Leading via Teacher Evaluation: The Case of the Missing Clothes?

Joseph Murphy, Philip Hallinger and Ronald H. Heck

EDUCATIONAL RESEARCHER 2013 42: 349 originally published online 30 July 2013
DOI: 10.3102/0013189X13499625

Leading via Teacher Evaluation: The Case of the Missing Clothes?

Joseph Murphy[1], Philip Hallinger[2], and Ronald H. Heck[3]

Over the last decade, teacher evaluation has moved onto center stage in efforts to strengthen schooling. In this article, we address the question of whether focusing on this administrative process is likely to accomplish what reformers hope. We answer that question by examining the available evidence, both direct and indirect. We deepen the empirical narrative that emerges there by highlighting workplace norms and conditions of work in schools.

Keywords: administration; descriptive analysis; educational reform; evaluation

Introduction

Since the turn of the century, considerable attention has been devoted to using teacher evaluation to strengthen schooling. The question at hand is: is teacher evaluation up to the task, at least in terms of enhancing student learning? We answer that question below by examining the available evidence, both direct and indirect. We deepen the empirical narrative that emerges there by foregrounding workplace norms and conditions of work in schools.

Before we begin that analysis, however, we provide a few words to explain why teacher evaluation has been given a starring role in the current school improvement play. To begin with, over the last two decades, considerable evidence has accumulated about the cardinal position of teachers and teaching in the school improvement equation (Lewis, 2008; Louis, Dretzke, & Wahlstrom, 2010), findings that are even more robust for students placed at risk by society and its institutions (Clotfelter, Ladd, & Vigdor, 2005). Not surprisingly, efforts to strengthen teaching have been on the rise. A good portion of that energy is being devoted to teacher evaluation.

A second explanation is found in the new era of accountability in education that began to take root in the 1990s. The press for responsibility has been ratcheted up considerably. That press has been accompanied by a sea change in focus as well, from attention to inputs and processes to concern for organizational outcomes, especially student learning results (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012). More specifically, in this new era of accountability a series of optimistic conclusions concerning the strength of teacher effects on student learning (e.g., Sanders, Ashton, & Wright, 2005; Sanders & Rivers, 1996; Wright, Horn, & Sanders, 1997) has led policy makers to conclude that individual teachers can and should be held accountable for year-to-year gains in their students' achievement. Consequently, a new generation of teacher evaluation models is emerging, one that incorporates measures of growth in student achievement as a criterion of teacher effectiveness (e.g., Danielson, 2007; Gates Foundation, 2013; Kimball, White, Milanowski, & Borman, 2004; Milanowski, 2004).[1]

The third explanation focuses on shifts in the context in which teacher evaluation plays out. Historically, teacher unions have been able to buffer teachers from outside forces, especially as regards the core technology of the profession. As we have moved toward a market-oriented phase of our development, that influence is wearing thin (Murphy, 2012). Concurrently, the context is shifting in ways that privilege ideology and tools from the corporate HR culture.

One last introductory assignment remains: providing a bit of understanding about the construct, teacher evaluation, in play. Our definition is the classic one crafted by Castetter (1976): *assessments by a supervisor of a subordinate to draw conclusions about the performance of the subordinate to improve performance and to make decisions about terms of employment*. Teacher evaluation is thus one dimension of the hierarchical architecture of schooling, with principals filling the role of supervisor and teachers occupying the role of subordinates. In recognition of the "professional nature" of teaching, considerable effort has been devoted over the years to "soften" the bureaucratic language of teacher evaluation—in the practice, development, and academic wings of the profession if not in the policy arena. But it is

[1]Vanderbilt University, Nashville, TN
[2]Hong Kong Institute of Education, Hong Kong
[3]University of Hawaii-Manoa, Honolulu, Hawaii

Educational Researcher, Vol. 42 No. 6, pp. 349–354
DOI: 10.3102/0013189X13499625

important to remember that the core of teacher evaluation is what it is, a system of performance management that is scaffolded on hierarchical authority and control over labor (Lipham, 1964; Thompson, 1961).

The Direct Evidence on the Effectiveness of Teacher Evaluation

Principals have been evaluating teachers for over a century. Not surprisingly, a good deal of descriptive data and prescriptive opinion has accumulated in this domain. More recently, there has been a sharp increase in research to map the domains of quality teaching and to evaluate their effects on student learning (Seidel & Shavelson, 2007). Evaluation objects range from teachers' preparation and credentials to their attributes, cognitive strategies, instructional skills, and behavior, whereas data are as diverse as observations and videos of teaching performance, student and teacher surveys, "value-added" estimates of teaching effectiveness, and various combinations.

During the past decade, value-added models (VAM) of teacher evaluation, which focus on estimating teachers' contribution to achievement gains after statistically removing extraneous sources of influence (e.g., student background, classroom composition), and direct classroom observations have become primary means of evaluating teachers' instructional effectiveness. Although each approach has produced evidence suggesting that some teachers contribute more to student learning than others over time (e.g., McCaffrey, Sass, Lockwood, & Mihaly, 2009; Milanowski, 2004; Rivkin, Hanushek, & Kain, 2005; Rockoff & Speroni, 2010; Taylor & Tyler, 2011), they have also raised questions regarding the validity, accuracy, and equity of value-added scores assigned to individual teachers (Darling-Hammond et al., 2012), the consistency of ratings across individuals responsible for making evaluations (Rockoff & Speroni, 2010), as well as the optimal mix of instructional strategies, skills, and behavior that might account for differences in student learning (Seidel & Shavelson, 2007), given broad variation in student characteristics and the purpose, focus, and technical details of each specific evaluation model. It remains unclear how well individual studies can be integrated and what the implications of their findings are for designing new evaluation systems (McCaffrey et al., 2009).

Previous evaluation research has also raised concern about the challenges of achieving sufficient implementation fidelity regarding who can and should be involved in observing teachers (e.g., administrators, resource teachers, outside evaluators) or the potential costs to train evaluators to achieve a reasonable inter-rater reliability threshold. Where evaluators may differ in content area expertise and rating agreement, or where the evaluations do not adequately reflect teachers' classroom skills or demonstrate stability with respect to student achievement, Kimball and Milanowski (2009) conclude that teachers can receive evaluations that are not defensible:

> We had hoped that we could identify evaluator practices associated with higher validity, which districts could then use to train evaluators to follow. Although disappointing, our failure to find such practices is important because it shows the complexity in identifying and assuring the use of good evaluation practice. . . . If policy makers and program designers want evaluation scores to be more highly related to some criterion such as student achievement, it will take more than specific rubrics and basic training of evaluators in the process to achieve a strong relationship. (p. 65)

When data on staff performance are to be used in making personnel decisions that have important consequences, it is incumbent upon system designers and evaluators to demonstrate the instruments and methods yield results that meet widely accepted standards of validity. Particularly prominent is the "evaluation of any adverse consequences for individuals and groups that are associated with bias in test scoring and interpretation or with unfairness in test use" (Messick, 1994, p. 21). To date, evidence from standards-based teacher/VAM evaluation systems has yet to meet this necessary standard for making accurate and equitable high stakes personnel decisions for individual teachers. Notably, our conclusion concerns:

- the stability of VAM teacher estimates across subjects, grade levels, and time;
- the present capacity of school administrators to produce valid evaluations when employing these new tools without extensive training; and
- the efficacy of teacher evaluation to yield improvements in teaching practice.

Indirect Evidence on the Effectiveness of Teacher Evaluation

So here is what we know so far. Teacher evaluation for most of the twentieth century had very little influence on much of anything of substance. It provided little more than a patina of symbolism, a layer of organizational legitimacy to buffer the core technology of education (Weis, 1990). The newer, more substantive teacher evaluation systems of the last 15 years have not been shown to power school improvement, as defined in terms of student learning either, although given the limited number of studies the question is hardly settled.

In cases such as this, it seems appropriate to expand the search for evidence on possible linkages between teacher evaluation and school improvement to the indirect evidence. We believe that this indirect evidence is most likely to be lurking in studies of school improvement, that is, in analyses of schools that help all children and youngsters reach ambitious targets of performance. When we conducted this deep assessment, however, we found that teacher evaluation as an explanatory or ancillary variable of interest in studies of each of the following broad domains of school improvement was conspicuous by its absence: effective schools and school improvement (Teddlie & Reynolds, 2000); instructional leadership (Bossert, Dwyer, Rowan, & Lee, 1982); leadership in general (Louis, Leithwood, Wahlstrom, & Anderson, 2010); transformational leadership (Leithwood & Jantzi, 2005); restructuring schools (Murphy, 1991); comprehensive school reform (Herman & Stringfield, 1997); effective use of data (Supovitz & Klein, 2003); sustainability, scaling up, and implementation (Borman, 2005); special populations (Reyes, Scribner, & Scribner, 1999); achievement gaps/at-risk

students (Murphy, 2010); turnaround schools (Leithwood, Harris, & Strauss, 2010); and teacher effects (Hattie, 2009).

Evidence From the Nature of Schooling

It seems to us that an investigation of the issue at hand would be more complete, and more satisfactory, if we also factored in some analysis of the organizational dynamics of schooling, what sociologists refer to as the occupational norms and workplace conditions of schools (Hamilton, Stecher, Russell, Marsh, & Miles, 2008; Lortie, 1975). One aspect of this context is the difficulty at the heart of the supervision function, the tension between managerial efforts to deepen performance evaluation and strive to engage in the development of professionally oriented work (Curry, 2008; Goldstein, 2004). Second, while it is beyond our ability to treat them in detail here, it is reasonable to expect that there are costs associated with managerial systems such as teacher evaluation (e.g., reduced organizational commitment, a decline in professionalism). These costs receive almost no treatment in the current literature (Weis, 1990).

It merits notice too that throughout time school leaders have rarely been found to have the skills to operate the teacher evaluation machinery well, that is, deeply and meaningfully. Efforts over the last 30 years, since the start of the school reform revolution, to change that organizational reality have been largely unsuccessful (Blase & Kirby, 2009). What this means is this. Teachers are influenced by those they perceive as credible sources of knowledge on instructional issues (Friedkin & Slater, 1994), especially those with content-based knowledge (Printy, 2008; Supovitz, 2008). School leaders rarely fall into this category. For a variety of substantive reasons, leaders are poorly positioned to make teacher evaluation work well.

Additionally, even if the tool could work and leaders had abundant knowhow about its use, there are reasons why administrators are unlikely to use teacher evaluation in the service of school improvement. To begin with, there are powerful organizational explanations why principals have not and do not exercise tight control over teachers, especially in the domain of instruction (Meyer & Rowan, 1975). Principals require the support of teachers to ensure that the school "runs" well and that conflict is corralled inside classrooms, or at least inside the school (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010). They know that a powerful way to garner that support is to provide teachers with autonomy over their individual classrooms (Blase & Kirby, 2009). Additionally, teachers have willingly participated in this exchange, trading influence over school-level activities for freedom in those classrooms. This exchange norm is woven deeply into the fabric of the schooling tapestry (Meyer & Rowan, 1975; Powell, 1991). It is not subject to alteration by appeals or demands from the policy world. As Hamilton and colleagues (2008) help us see, workplace conditions need to change before claims about linkages between teacher evaluation and school improvement can even be considered.

Second, even if the teacher evaluation machinery was functional and managers had the skills to operate the system, it is nearly impossible to believe that they have the time to undertake the work (Crum & Sherman, 2008). Spans of control in schools, i.e., organizational conditions, make anything beyond compliance checking and ritual problematic (Weis, 1990). Here is a "normal" situation in schools. Recent studies reveal that the average principal spends around 18 percent of his or her time in the area of instruction and curriculum (May & Supovitz, 2011), and around 3 percent of total time on teacher evaluation, numbers largely unchanged after 30 years of concentrated efforts to increase them (see Murphy, 1987, for a review of early time analyses in this area). The average elementary school has 475 students, 20 students per teacher, one principal (and no assistant principal), and a small cadre of other professional educators and staff that require the principal's attention. Let us assume a nine-hour day (2700 minutes per week). This means that the average principal spends about 80 minutes a week on teacher evaluation, about 3 minutes per teacher per week.

In summary, we are left with some realities that make the odds on this wager long. First, managers, by and large, are not qualified to do this work. They are not "head teachers." Nor are they "managing partners." We have proceeded down this path before, only to end up with narrow, decontextualized perspectives of learning and teaching. Second, managers have no appetite to do this work. In the well-choreographed play called "schooling," leaders avoid interfering with the work of teachers, especially inside classrooms. It is a production they know is not wise to change in any substantive way. Third, when we run the analyses, we find that even if points one and two were to be altered, there is little time to do this work. The resources to do this well cannot be mandated or wished into existence. And discussions of providing assistance for principals to undertake this work overlook the existing evidence on the willingness of the public to increase administrative costs in education.

Conclusion

We return to the central question of the article. Is teacher evaluation a good candidate to power school improvement? The evidence that we reviewed from multiple perspectives leads us to suggest caution in this area. Relatedly, it merits notice that teacher evaluation has been reinvented numerous times across the last century. During that time, we have witnessed a major shift from traits and characteristics of teachers to goals. We have seen the evolution from high inference judgments to objective-driven evaluations. Additionally, one does not need to be too old to remember how clinical approaches to supervision and evaluation were to sweep the old understanding of teacher evaluation aside in favor of revolutionary new perspectives that would create more effective schools. Our investigations tell us that we should be cautious in accepting claims about the ability of teacher evaluation to power significant school improvement simply because it is equipped with a fourth ("new" views of effective teaching) or fifth (student test scores) generation engine.

Finally, let us put aside our concerns and entertain the possibility of a world in which teacher evaluation is transformed into a muscled tool of school improvement, leadership is linked to the core technology, and close administrative control over teaching moves from negatively to positively yoked to student learning. Surely, it would be appropriate to follow this approach. Perhaps, but perhaps not. Throughout the analyses and debates on teacher evaluation over the last decade, we have uncovered

very little discussion about what this tool actually is: an instrument of industrial-era management, of well-informed managers directing the work of the laboring class toward greater efficiency. There is, of course, nothing wrong with this per se, although few talk about teacher assessment in this manner because it is indelicate and off-putting. The real problem here is that it privileges organizational architecture (bureaucracy, hierarchy, and institutionalism) under a very thin veneer of professionalism, a framework with limited linkages to school improvement. There is a new world taking shape in education, but it is not one in which the pillars of industrial management fit particularly well (Grubb & Flessa, 2006; Wenger, 2000).

Equally important, there is a robust body of empirical work that informs us that if school improvement is the goal, school leaders would be advised to spend their time and energy in areas other than teacher evaluation. Many leader initiatives can positively impact student achievement even if instructional quality remains unchanged, i.e., through actions that substitute for and/or enhance teaching (Supovitz, Sirinides, & May, 2009; Witziers, Bosker, & Kruger, 2003). Work here includes establishing a powerful sense of vision, with strong academic mission and challenging organizational goals and expectations (Leithwood & Jantzi, 2005); enhancing student opportunity to learn (Harris & Herrington, 2006); developing and using data systems to inform and monitor decisions (Lachat & Smith, 2005); creating personalized learning environments in which all youngsters are cared for, participate in, and have ownership of the school (Crosnoe, 2011); developing a school culture conducive to learning (Sebastian & Allensworth, 2012); and providing alignment and cohesiveness to all school actions (Bryk, Sebring, & Allensworth, 2010).

Studies also tell us that school administrators will be more likely to positively impact instructional quality if they allocate their direct efforts with teachers into facilitative channels. Studies highlight four bundles of actions in particular: providing actionable feedback to teachers (Hattie, 2009); developing communities of practice in which teachers share goals, work, and responsibility for student outcomes (Wahlstrom & Louis, 2008); offering abundant support for the work of teachers (Leithwood & Jantzi, 2005); and creating systems in which teachers have the opportunity to routinely develop and refine their skills (Bryk et al., 2010). A cardinal point here is the primacy of the facilitative role of leaders, an approach with considerably more empirical linkage to learning outcomes than direct one-on-one teacher evaluation work (Crum & Sherman, 2008; Louis et al., 2010).

NOTE

[1]Our conclusions apply to the linkages between teacher evaluation and school improvement in general, regardless of the components or elements of the "system." It is important to note, however, that we find the evidence for using VAM as a significant element in teacher evaluation systems to be insufficient. We conclude that both the magnitude and sustainability of effects of teacher-related variables on student learning outcomes remain inconsistent, and at times overstated (Darling-Hammond et al., 2012; Rothstein, 2009). This caution applies in particular to the value-added approach to measuring gains in student learning (see also Baker et al., 2010; McCaffrey et al., 2003; Darling-Hammond et al., 2012).

REFERENCES

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., . . . Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers* (EPI Briefing Paper 278). Washington, DC: Economic Policy Institute.

Blase, J., & Kirby, P. (2009). *Bringing out the best in teachers: What effective principals do*. Thousand Oaks, CA: Corwin.

Borman, G. D. (2005). National efforts to bring reform to scale in high-poverty schools: Outcomes and implications. *Review of Research in Education*, *29*, 1–27.

Bossert, S., Dwyer, D., Rowan, B., & Lee, G. (1982). The instructional management role of the principal. *Educational Administration Quarterly*, *18*(3), 34–64.

Bryk, A., Sebring, P., Allensworth, E., Luppescu, S., & Easton, J. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.

Castetter, W. B. (1976). *The personnel function in educational administration*. New York, NY: Macmillan.

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review*, *24*, 377–392.

Crosnoe, R. (2011). *Fitting in, standing out: Navigating the social challenges of high school to get an education*. Cambridge, UK: Cambridge University Press.

Crum, K. S., & Sherman, W. H. (2008). Facilitating high achievement: High school principals' reflections on their successful leadership practices. *Journal of Educational Administration*, *46*, 562–580.

Curry, M. (2008). Critical friends groups: The possibilities and limitations embedded in teacher professional communities aimed at instructional improvement and school reform. *Teachers College Record*, *110*, 733–774.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, *93*(6), 8–15.

Friedkin, N. E., & Slater, M. R. (1994). School leadership and performance: A social network approach. *Sociology of Education*, *67*, 139–157.

Gates Foundation. (2013). *Measures of effective teaching (MET)*. Retrieved from http://www.gatesfoundation.org/united-states/Pages/measures-of-effective-teaching-fact-sheet.aspx

Goldstein, J. (2004). Making sense of distributed leadership: The case of peer assistance and review. *Educational Evaluation and Policy Analysis*, *26*, 173–197.

Grubb, W. N., & Flessa, J. J. (2006). "A job too big for one": Multiple principals and other nontraditional approaches to school leadership. *Educational Administration Quarterly*, *42*, 518–550.

Hamilton, L. S., Stecher, B. M., Russell, J. L., Marsh, J. A., & Miles, J. (2008). Accountability and teaching practices: School-level actions and teacher responses. In B. Fuller, M. K. Henne, & E. Hannum (Eds.), *Strong states, weak schools: The benefits and dilemmas of centralized accountability* (Research in the Sociology of Education, Vol. 16, pp. 31–66). St. Louis, MO: Emerald Group.

Harris, D. N., & Herrington, C. D. (2006). Accountability, standards, and the growing achievement gap: Lessons from the past half century. *American Journal of Education*, *112*, 209–238.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Oxon, UK: Routledge.

Herman, R., & Stringfield, S. (1997). *Ten promising programs for educating all children: Evidence of impact*. Arlington, VA: Education Research Service.

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

Lachat, M. A., & Smith, S. (2005). Practices that support data use in urban high schools. *Journal of Education for Students Placed at Risk*, *10*, 333–349.

Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research 1996-2005. *Leadership and Policy in Schools*, *4*, 177–199.

Leithwood, K., Harris, A., & Strauss, T. (2010). *Leading school turnaround: How successful leaders transform low-performing schools*. San Francisco, CA: Jossey-Bass.

Lewis, A. (2008). *Add it up: Using research to improve education and minority students*. Washington, DC: Poverty and Race Research Action Council.

Lipham, J. (1964). Organizational character of education: Administrative behavior. *Review of Educational Research*, *34*, 435–454.

Lortie, D. (1975). *School-teacher: A sociology study*. Chicago, IL: University of Chicago Press.

Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School Effectiveness and School Improvement*, *21*, 315–336.

Louis, K. S., Leithwood, K., Wahlstrom, K. L., & Anderson, S. E. (2010). Investigating the links to improved student learning: Final report of research findings. University of Minnesota, USA. Retrieved from http://www.cehd.umn.edu/carei/Leadership/Learning-from-Leadership_Final-Research-Report_July-2010.pdf

May, H., & Supovitz, J. A. (2011). The scope of principal efforts to improve instruction. *Educational Administration Quarterly*, *47*, 332–352.

McCaffrey, D. L., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, *4*, 572–606.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*, 13–23.

Meyer, J. W., & Rowan, B. (1975, August). *Notes on the structure of educational organizations: Revised version*. Paper presented at the annual meeting of the American Sociological Association, San Francisco.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education*, *79*(4), 33–53.

Murphy, J. (1990). Principal instructional leadership. In L. S. Lotto & P. W. Thurston (Eds.), *Advances in educational administration: Changing perspectives on the school.* (Volume 1, Part B). Greenwich, CT: JAI Press.

Murphy, J. (1991). *Restructuring schools: Capturing and assessing the phenomena.* New York: Teachers College Press.

Murphy, J. (2010). *The educator's handbook for understanding and closing achievement gaps.* Thousand Oaks, CA: Corwin.

Murphy, J. (2012). *Homeschooling in America.* Thousand Oaks, CA: Corwin.

Powell, Douglas R. (1991). How schools support families: Critical policy tensions. *Elementary School Journal*, *91*, 307–319.

Printy, S. M. (2008). Leadership for teacher learning: A community of practice perspective. *Educational Administration Quarterly*, *44*, 187–226.

Reyes, P., Scribner, J., & Scribner, A. (1999). *Lessons from high-performing Hispanic schools: Creating learning communities*. New York, NY: Teachers College Press.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*, 417–58.

Rockoff, J., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *American Economic Review*, *100*: 261–266.

Rothstein, J. (2009, January). *Student sorting and bias in value added estimation: Selection on observables and unobservables* (NBER Working Paper, 14666). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w14666

Sanders, W., Ashton, J., & Wright, S. (2005). *Comparison of the Effects of NBPTS-certified teachers with other teachers on the rate of student academic progress*. Washington, DC: U.S. Department of Education and National Science Foundation.

Sanders, W., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville: University of Tennessee Value-Added Research and Assessment Center.

Sebastian, J., & Allensworth, E. (2012). The influence of principal leadership on classroom instruction and student learning: A study of mediated pathways to learning. *Educational Administration Quarterly*, *48*, 626–663.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis research. *Review of Educational Research*, *77*, 454–499.

Supovitz, J. (2008). Instructional influence in American high schools. In M. M. Mangin, & S. R. Stoelinga (Eds.), *Effective teacher leadership: Using research to inform and reform* (pp. 144–162). New York, NY: Teachers College Press.

Supovitz, J. A., & Klein, V. (2003). *Mapping a course for improved student learning: How innovative schools systematically use student performance data to guide improvement*. Philadelphia, PA: Consortium for Policy Research in Education.

Supovitz, J., Sirinides, P., & May, H. (2009). How principals and peers influence teaching and learning. *Educational Administration Quarterly*, *46*, 31–56.

Taylor, E. S., & Tyler, J. H. (2011, March). The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers (NBER Working Paper, No. 16877). Cambridge, MA: National Bureau of Economic Research.

Teddlie, C., & Reynolds, D. (2000). *The international handbook of school effectiveness research*. New York, NY: Falmer Press.

Thompson, Victor. (1961). *Modern organization*. New York, NY: Knopf.

Wahlstrom, K. L., & Louis, K. S. (2008). How teachers experience principal leadership: The roles of professional community, trust, efficacy, and shared responsibility. *Educational Administration Quarterly*, *44*, 458–495.

Weis, L. (1990). *Working class without work: High school students in a de-industrializing economy*. New York, NY: Routledge.

Wenger, E. (2000). Communities of practice and social learning systems. *Organization*, *7*, 225–246.

Witziers, B., Bosker, R. J., & Kruger, M. L. (2003). Educational leadership and student achievement: The elusive search for an association. *Educational Administration Quarterly*, *39*, 398–425.

Wright, S., Horn, S., & Sanders, P. (1997). Classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*, 57–67.

AUTHORS

JOSEPH MURPHY, PhD, is the Frank W. Mayborn chair in Vanderbilt University's Peabody College of Education, 230 Appleton Place, Nashville, Tennessee 37203-5701; *joseph.f.murphy@Vanderbilt.edu.* His research is in the area of school improvement.

PHILIP HALLINGER, EdD, is the Joseph Lau Chair Professor of Leadership and Change at the Hong Kong Institute of Education in Hong Kong; *hallinger@gmail.com*. His research focuses on school leadership, change, and improvement.

RONALD H. HECK, PhD, is a professor of educational administration at the University of Hawaii at Manoa, 1776 University Avenue, Honolulu, Hawaii 96822; *rheck@hawaii.edu*. His research focuses on educational leadership and policy.

Manuscript received August 16, 2012
Revisions received February 1, 2013, and April 26, 2013
Accepted July 5, 2013