Does video technology improve the classroom observation process?
Results from a randomized experiment

David Blazar
University of Maryland

Thomas J. Kane
Daniel Thal
Harvard Graduate School of Education

Background

School districts around the country are implementing new teacher evaluation systems in an effort to improve upon "cursory" systems (Hill & Grossman, 2013, p. 371) that did little to differentiate teachers with regard to performance standards (Kraft & Gilmour, 2016; Weisberg, Sexton, Mulhern, & Keeling, 2009). Designing and implementing observation and feedback cycles that improve teachers' instruction and, in turn, students' academic performance are complex tasks that require schools to wrestle with several key challenges (Darling-Hammond, 2013; Papay, 2012). Schools need to identify the appropriate instrument(s) to use to assess teachers' skills, and build expertise of personnel and school leaders who serve as the primary observers to use these rubrics. Logistically, schools need the capacity to visit classrooms and collect sufficient data on each teacher in a way that can be used to inform ongoing improvement efforts (Hill & Grossman, 2013). Teachers must also be open to receiving feedback, particularly from school principals who often straddle the line between being a formal evaluator and an instructional leader (Herman & Baker, 2009; Kraft & Gilmour, in press).

Purpose

In this paper, we evaluate one strategy thought to address these key challenges: allowing teachers to submit their own recorded lesson videos for evaluation in lieu of in-person observations (Jacobs, Doherty, Lakis, Lasser, & Staresina, 2014). We hypothesized that digital video would offer a number of potential advantages over in-person observations. Video would provide a more detailed, objective record of what occurred in teachers' classrooms than an observer's written notes. In turn, feedback cycles could focus on this record of teachers' classroom practice in order to identify areas of strength versus areas of weakness in need of improvement. Reliance on objective evidence could also help build trust between teachers and principals. For example, teachers may be more assured that evaluators' feedback reflected what actually happened in the classroom. Finally, video would allow principals to time-shift their observational duties to quieter times of the day or week, and facilitate the use of external observers and content experts. Ultimately, we hypothesized that evaluation feedback would be most effective if paired with a video record.

Intervention and Research Design

We tested this hypothesis through an experimental evaluation in which treatment teachers across four U.S. states were given a special video camera and access to a secure website for sorting and viewing their own recorded lessons. In Delaware, Georgia, Colorado, and California, 433 teachers and 134 administrators volunteered to participate in our study across two cohorts. A secure software platform allowed observers, including both formal evaluators and content experts, to watch the videos and provide time-stamped comments aligned to specific moments in the videos. These videos and comments were used in one-on-one discussions between teachers and principals or external content experts. We compare the efficacy of this approach to in-person classroom observation and feedback cycles, as assessed on teacher- and principal-reported surveys, student surveys, teacher turnover, and student achievement.

Findings

We find that giving control of the cameras to teachers successfully shifted the mode of the classroom observations from in-person to video and resulted in more productive feedback cycles. The opportunity to watch their own lessons resulted in treatment teachers becoming more self-critical. Teachers in the treatment group perceived their supervisors to be more supportive and their observations to be fairer. They reported fewer disagreements on the ratings they received and were more likely to describe a specific change in their practice resulting from their post-observation conference. Treatment administrators reported spending more time observing and less time on paperwork. Moreover, the ability to watch video allowed supervisors to time-shift their observation duties. In addition to improving several dimensions of the classroom observation process, relying on video also improved measures of teacher retention. Treatment teachers were more likely to remain in their teaching assignment in the year following the intervention, as well as remain in their same school or their district.

However, the intervention did not improve students' reports of their classroom experiences, or their academic achievement as measured on end-of-year state tests in math and reading. One possible explanation for these null results is that effects on the observation process were not large enough to result in improved student outcomes – a frequent trend in the academic literature on teacher development. For example, several large-scale experimental interventions of content-focused professional development have found that these efforts can improve teachers' content knowledge or classroom instruction but often fail to improve students' academic performance (Garet et al., 2008; Garet et al., 2011; Garet et al., 2016; Glazerman et al., 2010). In a recent meta-analysis of the causal evidence on teacher coaching – an intervention that also relies on in-depth observation and feedback cycles – Kraft, Blazar, and Hogan (2016) found that changes in student achievement appeared to require relatively large improvements in instructional quality. In this study, improvements in proximal outcomes likely were not large enough or sufficient to observe improvements in outcomes that were more distal.

Conclusion

Where does this leave schools and districts as the continue to experiment with ways to implement teacher evaluation? Even without statistically significant effects on students' academic performance, we believe that there are several reasons to incorporate video in observation and feedback cycles. We document that this approach has meaningful effects on the daily operations of schools and relationships between teachers and school leaders. It also is possible that these effects may show up on student test scores after several years, as effects on the school environment continue to accumulate; or, that this approach improves other student outcomes that we were not able to measure in this study, such as their behavior or engagement. Collecting video for the main purpose of formal evaluation could also be useful for other instructional improvement efforts. For example, these data could be used in the process of delivering teacher coaching. In turn, the cost of collecting video could be shared between budgets allocated to evaluation and to professional development.

References

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2016). *Teacher turnover, teacher quality, and student achievement in DCPS* (No. w21922). National Bureau of Economic Research.

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science*, *333*(6045), 1034-1037.

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2014). A helping hand? Teacher quality and learning outcomes in kindergarten. *Banco Interamericano de Desarrollo, Washington, DC. Inédito*.

Atteberry, A., Loeb, S. L., & Wyckoff, J. (2017). Teacher churning within schools: Impacts on student achievement. *Educational Evaluation and Policy Analysis, 39*(1), 3-30.

Blazar, D. (2015). Effective teaching in elementary mathematics: Identifying classroom practices that support student achievement. *Economics of Education Review, 48*, 16-29.

Blazar, D. (2015). Grade assignments and the teacher pipeline: A low-cost lever to improve student achievement?. *Educational Researcher*, *44*(4), 213-227.

Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment*, *17*(2-3), 62-87.

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement.* New York: Teachers College Press.

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, *34*(2), 267-297.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W…. & Sztejnberg, L. (2008). The impact of two professional development interventions on early reading instruction and achievement (NCEE 2008-4030). Washington, D.C.: *National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.* Retrieved from http://ies.ed.gov/ncee

Garet, M. S., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K…. & Doolittle, F. (2011). Middle school mathematics professional development impact study: Findings after the second year of implementation (NCEE 2011-4024).. Washington, DC: *National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.* Retrieved from http://ncee.ed.gov

Garet, M. S., Heppen, J.B., Walters, K., Parkinson, J., Smith, T.M., ..., Wei, T.E. (2016, November). *Focusing on mathematical knowledge: The impact of content-intensive*

teacher professional development. (NCEE 2016-4010). Washington DC: *National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.*

Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., & Jacobus, M. (2010). *Impacts of comprehensive teacher induction: Final results from a randomized controlled study* (No. 691d9603eb074051b57684e4affae4d4). Mathematica Policy Research.

Hamre, B. K., Pianta, R. C., & Chomat-Mooney, L. (2009). Conducting classroom observations in school-based research. In L. Dinella (Ed), *Conducting psychology research in school-based settings: A practical guide for researchers conducting high quality science within school environments* (pp. 79-106). Washington, DC: APA Press.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191–205.

Ho, A. D., & Kane, T. J. (2013). *The reliability of classroom observations by school personnel.* Seattle, WA: Bill & Melinda Gates Foundation.

Herman J. L., & Baker, E. L. (2009). Assessment policy: Making sense of the Babel. In G. Sykes, B. Schneider, and D. N. Plank (Eds.), *Handbook of educational policy research* (pp. 176-190). New York: Routledge.

Hill, H. C. (2015). *Randomized trial meets the real world: Exploring and explaining null results in recent federally-funded RCTs*. Working Paper. Harvard University.

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough teacher observation systems and a case for the generalizability study. *Educational Researcher, 41*(2), 56-64.

Hill, H. C., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review*, *83*(2), 371-384.

Jacob, B. A. (2015). *Harnessing the value of "failure."* Brookings Institute. Retrieved from: https://www.brookings.edu/research/harnessing-the-value-of-failure/

Jacob, B. A. (2015). When evidence is not enough: Findings from a randomized evaluation of evidence-based literacy instruction (EBLI). Working Paper 21643. Cambridge, MA: National Bureau of Economic Research.

Jacobs, S., Doherty, K. M., Lakis, K., Lasser, P., & Staresina, L. N., (2014). *2015 state teacher policy yearbook.* Washington, DC: National Center on Teacher Quality.

Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, *114*(10), 1-39.

Kane, T., & Staiger, D. O. (2010). *Learning about teaching: Initial findings from the measures of effective teaching project*. Seattle, WA: Bill & Melinda Gates Foundation.

Kane, T., & Staiger, D. O. (2012). *Gathering feedback on teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation.

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.* Seattle, Seattle, WA: Bill & Melinda Gates Foundation.

Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

Kraft, M.A., Blazar, D., Hogan, D. (2016). The effect of teaching coaching on instruction and achievement: A meta-analysis of the causal evidence. Brown University Working Paper.

Kraft, M.A. & Gilmour, A. (in press) Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly.*

Kraft, M. A., & Gilmour, A. F. (2016). *Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness*. Working paper). Providence, RI: Brown University.

Kraft M. A., Marinell W. M., & Yee D. (2016). School organizational contexts, teacher turnover, and student achievement: Evidence from panel data. *American Educational Research Journal, 53(5*), 1411-1499.

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*, *36*(4), 476-500.

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*(1), 29-45.

Ost, B. (2014). How do teachers improve? The relative importance of specific and general human capital. *American Economic Journal. Applied Economics*, *6*(2), 127.

Papay, J. P. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review, 82*(1), 123-141.

Papay, J. P., Taylor, E. S., Tyler, J. H., & Laski, M. (2016). *Learning job skills from colleagues at work: Evidence from a field experiment using teacher performance data* (Working Paper No. W21986). Cambridge, MA: National Bureau of Economic Research.

Porter, A. C. & Smithson, J. L. (2001). *Defining, developing, and using curriculum indicators*. Philadelphia, PA: Consortium for Policy Research in Education.

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on Head Start teachers and children. *Journal of Educational Psychology, 102*(2), 299-312.

Scher, L., & O'Reilly, F. (2009). Professional development for K–12 math and science teachers: What do we really know?. *Journal of Research on Educational Effectiveness*, *2*(3), 209-249.

Steinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's Excellence in Teaching project. *Education Finance and Policy*, *10*(4), 535-572.

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

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *The New Teacher Project*.

Young, A. (2016). *Channelling Fisher: Randoization tests and the statistical insignificance of seemingly significant experimental results*. Working Paper. London School of Economics. Retrieved from: http://personal.lse.ac.uk/YoungA/ChannellingFisher.pdf