

NEPC REVIEW: MAKING A DIFFERENCE: SIX PLACES WHERE TEACHER EVALUATION SYSTEMS ARE GETTING RESULTS (NATIONAL COUNCIL ON TEACHER QUALITY, OCTOBER 2018)



Reviewed by:

Amy N. Farley and Leah M. Chamberlain
University of Cincinnati

December 2018

National Education Policy Center

School of Education, University of Colorado Boulder
Boulder, CO 80309-0249
(802) 383-0058
nepc.colorado.edu

Acknowledgements

NEPC Staff

Kevin Welner
Project Director

William Mathis
Managing Director

Alex Molnar
Publications Director

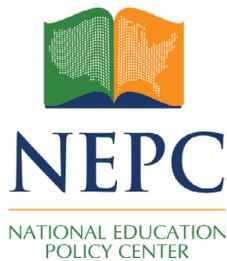
Suggested Citation: Farley, A.N. & Chamberlain, L.M. (2018). *NEPC Review: “Making a Difference: Six Places Where Teacher Evaluation Systems Are Getting Results.”* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-teacher-eval>.

Funding: This review was made possible in part by funding from the Great Lakes Center for Educational Research and Practice.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

This publication is provided free of cost to NEPC’s readers, who may make non-commercial use of it as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.



NEPC REVIEW: MAKING A DIFFERENCE: SIX PLACES WHERE TEACHER EVALUATION SYSTEMS ARE GETTING RESULTS (NATIONAL COUNCIL ON TEACHER QUALITY, OCTOBER 2018)

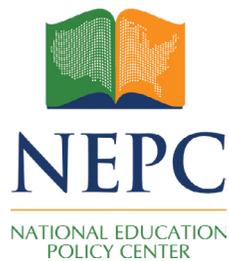
Reviewed by:

Amy N. Farley and Leah M. Chamberlain
University of Cincinnati

December 2018

Executive Summary

A new report from the National Council on Teacher Quality (NCTQ) highlights six teacher evaluation systems claimed to be “yielding substantial benefits.” This comes at the end of a decade when reformed teacher evaluation systems that link teacher performance to measures of student growth have been at the center of educational debate. Disagreements range from the theoretical (e.g., *is teacher quality fundamentally related to inequalities in student outcomes?*) to technical (e.g., *which measures should be included and how should they be defined?*) to practical (e.g., *how should ratings be used for personnel decisions?*). Overall, the research regarding teacher evaluation is mixed, at best. Most notably, a recent multi-year RAND report suggests that a \$500 million investment in teacher evaluation that heavily weighted student growth measures, largely funded by the Gates Foundation, did not improve student outcomes and, in some cases, exacerbated unequal access to effective teachers for low-income students and students of color. The new NCTQ report, while clearly promoting such teacher evaluation, does not seriously counter the groundswell of academic literature critiquing these systems. It does not address the relevant literature, present a compelling justification for its site selection or the inclusion criteria for evidence, or adequately consider disconfirming or contradictory evidence. These methodological flaws limit the validity of the report’s findings and conclusions, which ultimately diminishes the usefulness of the report for policy and practice.



NEPC REVIEW: MAKING A DIFFERENCE: SIX PLACES WHERE TEACHER EVALUATION SYSTEMS ARE GETTING RESULTS (NATIONAL COUNCIL ON TEACHER QUALITY, OCTOBER 2018)

Reviewed by:

Amy N. Farley and Leah M. Chamberlain
University of Cincinnati

December 2018

I. Introduction

Teacher evaluation has been the topic of attention and debate over the past decade. Former evaluation systems were criticized for their inability to differentiate among teachers and for doing little to inform instructional improvement or personnel decisions.¹ Inspired by federal incentives like Race to the Top grant funding and waivers from No Child Left Behind requirements, evaluation reform has become nearly ubiquitous. Nearly all states and many districts² have adopted new or revised teacher evaluation systems, with a focus on student growth measures tied to individual teachers. This rapid adoption has been met with a nearly equal growth in related scholarship.

Numerous converging influences have provoked a backlash to this quick escalation of teacher evaluation reform. First, the 2015 Every Student Succeeds Act relaxed federal control over teacher evaluation.³ At the same time, teachers voiced persistent reservations about revised evaluation systems in general, and student growth measures in particular.⁴ These concerns, as well as the surge of lawsuits challenging evaluation systems across the country,⁵ have driven districts to rethink or revise their adopted models. Finally, it has become increasingly difficult to justify the continued use of measures that attribute student outcomes to individual educators amid warnings from researchers⁶ and professional organizations⁷ regarding their defensibility and reliability.

Another considerable blow to teacher evaluation reform may have been the June 2018 RAND evaluation report, which concluded that Gates' investment in teacher evaluation systems had proven almost entirely unsuccessful.⁸ The RAND report concluded that student outcomes did not improve and low-income and minority students had no greater access to

effective teachers, despite a multiyear, \$500+ million investment. To make matters worse, in at least one system (i.e., Hillsborough County Public Schools), low-income and minority students were *less likely* to have access to highly effective teachers.⁹ Further, old problems remained: very few teachers were classified as ineffective, and effective teachers were no more likely to be retained than before the reform. As a result, the Gates Foundation is no longer investing in teacher evaluation, although Bill Gates has said it remains interested in researching its effects.¹⁰

In October 2018, a new report from the National Council on Teacher Quality (NCTQ) was released. *Making a difference: Six places where teacher evaluation systems are getting results*, written by Hannah Putman, Kate Walsh, and Elizabeth Ross, highlights six places that have “surmounted challenges to implement successful teacher evaluation systems that are yielding substantial benefits.”¹¹ The report concludes that, unlike the sites studied in the RAND report, these six effective evaluation systems have generated improvements in the overall quality of the teacher workforce.

In this review, we summarize the findings and conclusions of the NCTQ report, with special attention to its use of literature and the justifications given for the conclusions it draws. We argue that the report does not sufficiently address the evidence surrounding teacher evaluation and does not adequately consider disconfirming or contradictory evidence. These methodological flaws limit the validity of the report’s findings and conclusions, reducing the usefulness of the report for policy and practice.

II. Findings and Conclusions of the Report

The NCTQ report analyzes six teacher evaluation systems reporting positive results, including four districts (Dallas, Denver, District of Columbia, and Newark) and two states (Tennessee and New Mexico).¹²

The report stresses the importance of evaluation systems that differentiate between varying levels of teacher effectiveness and use that information for personnel decisions. It lauds the six systems for continued progress in several key areas, including an increase in the frequency of evaluation and the number of rating levels used to evaluate teachers; the use of multiple measures, including “objective measures of student growth,”¹³ to determine teacher ratings; and the use of evaluation results in setting compensation. The report also highlights what these six systems have in common, organized by four principles of strong practice: (a) adherence to core principles, (b) a thoughtful approach to weighting individual components, (c) linking evaluation to personnel decisions, and (d) linking evaluation to compensation.

Adherence to core principles

NCTQ praises all six systems for following strong practices. Most notably, they all maintain the use of multiple measures, including student surveys and student growth measures. The

systems also used at least three teacher-rating categories, utilized annual evaluations and observations, and linked professional development to evaluation results.

Thoughtful weighting of individual components

Components of the various evaluation systems included observations, student achievement and/or growth, student surveys, professionalism, commitment to the school community, and school attendance. All systems at a minimum included observations and measures of student achievement/growth. The weight given to student achievement/growth ranged from 15-50%.¹⁴ The weight for student surveys varied from 0-15%.¹⁵ Two of the profiled systems also included a measure of professionalism, and one considered commitment to the school community and one included teacher attendance as part of the overall evaluation rating.

Linking evaluation to personnel decisions

All six systems tailored professional development in response to evaluation results. Some sites also used evaluations to select mentors for student teachers and teachers for leadership opportunities. Others based teacher dismissal decisions on evaluations or provided financial incentives for highly rated teachers who work in high-need schools.

Linking evaluation to compensation decisions

A high evaluation rating leads to a raise, bonus, or award in all six evaluation systems. Five of the six systems also compensate teachers for teaching a hard-to-staff subject or teaching in a high-need school.

Report conclusions

Although the report itself does not clearly delineate a set of conclusions, the NCTQ press release is more forthcoming. It asserts that the six evaluation systems have produced a meaningful and realistic measure of teacher talent. Thus, districts have been able to retain strong teachers and increase the rate at which weaker teachers choose to leave.¹⁶ The press release highlights increases in student proficiency that coincide with implementation of the new evaluation systems, suggesting that those reformed systems are responsible for the improved outcomes.¹⁷ It also suggests that the most significant component of each of the six systems is the link between teacher evaluation results and compensation,¹⁸ a suggestion that does not appear plainly in the full report.

III. The Report's Rationale for Its Findings and Conclusions

The report does not offer an explicit rationale for its conclusions, which are presented pri-

marily in the introduction and accompanying press releases rather than in the text. There is not a clear connection between these conclusions and the data presented in the case studies.

IV. The Report's Use of Research Literature

The majority of references are drawn from the six teacher evaluation systems, and include policy documents, personal communication with stakeholders, and internal evaluations. Beyond those, the report cites news outlets, including *Education Week* and the *Washington Post*; publications from advocacy-oriented organizations, like TNTP and the Association for Supervision and Curriculum Development (ASCD); and research centers like the Center for Economic and Policy Research (CEPR) at Harvard University and the RAND Corporation. Seven endnotes refer to other NCTQ documents or projects.

The report does not sufficiently utilize academic or peer-reviewed work. Despite a robust literature base on teacher evaluation, only nine of the 127 endnotes in the report include unique citations from peer-reviewed journals.¹⁹ Although the report seems to draw on an unstated conceptual framework that prioritizes evaluation, high-stakes testing, accountability, and pay-for-performance, it almost completely ignores the extensive literature regarding each reform, which presents mixed evidence at best. The report inadequately discusses the empirical evidence—or lack thereof—regarding the use of these reforms, the associated effect on teachers or the workforce, or their ultimate impact on student outcomes. In particular, it fails to address the body of work related to the challenges of attributing student growth to teachers, particularly more recent research that presents a less positive picture of evaluation reforms.²⁰ It also does not present any meaningful discussion of the role of evaluation on teacher development and improvement.

V. Review of the Report's Methods

The NCTQ report does not discuss its methodology. This results in two major uncertainties. First, the report does not detail the selection criteria used for the inclusion of districts and states, simply noting that it was able to “identify six pioneers ... [whose] results are setting them apart.”²¹ There is no discussion of how these systems were selected, what other locales were considered, or how the evidence base contributed to those decisions. In the absence of that information, one is left to wonder if NCTQ hand-selected locations based on instincts, anecdotes, or external pressures. Several of these locales have personal connections to NCTQ²² and five of the six have received significant financial support from the Bill and Melinda Gates Foundation,²³ as has NCTQ.²⁴ The lack of information about the selection of these locales significantly hampers the ability to interpret how these systems’ results “set them apart” from anyplace else.

The second major methodological shortcoming is the approach used to synthesize data about the teacher evaluation systems in each district and state. NCTQ seems to imply that its

findings come from its own empirical investigation.²⁵ A closer review suggests that the data presented do not in fact stem from original analyses, but are instead synthesized from “official policies, reviews of the evaluation system that these states and districts have conducted, independent studies, and interviews with district and state staff.”²⁶ However, high-quality research syntheses should seek to “integrate empirical research for the purpose of generalizations.”²⁷ When done well, research syntheses are complex approaches requiring numerous methodological decisions (e.g., how to select relevant studies, code their findings, and analyze their results to synthesize the most salient themes).

Nowhere does the report describe the approach it took for each of these methodological decisions, although we were able to recreate an inventory of the studies and data sources utilized, presented in Table I.²⁸

Table 1: Data Sources Used in NCTQ Analysis

| Locale | Interview/ personal communication | Internal research or other internal documents | External source (not peer-reviewed) | External source (peer-reviewed) |
|--|---|---|--|---|
| Dallas Independent School District (DISD) | Suzy Smith, Manager of Performance Management John Vega, Deputy Chief, Human Capital Management | 2015-16 survey results report ²⁹ 2016-17 TEI evaluation ³⁰ and guidebook ³¹ 2015-16 ACE evaluation ³² * ACE website ³³ | ** Commit! ³⁴ State achievement data ³⁵ ** <i>Dallas News</i> article re: state school accountability ratings ³⁶ | None |
| Denver Public Schools (DPS) | Nicole Wolden, Director of Growth and Performance ³⁷ | * ** press release re: student growth in 2018 ³⁸ | ** CADRE Center evaluation of Denver ProComp ³⁹ * ERS presentation to DPS leadership ⁴⁰ | None |
| District of Columbia Public Schools (DCPS) | Betsy Press, Former Deputy Chief of IMPACT Christopher Lewis, Deputy Chief of IMPACT ⁴¹ | * 2010-11 ⁴² and 2017-18 ⁴³ IMPACT system overview * 2017-18 LIFT guidebook ⁴⁴ * ** Press release re: NAEP results ⁴⁵ * ** FY 2018 salary schedule ⁴⁶ | FutureEd analysis of DCPS teacher reforms ⁴⁷ Bellwether Education’s analysis of DCPS exit survey ⁴⁸ Mathematica analysis of DCPS effectiveness under IMPACT ⁴⁹ * <i>Washington Post</i> article re: IMPACT scores ⁵⁰ ** Nation’s Report Card 2009 vs 2017 NAEP results ⁵¹ | Adnot et al. (2017) article re: teacher turnover under IMPACT ⁵² Dee et al. (2015) article re: teacher retention/performance under IMPACT ⁵³ |
| Newark Public Schools (NPS) | Larisa Shambaugh, Former Chief Talent Officer ^{54, 55} | 2016 state-operated school district annual report to the NJ State Board of Education ⁵⁶ | AIR 2016 external evaluation: NPS teacher contract ⁵⁷ CEPR 2017 study: NPS reform effect on achievement ⁵⁸ MarGrady 2017 analysis of NPS achievement trends ⁵⁹ | None |

| Locale | Interview/ personal communication | Internal research or other internal documents | External source (not peer-reviewed) | External source (peer-reviewed) |
|--|---|---|--|---|
| New Mexico Public Education Department (NMPED) | Matt Montañó, Former Deputy Cabinet Secretary, Teaching and Learning Christopher Eide, Director, Educator Quality ⁶⁰ | * 2017 ESSA plan ⁶¹ * 2018 NMTEACH results ⁶² * Press release re: 2018 NMTEACH results ⁶³ | <i>Albuquerque Journal</i> article re: alterations to NMTEACH ⁶⁴ <i>Albuquerque Journal</i> article re: PARCC student gains ⁶⁵ | Kraft & Gilmour (2017) article re: the “widget effect” and teacher evaluation distributions ⁶⁶ |
| Tennessee Department of Education (TDOE) | Paul Fleming, Assistant Commissioner of Teachers and Leaders | 2012 teacher evaluation report (year 1) ⁶⁷ 2017 ⁶⁸ and 2018 ⁶⁹ educator survey results TVAAS data ⁷⁰ 2014 research brief: Teacher retention ⁷¹ * **Salary schedule ⁷² | NCTQ state policy yearbook ⁷³ TERA 2018 research brief re: teacher improvement in Tennessee ⁷⁴ <i>Education Week</i> article re: Tennessee state testing ⁷⁵ | None |

* These cited sources are not research reports nor do they present evidence or data regarding the evaluation systems.

** These cited sources do not directly address the evaluation system in a given locale.

This summary makes clear that the majority of district and state profiles relied heavily on interviews with one or two staff members within each agency. In at least one location (Newark Public Schools), NCTQ was not able to collect any data directly from current staff, instead relying exclusively on data from a former staff member, in addition to data drawn from one PowerPoint presentation to the New Jersey Board of Education⁷⁶ and three external reports.⁷⁷ This summary also elucidates how little evidence was gathered from “independent studies”⁷⁸ or external research: With the exception of DC Public Schools and New Mexico, each case study was created without the benefit of any data or evidence drawn from peer-reviewed empirical research.

While our summary of evidence may provide a better understanding of the data included in the report, it does not answer other important questions about NCTQ’s data-gathering approach, such as how thorough or selective this data-gathering was. For example, the reader cannot ascertain whether NCTQ conducted an exhaustive search of all publicly available evidence or if they confined themselves to seeking out evidence related to the report’s key principles. If it is the latter, the evidence and associated conclusions are largely invalid.

VI. Review of the Validity of the Findings and Conclusions

Below, we briefly evaluate the NCTQ report using Cooper and Hedges’ suggestions for sound research syntheses.⁷⁹

The report pays little attention to relevant theory

Good research practice would entail drawing on the relevant prior literature, including scholarship regarding teacher evaluation, professional improvement, and performance pay. However, as noted above, NCTQ does not adequately address or reference the prior literature.

The report relies on an underdeveloped and incomplete analysis of the relevant research, and fails to address disconfirming evidence

In addition to using a narrow set of sources, there are several examples where the report seems to misinterpret or mischaracterize findings, or leave out data that was less than positive. For example, survey results cited within the report suggest that only 41% of teachers in Dallas Independent School District were satisfied with the system.⁸⁰ Although the report praises Dallas because this number represents an increase over prior years' data, a 41% satisfaction rate simply fails to represent the perceptions of teachers in a well-functioning system. The report does not mention that nearly half the teachers cite concerns about the system's fairness.⁸¹ To its credit, the report does highlight a concern among Dallas teachers that the system may disadvantage teachers in lower-performing schools, citing this as an opportunity for improvement in the district. However the report mischaracterizes this finding, claiming that "roughly a quarter of teachers felt the evaluation system was unfair to teachers facing greater challenges in low-performing schools."⁸² In fact, 51% of teachers surveyed about this issue selected it as a concern, although the question was asked of only a subset of teachers. In total, these omissions and mischaracterizations suggest that the report does not take a critical or objective lens regarding the evidence base.

The report makes only oblique references to future research

Although the report acknowledges that "the knowledge base for building a strong evaluation system is still young and is continuously being refined,"⁸³ it does not return to this in its conclusions and fails to identify areas for future research or acknowledge limitations or uncertainties that warrant additional study.

VII. Usefulness of the Report for Guidance of Policy and Practice

Had the report more fully considered the body of evidence surrounding teacher evaluation in each of these locales and presented a more holistic narrative of its successes and challenges, it might have provided a meaningful contribution. Instead, its narrow focus and questionable inclusion criteria make it more misleading than helpful.

Teacher evaluation reform has not yet been sufficiently vetted in the empirical literature

and it has certainly not resulted in a demonstrable effect on student achievement.⁸⁴ The NCTQ report does little to extend that literature. Given the timing of its release immediately following the RAND evaluation, it is worth considering whether the NCTQ publication is a thinly veiled attempt to shift the national discourse away from negative press that characterized teacher evaluation reform as an “expensive experiment,”⁸⁵ “a bust,”⁸⁶ and “one of the biggest failures in K-12 philanthropy.”⁸⁷

While not the focus of this review, it is noteworthy that the RAND evaluation provides more thoughtful guidance for policy and practice than is noted here. It acknowledges that while it is difficult to determine the exact cause of the lackluster results, they may be attributable to the reform’s narrow focus: in order for transformative change to be possible, RAND points out, “many other factors might need to be addressed, ranging from early childhood education, to students’ social and emotional competencies, to the school learning environment, to family support.”⁸⁸ We agree but also encourage policymakers and practitioners to think beyond short-term interventions aimed at improving schools or teachers like those described in the NCTQ report. Instead, real educational change may require greater understanding of how structural inequalities and oppressive systems create and perpetuate opportunity gaps with real consequences for students.⁸⁹

The report does not address teacher development or professional learning in a meaningful way, and instead celebrates the six systems without offering comprehensive empirical evidence of their putative successes. It also ignores the influx of additional funding that enabled these agencies to design and implement the profiled systems, including at least \$75 million dollars in direct grants from the Bill and Melinda Gates Foundation.⁹⁰ This omission is noteworthy because it suggests that even if these reforms are effective, those results may not be possible elsewhere without a similar external investment. Although NCTQ tries to suggest otherwise, perhaps that money would have been better spent investing in teacher and student learning and programs to address persistent opportunity gaps. As it stands, this report does little to enrich an already tired conversation about linking teacher evaluation to student achievement.

Notes and References

- 1 Kraft, M.A., & Gilmour, A.F. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Researcher*, 46(2), 234-249;

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.
- 2 This exact number is somewhat contested in the literature. Most studies show anywhere from 40 to 46 states. Two examples are listed here: Walsh, K., Joseph, N., Lakis, K., & Lubell, S. (2017). *Running in place: How new teacher evaluations fail to live up to promises*. Washington, D.C.: National Council on Teacher Quality.

Steinberg, M.P., & Donaldson, M.L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy*, 11(3), 340-359.
- 3 Croft, M., Guffy, G., & Vitale, D. (2018). *The shrinking use of growth: Teacher evaluation legislation since ESSA*. Iowa City, IA: ACT. Retrieved October 30, 2018, from <https://www.act.org/content/dam/act/unsecured/documents/teacher-evaluation-legislation-since-essa.pdf>;

Pennington, K. & Mead, S. (2016, December). *For good measure?: Teacher evaluation policy in the ESSA era*. Washington, D.C.: Bellwether Education Partners. Retrieved November 5, 2018, from <http://bellwethereducation.org/publication/good-measure-teacher-evaluation-policy-essa-era>
- 4 Balch, R. & Koedel, C. (2014). Anticipating and incorporating stakeholder feedback when developing value-added models. *Education Policy Analysis Archives*, 22(97), 1-17;

Collins, C. (2014). Houston, we have a problem: Teachers find no value in the SAS Education Value-Added Assessment System (EVAAS). *Education Policy Analysis Archives*, 22(98), 1-41;

Hewitt, K.K. (2015). Educator evaluation policy that incorporates EVAAS value-added measures: Undermined intentions and exacerbated inequities. *Education Policy Analysis Archives*, 23(76), 1-49;

Jiang, J.Y., Sporte, S.E., Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, 44(2), 105-116.
- 5 Close, K., & Amrein-Beardsley, A. (2018). Learning from what doesn't work in teacher evaluation. *Phi Delta Kappan*, 100(1), 15-19. Retrieved November 3, 2018, from <http://www.kappanonline.org/learning-from-what-doesnt-work-in-teacher-evaluation/>;

Education Week. (2015). *Teacher evaluation heads to the courts*. Retrieved November 3, 2018, from <https://www.edweek.org/ew/section/multimedia/teacher-evaluation-heads-to-the-courts.html>
- 6 This literature is sizable; the following references provide an introduction to the key concerns: Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. New York, NY: Routledge;

Ballou, D., & Springer, M.G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher*, 44(2), 77-86;

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, 44(2), 132-137;

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

Harris, D.N., & Herrington, C.D. (2015). Editors' introduction: The use of teacher value-added measures in

schools: New evidence, unanswered questions, and future prospects. *Educational Researcher*, 44(2), 71-76;

Raudenbush, S.W. (2015). Value added: A case study in the mismatch between education research and policy. *Educational Researcher*, 44(2), 138-141;

Rowan, B. & Raudenbush, S.W. (2016) Teacher evaluation in American schools. In D.H. Gitomer & C.A. Bell (Eds.), *Handbook of research on teaching (5th ed.)* (pp. 1159-1216). Washington, D.C.: American Education Research Association.

- 7 American Educational Research Association Council (2015, November). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher*, 44(8), 448-452.

American Statistical Association (2014, April). ASA statement on using value-added models for educational assessment. Alexandria, VA: Author. Retrieved January 25, 2017, from <http://www.amstat.org/asa/files/pdfs/POL-ASAVAM-Statement.pdf>

- 8 Stecher, B.M., Holtzman, D.J., Garet, M.S., Hamilton, L.S., Engberg, J., Steiner, E.D., & Chambers, J. (2018). *Improving teaching effectiveness: Final report*. Santa Monica, CA: RAND Corporation.

- 9 Stecher, B.M., Holtzman, D.J., Garet, M.S., Hamilton, L.S., Engberg, J., Steiner, E.D., & Chambers, J. (2018). *Improving teaching effectiveness: Final report*. Santa Monica, CA: RAND Corporation.

- 10 Loewus, L.H. (2017, October 23). Gates ends investment in teacher evaluation: What that means for the field. *Education Week*. Retrieved November 6, 2018, from http://blogs.edweek.org/edweek/teacherbeat/2017/10/gates_ends_investment_in_teacher_evaluation_what_does_it_mean.html;

Will, M. (2018). 'An expensive experiment': Gates teacher-effectiveness program shows no gains for students. *Education Week*. Retrieved November 1, 2018, from <https://www.edweek.org/ew/articles/2018/06/21/an-expensive-experiment-gates-teacher-effectiveness-program-show.html>

- 11 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference

- 12 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference

- 13 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality, p. 1. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference

- 14 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference

- 15 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference

- 16 Gerber, N. (2018). *Strong teacher evaluation systems go hand-in-hand with improved teacher quality*. Washington, D.C.: National Council on Teaching Quality. Retrieved October 30, 2018, from https://www.nctq.org/dmsView/Making_a_Difference_Press_Release_FINAL

- 17 Gerber, N. (2018). *Strong teacher evaluation systems go hand-in-hand with improved teacher quality*. Washington, D.C.: National Council on Teaching Quality. Retrieved October 30, 2018, from https://www.nctq.org/dmsView/Making_a_Difference_Press_Release_FINAL

org/dmsView/Making_a_Difference_Press_Release_FINAL

- 18 Gerber, N. (2018). *Strong teacher evaluation systems go hand-in-hand with improved teacher quality*. Washington, D.C.: National Council on Teaching Quality. Retrieved October 30, 2018, from https://www.nctq.org/dmsView/Making_a_Difference_Press_Release_FINAL
- 19 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference
- 20 Stecher, B.M., Holtzman, D.J., Garet, M.S., Hamilton, L.S., Engberg, J., Steiner, E.D., & Chambers, J. (2018). *Improving teaching effectiveness: Final report*. Santa Monica, CA: RAND Corporation.
- 21 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality, p. 1. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference
- 22 For example, NCTQ's board of directors currently includes F. Mike Miles, the former superintendent of Dallas Independent School District who was largely credited with the development and implementation of the teacher evaluation system there (<https://www.dallasnews.com/news/education/2015/06/23/dallas-isd-chief-mike-miles-announces-resignation-after-contract-changes-rejected>). Former chancellor of education Michelle Rhee—who instituted much of the reform in DC Public Schools—also served on their technical advisory board (https://www.washingtonpost.com/blogs/answer-sheet/post/ravitch-what-is-nctq-and-why-you-should-know/2012/05/23/gJQA7CrlU_blog.html?utm_term=.ef9c9aa604c1).
- 23 According to their publicly available database of funded projects, Gates provided over \$8 million in direct grants to Dallas Independent School District, \$28 million to Denver Public Schools and the Denver Public Schools Foundation, \$7 million to the DC Public Education Fund, and \$7 million to the New Schools Venture Fund to support the development of teacher evaluation reforms and data systems in Newark and Washington D.C. In Tennessee, Gates' investment has been even larger, with a combined total of over \$140 million in education grants awarded to various districts, agencies, and foundations in the state. Of that \$140 million, \$14 million was awarded directly to the Tennessee Department of Education, and another \$15 million to the Tennessee State Collaborative on Reforming Education. Bill and Melinda Gates Foundation. (n.d.). *Grantmaking awarded grants*. Seattle, WA: Author. Retrieved October 15, 2018, from <https://www.gatesfoundation.org/How-We-Work/Quick-Links/Grants-Database>
- 24 NCTQ has also received a significant investment from Gates, with seven grants totaling more than \$15 million.
- 25 The report's website explicitly refers to the report as a study (e.g., "In this study, NCTQ examines evidence of the impact of teacher evaluation in six places...") and the formal press release refers to it as an original analysis (i.e., "Our analysis suggests that moving forward with the teacher evaluation systems presents students and teachers with a huge opportunity"). National Center for Teacher Quality (n.d.). *Making a difference: Six places where teacher evaluation systems are getting results* (web landing page). Washington, D.C.: Author. Retrieved October 30, 2018, from <https://www.nctq.org/publications/Making-a-Difference>;
Gerber, N. (2018). *Strong teacher evaluation systems go hand-in-hand with improved teacher quality*. Washington, D.C.: National Council on Teaching Quality. Retrieved October 30, 2018, from https://www.nctq.org/dmsView/Making_a_Difference_Press_Release_FINAL
- 26 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality, p. 1. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference
- 27 Cooper, H., & Hedges, L.V. (2009). Research synthesis as a scientific process. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis (2nd ed.)*, pp. 3-16. New York,

NY: Russell Sage Foundation, p. 4.

- 28 It is worth noting that sources cited within each profile but which were not used as data sources or pieces of evidence about the teacher evaluation system in question are not included in this table. For example, the Dallas profile cites an ASCD article on the provision of feedback to improve teaching practice; this source is used to buttress the claim that the district's short "spot evaluations" have the power to "save principals time and give them more real-time opportunities to engage with teachers and learn about their strengths and areas for growth" (p. 5). However, the ASCD article does not actually present evidence about Dallas—it instead addresses effective feedback practices in general—so it does not appear in the table.
- 29 Douglas, N.E., & Barton M. (2017). *Evaluation of the 2015-16 Teacher Excellence Initiative (TEI): Survey results report* (EA17-539-2.2). Dallas, TX: Dallas Independent School District, Department of Evaluation and Assessment. Retrieved October 30, 2018, from https://www.dallasisd.org/cms/lib/TX01001475/Centricity/domain/98/evaluation/15-16/finalrpts/EA17-539_TEI_SurveyAAG.pdf
- 30 Barton, M., & Palladino, D. (2018). *2016-17 Evaluation of Teacher Excellence Initiative (TEI)*. Dallas, TX: Dallas Independent School District, Department of Evaluation and Assessment. Retrieved October 30, 2018, from <https://www.dallasisd.org/cms/lib/TX01001475/Centricity/domain/98/evaluation/16-17/finalrpts/EA18-539-2%20Full%202017-18%20TEI%20Report.pdf>
- 31 Dallas Independent School District (2016). *2016-2017 TEI teacher guidebook: Dallas ISD's guide for defining, supporting and rewarding teacher excellence*. Dallas, TX: Dallas Independent School District. Retrieved October 30, 2018, from <https://tei.dallasisd.org/home/resources/>
- 32 Accelerating Campus Excellence (ACE) is a district effort to "incentivize top teachers and principals to relocate to seven of the district's Improvement Required (IR) campuses" (p. 1). Palladino, D.K. (2016). *Evaluation of Accelerating Campus Excellence (ACE) 2015-16*. Dallas, TX: Dallas Independent School District. Retrieved October 30, 2018, from <https://www.dallasisd.org/site/handlers/filedownload.ashx?moduleinstanceid=66303&dataid=88235&FileName=Evaluation%20of%20Accelerating%20Campus%20Excellence%20ACE%202015-16.pdf>
- 33 Dallas Independent School District. (2018). *ACE accomplishments*. Dallas, TX: Author. Retrieved October 30, 2018, from <https://www.dallasisd.org/Page/46766>
- 34 Commit! is a Dallas-based education non-profit that works "to ensure that all DFW students receive an excellent and equitable education that prepares them to flourish in college and career," with three key focus areas, including "improving early childhood education, preparing and retaining effective educators, and increasing college completion rates" (Commit!, 2018). Commit! (2018). *2017-2018 impact report*. Dallas, TX: Author. Retrieved October 30, 2018, from https://commitpartnership.org/uploads/pdfs/Commit_ImpactReport_2018_FINAL-1.pdf
- 35 Commit! (2017). *Overview of Dallas ISD STAAR achievement at "meets" post-secondary standard across various demographics and subjects 2012-2017*. Dallas, TX: Author. (Just district data compared to state data).
- 36 Smith, C. (2018, June). How many Dallas ISD schools failed to meet state accountability? DISD leader reveals an amazing number. *Dallas News*. Retrieved October 30, 2018, from <https://www.dallasnews.com/news/education/2018/06/22/many-dallas-isd-schools-failed-meet-state-accountability-disd-leader-reveals-amazing-number>
- 37 The vast majority of the data presented in the profile of Denver Public Schools (DPS) appears to have been drawn from a single interview—and subsequent personal communication—with Nicole Wolden, the Director of Growth and Performance at DPS. Her interview and other personal communications are cited 10 times in the report, while other references regarding the DPS system are only cited three times. The report also acknowledges Sarah Almy, DPS Executive Director of Talent Management, but the district profile

never references any data or evidence based on an interview with her. It is therefore unclear whether she participated in an interview or simply provided feedback on the final report.

- 38 Denver Public Schools (2018). *DPS students again outpace state in academic growth on 2018 CMAS*. Denver, CO: Author. Retrieved October 30, 2018, from <https://www.dpsk12.org/dps-students-again-outpace-state-in-academic-growth-on-2018-cmas/>
- 39 Atteberry, A., Briggs, D.C., LaCour, S., & Bibilos, C. (2015). *Year 2 Denver ProComp evaluation report: Teacher retention and variability in bonus pay, 2001-02 through 2013-14*. Boulder, CO: Colorado Assessment Design Research and Evaluation (CADRE) Center.
- 40 Education Resource Strategies (ERS) is a national non-profit that works with school, district, and state leaders to “transform how they use resources (people, time, and money) to create strategic school systems” (https://www.erstrategies.org/info/who_we_are). One of their products is the School System 20/20 diagnostic, which they presented to DPS leaders on March 15, 2017. Education Resource Strategies (2017). *Denver Public Schools: Leveraging system transformation to improve student results*. Watertown, MA: Author. Retrieved October 30, 2018, from <https://www.slideshare.net/ERSslides/ers-system-2020-analysis-of-denver-public-schools-march-2017>
- 41 The NCTQ report’s endnotes only reference personal communication—not an interview—with Christopher Lewis, the then current Deputy Chief of IMPACT in DCPS.
- 42 District of Columbia Public Schools (2010). *IMPACT: The District of Columbia Public Schools effectiveness assessment system for school-based personnel: 2010-2011*. Washington, DC: Author. Retrieved October 30, 2018, from https://www.nctq.org/dmsView/Impact_The_District_of_Columbia_Public_Schools_Effectiveness_Assessment_DCPS_Report
- 43 District of Columbia Public Schools (2017). *IMPACT: The District of Columbia Public Schools effectiveness assessment system for school-based personnel: 2017-2018*. Washington, DC: Author. Retrieved October 30, 2018, from <https://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/1%20%20%20Teachers%20Grades%204%20with%20Individual%20Value-Added%20Student%20Achievement%20and%20Student%20Survey%20Data.pdf>
- 44 NCTQ cites the 2017-18 guidebook, but the link appears to be inactive, although the 2018-19 guidebook is available. Citation: District of Columbia Public Schools. (2018). *LIFT: Leadership Initiative for Teachers: 2018-2019*. Washington, DC: Author. Retrieved October 30, 2018, from <https://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/2018-19%20LIFT%20Guidebook.pdf>
- 45 District of Columbia Public Schools (2018, April 20). *DC Public Schools continues steady growth on NAEP for ninth year*. Washington, DC: Author. Retrieved October 30, 2018, from <https://dcps.dc.gov/release/dc-public-schools-continues-steady-growth-naep-ninth-year>
- 46 District of Columbia Public Schools (2017). *FY 2018 ET 15 salary schedule*. Washington, DC: Author Retrieved October 30, 2018, from <https://dcps.dc.gov/sites/default/files/dc/sites/dcps/publication/attachments/ET-15%20FY%2018%20Pay%20Schedule.pdf>
- 47 Toch, T. (2018). *A policymaker’s playbook: Transforming public schools in the nation’s capital*. Washington, DC: FutureEd Georgetown. Retrieved October 30, 2018, from <https://www.future-ed.org/a-policymakers-playbook-for-transforming-teaching/>
- 48 Pennington, K., & Brand, A. (2018). *Retaining high performers: Insights from DC Public Schools’ teacher exit survey*. Washington, DC: Bellwether Education. Retrieved October 30, 2018, from https://bellwethereducation.org/sites/default/files/DCPS%20Retention%20Deck_5.7.18_FINAL.pdf
- 49 Walsh, E., & Dotter, D. (2014). *Longitudinal analysis of the effectiveness of DCPS teachers (No. 40185.533)*.

Washington, DC: Mathematica Policy Research. Retrieved October 30, 2018, from <https://www.mathematica-mpr.com/download-media?MediaItemId=%7B26D6EBC1-A507-4585-82C3-FE350C490EB4%7D>

- 50 Chandler, M.A. (2015, February). IMPACT scores ward-by-ward in D.C. *The Washington Post*. Retrieved October 30, 2018, from https://www.washingtonpost.com/news/local/wp/2015/02/24/impact-scores-ward-by-ward-in-d-c/?utm_term=.f69a4e47b0b5
- 51 The Nation's Report Card (2018). *District of Columbia Student Groups and Gaps Data*. Retrieved October 30, 2018, from https://www.nationsreportcard.gov/profiles/stateprofile/overview/DC?cti=PgTab_GapComparisons&chort=1&sub=RED&sj=DC&fs=Grade&st=MN&year=2009R3&sg=Race%2FEthnicity%3A+White+vs.+Black&sgv=Difference&ts=Cross-Year&tss=2017R3-2009R3&sfj=NP
- 52 Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, 39(1), 54-76.
- 53 Dee, T.S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, 34(2), 267-297.
- 54 Like Denver, much of the data presented in the profile of Newark Public Schools (NPS) appears to have been drawn from a single interview with Larisa Shambaugh, Former Chief Talent Officer at NPS. This interview is cited nine times in the 15 included endnotes. Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, DC: National Council on Teacher Quality.
- 55 Unlike Denver, NCTQ was not able to engage any current NPS staff in their data collection efforts, noting in endnote 72 that "current district leadership did not respond to requests to review the information before publication" (p. 23). Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, DC: National Council on Teacher Quality.
- 56 Newark Public Schools (2016, November). *Newark annual report*. Newark, NJ: Author. Retrieved from <http://www.nj.gov/education/sboe/meetings/2016/November/public/Newark%20Annual%20Report.pdf>
- 57 Fulbeck, E., Citkowicz, M., Hester, C., Manzeske, D., Yisak, M., & Eisner, R. (2016). *Newark Public Schools and Newark Teachers Union teacher contract evaluation: Year 1 report*. Washington, DC: American Institutes for Research. Retrieved October 30, 2018, from https://www.air.org/system/files/downloads/report/Newark-Public-Schools-Teacher-Contract-Evaluation-Year-1-February-2016_rev.pdf
- 58 Chin, M., Kane, T.J., Kozakowski, W., Schueler, B.E., & Staiger, D.O. (2017, October). *Assessing the impact of the Newark education reforms: The role of within-school improvement vs. between-school shifts in enrollment*. Cambridge, MA: Center for Education Policy Research. Retrieved from October 30, 2018, from https://cepr.harvard.edu/files/cepr/files/newark_ed_reform_report_synopsis.pdf
- 59 Margolis, J. (2017, October). Moving up: Progress in Newark's schools from 2010 to 2017. MarGrady Research. Retrieved October 30, 2018, from <http://margrady.com/wp-content/uploads/2017/10/Moving-Up-Progress-in-Newarks-Schools.pdf>
- 60 It appears Christopher Eide, the Director of Educator Quality at the New Mexico Public Education Department, participated in an email exchange with NCTQ in lieu of a full interview; see endnote 97 of Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, DC: National Council on Teacher Quality, pg. 24.
- 61 New Mexico Public Education Department (2017). *New Mexico rising: New Mexico's state plan for the Every Student Succeeds Act*. Santa Fe, NM: Author. Retrieved from <https://www2.ed.gov/admins/lead/account/stateplan17/nmcsa2017.pdf>

- 62 The NCTQ report draws several pieces of evidence regarding system performance from the following NMPED source, which unfortunately appears to have been removed from the NMPED website: New Mexico Public Education Department (2018). *Elevating the teaching profession in New Mexico: 2018 NMTEACH results, statewide overview*. Santa Fe, NM: Author.
- 63 New Mexico Public Education Department. (2018). *New Mexico has more Exemplary and Highly Effective teachers than ever before*. Santa Fe, NM: Author. Retrieved October 28, 2018, from <https://webnew.ped.state.nm.us/news/new-mexico-has-more-exemplary-and-highly-effective-teachers-than-ever-before/>
- 64 Burgess, K. (2017, April). PED alters teacher evaluation system after pushback. *Albuquerque Journal*. Retrieved from <https://www.abqjournal.com/981045/ped-alters-teacher-evaluation-system-after-pushback.html>
- 65 Albuquerque Journal Editorial Board (2018, July). Editorial: Student gains a strong reason to keep PARCC. *Albuquerque Journal*. Retrieved from <https://www.abqjournal.com/1196483/student-gains-a-strong-reason-to-keep-parcc.html>
- 66 Kraft, M.A., & Gilmour, A.F. (2017). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher. *Educational Researcher*, 46(5), 234-249.
- 67 Tennessee Department of Education (2012, July). *Teacher evaluation in Tennessee: A report on year 1 implementation*. Nashville, TN: Author. Retrieved November 1, 2018, from https://www.tn.gov/content/dam/tn/education/reports/rpt_teacher_evaluation_year_1.pdf
- 68 Bailey, I. (n.d.). *Educator insights: Takeaways from the 2017 Tennessee educator survey*. Nashville, TN: Tennessee Department of Education. Retrieved from https://www.tn.gov/content/dam/tn/education/data/data_survey_report_2017.pdf
- 69 Tennessee Department of Education. (2018). *Tennessee educator survey: Statewide results for main survey*. Nashville, TN: Author. Retrieved from <http://educatorsurvey.tnk12.gov/#1/all-districts/all-schools/0>
- 70 Tennessee Department of Education (n.d.). *Data downloads and requests. Additional data. TVAAS Composites. District wide TVAAS*. Retrieved from <https://www.tn.gov/education/data/data-downloads.html>
- 71 Office of Research and Policy Brief (2014, May). *Teacher retention in Tennessee: Are we keeping our best teachers?* Nashville, TN: Tennessee Department of Education. Retrieved from https://www.tn.gov/content/dam/tn/education/reports/rpt_teacher_retention.pdf
- 72 Tennessee Department of Education (2017). *State salary schedule* (Tenn. Code Ann. 49-3-306). Nashville, TN: Author. Retrieved October 30, 2018, from https://www.tn.gov/content/dam/tn/education/cpm/loc_fin/loc_fin_state_min_salary_sched_2017-18.pdf
- 73 National Center for Teacher Quality (2017). *State policy yearbook database: Dismissal*. Washington, DC: Author. Retrieved from <https://www.nctq.org/yearbook/national/Dismissal-79>
- 74 Papay, T.P., & Laski, M.E. (2018). *Exploring teacher improvement in Tennessee: A brief on reimagining state support for professional learning*. Nashville, TN: Tennessee Education Research Alliance. Retrieved October 30, 2018, from https://peabody.vanderbilt.edu/research/tnedresearchalliance/files/Exploring_Teacher_Improvement.pdf
- 75 Herold, B. (2018, May). Online state testing in 2018: Mostly smooth, with one glaring exception. *Education Week*. Retrieved from <https://www.edweek.org/ew/articles/2018/05/09/online-state-testing-in-2018-mostly-smooth.html>
- 76 Newark Public Schools (2016). *Newark annual report*. Newark, NJ: Author. Retrieved November 1, 2018, from <http://www.nj.gov/education/sboe/meetings/2016/November/public/Newark%20Annual%20Report.pdf>

- 77 Fulbeck, E., Citkowicz, M., Hester, C., Manzeske, D., Yisak, M., & Eisner, R. (2016). *Newark Public Schools and Newark Teachers Union teacher contract evaluation: Year 1 report*. Washington, D.C.: American Institutes for Research. Retrieved November 1, 2018, from https://www.air.org/system/files/downloads/report/Newark-Public-Schools-Teacher-Contract-Evaluation-Year-1-February-2016_rev.pdf;
- Chin, M., Kane, T.J., Kozakowski, W., Schueler, B.E., & Staiger, D.O. (2017, October). *Assessing the impact of the Newark education reforms: The role of within-school improvement vs. between-school shifts in enrollment*. Cambridge, MA: Center for Education Policy Research. Retrieved November 1, 2018, from https://cepr.harvard.edu/files/cepr/files/newark_ed_reform_report_synopsis.pdf;
- Margolis, J. (2017, October). *Moving up: Progress in Newark's schools from 2010 to 2017*. MarGrady Research. Retrieved November 1, 2018, from <http://margrady.com/wp-content/uploads/2017/10/Moving-Up-Progress-in-Newarks-Schools.pdf>
- 78 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality, p. 1. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference
- 79 Cooper, H., & Hedges, L.V. (2009). Research synthesis as a scientific process. In H. Cooper, L.V. Hedges, & J.C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis (2nd ed.)*, pp. 3-16. New York, NY: Russell Sage Foundation.
- 80 It appears that this piece of data is only the tip of the proverbial iceberg: a closer examination of Dallas' data suggests this dissatisfaction persists among the most highly rated teachers. Dallas teachers are rated into nine effectiveness levels, including unsatisfactory, progressing (I and II), proficient (I, II, and III), exemplary (I and II), and master. Among the 5,498 teachers in the system, only 54% of the 1,214 teachers rated in the top two proficient and top two exemplary categories reported being satisfied or very satisfied with the system. Barton, M., & Palladino, D. (2018). *2016-17 evaluation of Teacher Excellence Initiative (TEI)*. Dallas, TX: Dallas Independent School District, Department of Evaluation and Assessment. Retrieved October 30, 2018 from, <https://www.dallasisd.org/cms/lib/TX01001475/Centricity/domain/98/evaluation/16-17/finalrpts/EA18-539-2%20Full%202017-18%20TEI%20Report.pdf>
- 81 Barton, M., & Palladino, D. (2018). *2016-17 evaluation of Teacher Excellence Initiative (TEI)*. Dallas, TX: Dallas Independent School District, Department of Evaluation and Assessment. Retrieved October 30, 2018, from <https://www.dallasisd.org/cms/lib/TX01001475/Centricity/domain/98/evaluation/16-17/finalrpts/EA18-539-2%20Full%202017-18%20TEI%20Report.pdf>
- 82 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality, p. 5. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference
- 83 Putman, H., Walsh, K., & Ross, E. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. Washington, D.C.: National Council on Teacher Quality, p. 2. Retrieved October 13, 2018, from https://www.nctq.org/dmsView/NCTQ_Report_-_Making_a_Difference
- 84 Stecher, B.M., Holtzman, D. J., Garet, M.S., Hamilton, L.S., Engberg, J., Steiner, E.D., & Chambers, J. (2018). *Improving teaching effectiveness: Final report*. Santa Monica, CA: RAND Corporation.
- 85 Will, M. (2018). 'An expensive experiment': Gates teacher-effectiveness program shows no gains for students. *Education Week*. Retrieved November 1, 2018, from <https://www.edweek.org/ew/articles/2018/06/21/an-expensive-experiment-gates-teacher-effectiveness-program-show.html>
- 86 Strauss, V. (2018). Bill Gates spent hundreds of millions of dollars to improve teaching. New report says it was a bust. *Washington Post*. Retrieved November 1, 2018, from <https://www.washingtonpost.com/news/answer-sheet/wp/2018/06/29/bill-gates-spent-hundreds-of-millions-of-dollars-to-improve-teaching-new-report->

says-it-was-a-bust/

- 87 Hall, L.S., & Callahan, D. (2018). It's one of the biggest failures yet in K-12 philanthropy. What are the lessons? *Inside Philanthropy*. Retrieved November 1, 2018, from <https://www.insidephilanthropy.com/home/2018/7/54/another-lesson-in-k-12-philanthropy-the-gates-teacher-effectiveness-initiative>
- 88 Stecher, B.M., Holtzman, D.J., Garet, M.S., Hamilton, L.S., Engberg, J., Steiner, E.D., & Chambers, J. (2018). *Improving teaching effectiveness: Final report*. Santa Monica, CA: RAND Corporation, pp. xvii-xlviii.
- 89 For a good discussion of opportunity gaps and their influence on educational outcomes, see Carter, P.L., & Welner, K.G. (Eds.) (2013). *Closing the opportunity gap: What America must do to give every child an even chance*. Oxford: Oxford Press.
- 90 See Notes 23 and 24 for more information. Bill and Melinda Gates Foundation (n.d.). *Grantmaking awarded grants*. Seattle, WA: Author.