this?

---

[7] Eg school effectiveness research, and the 'process-product' literature that looks for correlations between school or classroom processes and outcomes

c. What models of feedback/dialogue and improvement mechanisms exist (eg appraisal/evaluation, information for self-evaluation, support and 'consultation' in interpreting and responding to feedback, goal-setting + feedback, etc)?

4. What evidence is there of the impact of any of these approaches on student outcomes?

   a. What high-quality (eg RCT) evaluations exist of interventions based on feedback of classroom quality evaluation?

   b. What claims exist, based on case-studies or other less-rigorous designs? What relevant work is currently underway?