

EQUITY IN EDUCATION SERIES

Quest for Quality: An Indicator System for Teaching

Gary Sykes
Courtney Bell
Bhavya Shukla

THE ETS CENTER FOR RESEARCH ON HUMAN CAPITAL AND EDUCATION



Table of Contents

A Missing Link in Our Appreciation of Teaching Quality and Equity..... 1
On Indicators2
Indicators of Health Care3
Federal Measures and the Quest for Quality4
Appendices8
 Appendix A: Indicator Type and Other Details8
 Appendix B: Types of Indicators on Federal Surveys 12
 Appendix C: Source List for Educational Surveys Used in this Report 13
About the Authors 14

This report was written by:

Gary Sykes
Courtney Bell
Bhavya Shukla

The views expressed in this report are those of the authors and do not necessarily reflect the views of the officers and trustees of Educational Testing Service.

Copyright © 2019

Educational Testing Service.

All rights reserved.

ETS, the ETS logo, and MEASURING THE POWER OF LEARNING are registered trademarks of Educational Testing Service (ETS). All other trademarks are the property of their respective owners.

June 2019

ETS Center for Research on
Human Capital and
Education

Research and Development
Educational Testing Service
Rosedale Road Princeton, NJ
08541-0001

The research in this report was made possible in part by a grant from the Spencer Foundation (#201700160). The views expressed are those of the authors and do not necessarily reflect the views of the Spencer Foundation.

A Missing Link in Our Appreciation of Teaching Quality and Equity

Ask any parents what they hope for their children's education and teachers will quickly emerge in the conversation. Whatever else might be happening in schools, parents hope their children will have the good fortune of capable and caring teachers. Research has recently caught up with such common sense: Teachers matter a great deal. In fact, existing research demonstrates that of all the schooling-related resources that account for student academic learning, the quality of teachers matters most.^{1 2}

Parents also recognize that learning is something that takes place over time. They can see the gradual growth of knowledge and skill in their children, the most visible product of teachers' work. Children need not one or two quality teachers during their years of schooling but a steady succession of them whose efforts combine over time to support children's growth and development. The effects of schooling are cumulative, and so are the effects of teachers. So what matters for all children is the trajectory of their learning and development.³

In the best of worlds, all children benefit from a succession of quality teachers, but the sad fact is that only some have such access. Massive inequalities in access to quality teachers is a prominent feature of schooling in America. The children least likely to have quality teachers are those who need them most—children of color, those from low-income families, and those learning to speak English.^{4 5 6}

But this national narrative of inequitable access to quality teachers overlooks a critical part of the story. There is a missing link regarding the connection of students to teachers. From a learning perspective, that critical link is the quality of instruction that students receive. It matters who teachers are,

but it matters even more what they do in their interactions with students and the subjects that students are learning.⁷

Contemporary scholarship on teaching reflects this point. For example, research on observations of teaching shows that the same teacher may provide different levels of teaching quality, depending on the students who are in the room.⁸ An older tradition focusing on teachers is giving way to a much broader emphasis on (1) what teachers do, (2) their interactions with students, (3) the content and goals of education, and (4) specific school and community contexts.⁹ These four features shape the crucible of learning, sometimes referred to as instruction.

Equity enters this portrait of teaching as well. On the one hand, we have clear and compelling evidence about the distribution of teacher qualifications, contributing to a powerful story of inequity.¹⁰ A robust set of indicators related to teacher characteristics has been shown to be inequitably distributed across students, including years of experience, certification status, and others.^{11 12 13 14} On the other hand, we have mapped student learning in a variety of ways, documenting in recent years how uneven and unequal achievement is among student groups.¹⁵ However, in terms of the emerging focus that takes the four critical features into account, we have little corresponding evidence about the distribution of instruction itself.¹⁶ We know that teaching varies in its instructional effectiveness, but we have not mapped how that variation relates to students' varying backgrounds. In the K-12 school context, we have no trustworthy indicators of teaching itself to guide policymaking.¹⁷ In addition to equity concerns, the absence of instructional indicators makes it challenging to improve teaching quality in a systematic fashion.

From a policy perspective, then, we have an inadequate account of the teaching-learning process that relies just on inputs and outcomes. In other spheres of American life, a much richer set of indicators measures and models critical processes. Health care is a prime example, as is regulation of the

economy, but advancements have been made in many fields. We cannot guide the pursuit of educational quality or educational equity absent an understanding of how learning is achieved through the interactions of teachers and students—in the classroom, around worthwhile content.

To address this missing link, we propose the development of indicators of teaching quality. Indicators, as we next describe, must meet certain criteria; they have proven to be an important resource in other fields as we describe in the area of health care. To assess the standing of indicators in education, we analyze existing indicators used in major surveys administered by the federal government. We conclude with thoughts about the future of indicators for teaching.

On Indicators

Indicators are a familiar idea, commonplace today in many walks of life, used by the public and policymakers to inform many kinds of decision-making. The idea of an indicator is simple. It is a statistic or combination of statistics designed to gauge progress toward some important end or outcome. Indicators differ from "mere" statistics in that they serve as the object of policymaking and can provide some valued direction.

For example, in the field of economics, the unemployment rate is a critical indicator of economic functioning. Policymakers attend carefully to this indicator as an outcome of interest and as one gauge of economic and social policies. In the health-care field, hospitals carefully track mortality and morbidity rates in response to treatments, while individual patients might select physicians to perform a surgery based on the number of such procedures the physician has performed. Indicators then can be used by consumers and professionals, individuals and organizations, and policy systems at local, state, and national levels.

Indicators have two broad uses. One is to track the status of some important value. Policymakers want to know, for example,

about trends in student learning in the United States. The National Assessment of Educational Progress (NAEP) provides such vital information about student academic achievement in key subject areas at multiple grade levels. The other purpose for an indicator is to aid in explanation. When multiple indicators are available, the combined information assists in identifying and interpreting trends of importance to policymakers.

Complexities arise when indicators are used for explanation and interpretation because they must rely on theory and evidence of some kind to support a claim about causal links—how "X is related to Y." The stronger this relationship, the greater confidence we have in the indicator.

Good indicators meet certain well-established criteria.¹⁸ Ideally:

- They rest on evidence that the activity monitored by an indicator is related to the production of valued outcomes.
- The collection and reporting of information on the activity does not lead to undesirable responses.
- They are minimally burdensome on those who supply the information.
- They are cost efficient.
- The information categories used are likely to be meaningful for a foreseeable future timeline, for example, the next 15 years.

In light of the growing recognition of the importance of teaching quality and its provision to all students, indicators may be one useful tool for addressing persistent inequalities in U.S. students' outcomes. This brief offers an argument for such indicators drawing on the model used in the health-care field. This case demonstrates the utility of a framework that combines three types of indicators—referred to as "structural," "process," and "outcomes" (described below)—to gauge the overall quality of health care. When we examine the measures of teaching drawn from federal surveys that are designed to collect information at the teacher level, we find almost no measures that track the quality of the instructional process itself.

Structural and outcome measures abound with no corresponding attention to process measures. Our recommendation is to begin exploring prospects for developing such measures that might yield a more comprehensive accounting of educational quality and its provision to all students.

Indicators of Health Care

For many decades, leaders in the health field have worked on defining and measuring quality of care. While disagreement exists, a general model, introduced to the field years ago, has proven durable and is widely used. The "Donabedian model," named after its author, parses the assessment of health care into structure, process, and outcomes.¹⁹ *Structure* denotes attributes of the settings in which care occurs. These may include material, human resources, or organizational features associated with settings. *Process* indicators denote what is actually done in giving and receiving care, including both patient and practitioner activities. *Outcomes* refer to effects of care on health status (e.g., mortality and morbidity rates) while also including changes in patient knowledge and behavior together with their satisfaction with the care provided.

An important point here is that this three-part approach is possible only because, as Donabedian (1988) has written, "good structure increases the likelihood of good process, and good process increases the likelihood of good outcomes. It is necessary, therefore, to have established such a relationship before any particular component or outcome can be used to assess quality" (p. 1745).

There are two sources of validation for measures of quality, associated with a logic of correlation or logic of correspondence. Ideally, criteria and standards for structural, process, and outcome indicators are based on sound, scientifically validated knowledge. Claims for indicators rest on their demonstrated association with valued outcomes, either directly or as mediated by process measures. Indicators must correlate with important factors. A causal logic based

on association or correlation among measures is the desired standard. But in cases where such knowledge is not available, Donabedian writes, criteria and standards "should represent the best informed, most authoritative opinion on any particular subject. Criteria and standards can also be inferred from the practice of eminent practitioners in a community. Accordingly, the criteria and standards vary in validity, authoritativeness, and rigor."²⁰ This second logic, then, rests on a claim of correspondence with best practice and the collective judgment of experts.

This model has provided the basis for a considerable set of indicators in the medical field, applying to a wide range of health professionals and institutions. The field relies on all three kinds of measures because each is important in evaluating cases of different kinds.

Before applying these ideas to teaching, it is important to note that the specific aspects of health quality being tracked and the purpose of the indicator are important. For example, for procedures that are high risk and relatively uncommon, structural measures are the preferred option because the procedures are performed too infrequently to support direct outcome measurement. In such cases, procedure volume—a structural measure—is likely the only practical quality indicator.

In cases of low risk and frequent health outcomes, process measures and patient-centered outcomes have the greatest advantages, typically risk-adjusted measures of morbidity and mortality. For such procedures, volume and other structural measures are not known to be major determinants of outcomes. Quality is best judged by process measures (where available) and by outcome measures such as functional health status that extend beyond morbidity and mortality rates.

The knowledge base for indicators in medicine is robust and steadily expanding. Indicators have proven to be a useful tool for managing health care, even as they involve limitations of various kinds and cannot be

applied with equal validity to all aspects of medical care. All three types of indicators have value and all are used in evaluating health care.

This brief excursion into the medical field prompts the question whether parallel possibilities present themselves in teaching. An analysis of one significant sample of teaching indicators demonstrates where the emphasis has been placed among the three types of indicators—and where gaps appear that may be leveraged to address the instructional inequalities across our nation's classrooms.

Federal Measures and the Quest for Quality

Given its utility in medicine, we now apply the three-part logic of "structure-process-outcome" to the practice of teaching. The federal government spends thousands of dollars every year collecting important indicators of teaching and learning through questionnaires of various kinds. This represents the most systematic national data we have on teaching. To determine whether and how these federal measures provide insight into the process of teaching quality, we collected and analyzed the primary ongoing measures the federal government employs to track trends in relation to teachers and teaching: teacher surveys administered by the National Center for Education Statistics (NCES).²¹

In the United States, the states have constitutional responsibility for education, coupled with a strong tradition of local control. States have responsibility for

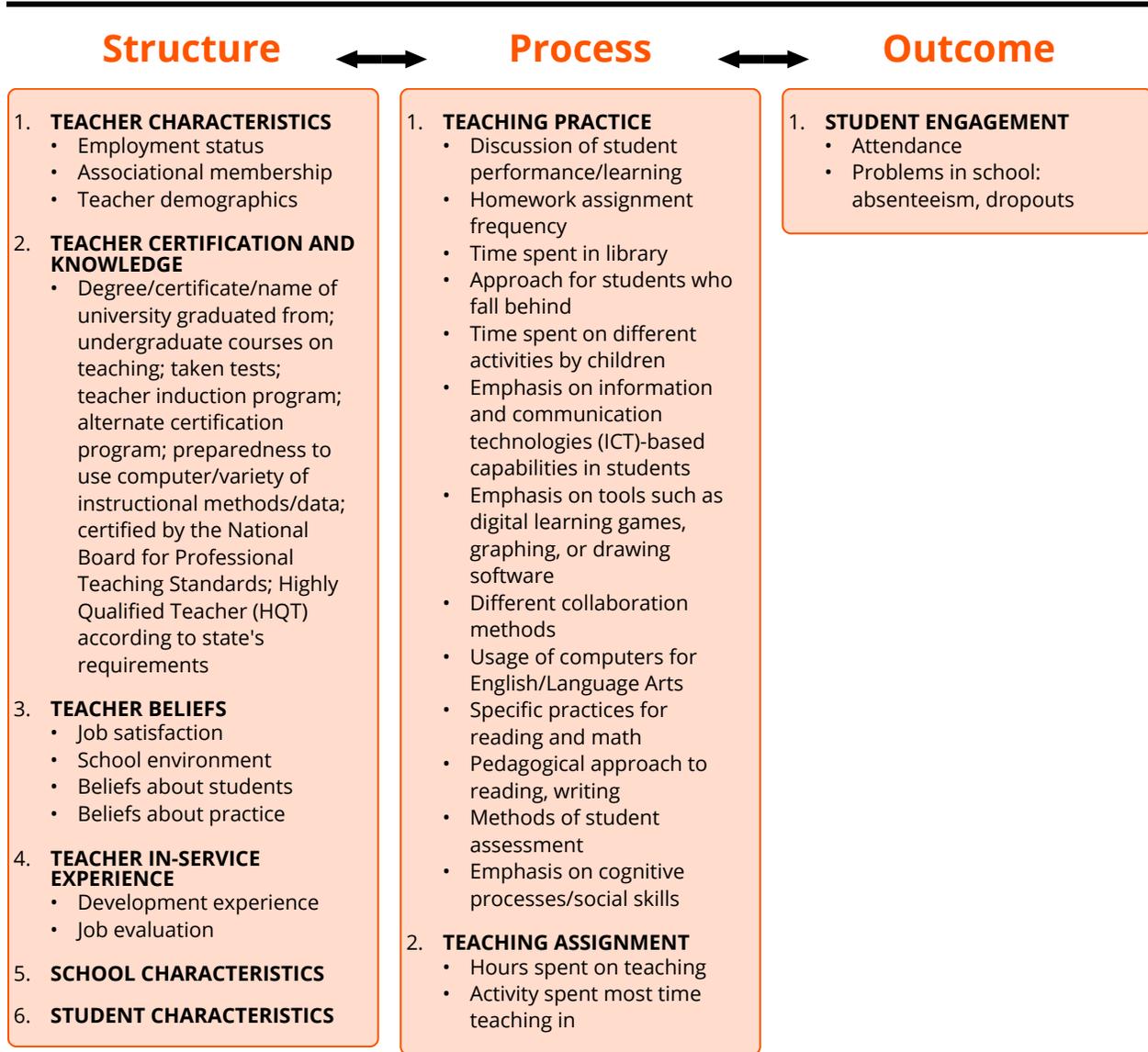
managing their teacher workforce through the labor market, determining education funding formulas, and teacher evaluation efforts, while the federal government has the mandate to conduct educational research, which has included the sponsorship of large-scale surveys and other information-gathering activities. Since there is no systematic way to look across state-level data except by collecting it state by state, we focus on a federal sample of surveys because it is the data that policymakers most likely can use to learn in a timely, systematic way about teaching quality across the nation.

Surveys of teachers accompany many of the large-scale assessments of student achievement that the NCES has sponsored, in addition to one standalone survey of teachers and principals, the National Principal and Teacher Survey (hereafter, we refer to questionnaires administered to teachers as "surveys" and the programs collecting information on student achievement as "assessments.") We downloaded these surveys from the NCES website, then reviewed them to form our initial categories. See Appendix C on the NCES website for the complete list.

In examining these potential sources for teaching quality indicators, we adopted the definitions of the three kinds of measures from the health-care field to explore this question: What proportion of items might be classified as structure, process, or outcome indicators?

Figure 1 summarizes the distinctions we used to organize the analysis of constructs and items included in this sample of surveys sponsored by the NCES.

Figure 1: Sorting and Classification of Indicators from Federal and International Surveys



Note: Names of constructs are numbered and listed in boldface.

The process of classifying items based on these distinctions took place in several steps. First, we organized the items in these surveys according to hypothesized constructs. For example, the construct identified as "teacher characteristics" included items that asked teachers for their employment status, demographic descriptors, and participation in teacher organizations, such as one of the two major teacher unions.

Then, one member of our team reviewed and classified all of the items as either structural, process, or outcome. To provide a check of inter-rater agreement, a second member of our team classified all of the items according to our distinctions. This yielded agreement on 99 percent of the items. Disagreements were reconciled and final classifications reflect that reconciliation.

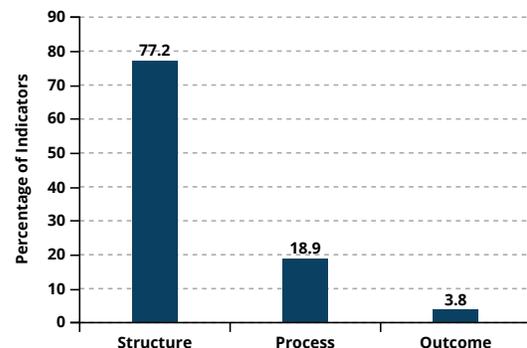
Across all of the surveys, we identified 79 items that measure 9 constructs. Figure 2 presents the results. Of the 79 items in this sample, 61 were structural, 15 were process, and only 3 were outcomes, featuring certain nonacademic measures. Appendix A arrays these items by indicator type and rationale for inclusion.

We offer several observations about these federal surveys. First, measures of academic achievement outcomes are most prominent in the large-scale assessments sponsored by the NCES. Based on such projects as NAEP, PISA (Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study), and others, we have a rich body of knowledge about the academic outcomes of schooling that provides for a range of comparisons with respect to subject matter, jurisdictions (e.g., states), and student groups. In terms of the Donabedian model, then, there has been considerable emphasis on the outcomes of teaching in terms of academic learning. At the same time, these projects limit attention to academic outcomes that can be measured at scale. We know much less about a broader range of outcomes that encompass, for example, social and emotional aspects of learning.

Structural aspects of teaching have also received attention in the teacher surveys. Indicators such as years of experience, certification status, out-of-field teaching, and others are valuable to track, serving as weak proxies for student learning. But as we hypothesized, there are few measures of process with respect to teaching. Figure 2 reveals that over three quarters of all items on the teacher surveys referred to some structurally related factors. Appendix B provides the number and percent for each of the items.

There are also very few outcome measures on these surveys. This is sensible because outcome measures of instructional quality usually come in the form of assessments of student learning. These are most often measures of academic achievement, frequently assessed on cyclical, federal, large-scale student data collections such as NAEP and PISA.

Figure 2: Distribution of Federal Potential Indicators According to Donabedian Model



In terms of the Donabedian model, then, constructs and items from these teacher surveys include minimal attention to teaching process, confirming a missing link between structural or input features and the measures of academic achievement included on the large-scale assessments.

A range of considerations explains this absence. Locating indicators of teaching process that meet the criteria set out above will offer challenges. Direct observation is one option but carries with it the obvious problems of expense, intrusiveness, and

technical adequacy. Other indicators based, for example, on teacher and student surveys may be weak proxies for instructional quality. Forming a consensus on a single set of measures that bears some predictive relation to student outcomes will surely be both a political and scientific challenge. But without process indicators, we lack crucial information of use in educational policymaking. We need to track trends in student achievement over time and in comparison with other countries, but we also need information to interpret the trends and to assist in explaining patterns in the outcome data. At present, such information is severely limited.

There have been prior efforts to identify a potential set of measures, but none recent enough to capitalize on the contemporary growth of knowledge about teaching effectiveness. We argue, then, that renewed attention to indicators of teaching quality is urgent, particularly from the perspective of equity. Already, a significant body of evidence, oriented largely to teacher qualifications, has demonstrated how inequitable the distribution of teachers is, to

the detriment of the many historically underserved students in our country. The real question, though, concerns whether all students are receiving high-quality instruction. It's the teaching that matters. What matters is what happens between teachers and students in pursuit of the crucial goals of learning as they encompass knowledge and skills, as well as social and emotional outcomes.

We conclude by issuing a call for work on this most critical policy issue despite its challenge. We believe a good start can be made based on recent advances in our understanding of teaching effectiveness and of methods for its measurement. Once a set of such indicators has been developed and trialed, it may be progressively refined as our knowledge of teaching continues to grow. Most importantly, though, such indicators will provide tangible commitment to the core ideals undergirding the American Dream—that all children be provided with the intellectual means for achieving a good life and contributing to our democracy. We need an enriched set of policy tools to pursue this ideal, and now is the time to start.

Appendix A: Indicator Type and Other Details

Abbreviations: CCD: Common Core of Data; ECLS: Early Childhood Longitudinal Study; HSLS: High School Longitudinal Study; ICILS: International Computer and Information Literacy Study; MGLS: Middle Grade Longitudinal Study; NAEP: National Assessment of Educational Progress; NTPS: National Teacher and Principal Survey; PIRLS: Progress in International Reading Literacy Study; PISA: Programme for International Student Assessment; SASS: Schools and Staffing Survey; TIMSS: Trends in International Mathematics and Science Study

Table A-1: Indicator Type, Associated Constructs, Specific Items, and Rationale

Indicators	Data Sources
Structure (This includes the attributes of material resources, human resources, and organizational structure.)	
Teacher Characteristics: Rationale/Logic — Aspects of teachers' working conditions pattern affect their work and interactions with students and her professionals.	
1. Employment Status	
Regular/full-time/part-time teacher	SASS, PISA, ECLS
Teacher experience/years worked	SASS, NTPS
Field/subject area taught in	SASS, NTPS
2. Associational Membership	
Member of teacher's union	SASS, NTPS
3. Teacher Demographics	
Gender	SASS, NTPS, ECLS, HSLS, CCD, ICILS
Marital status	SASS, NTPS, ECLS, HSLS, CCD, ICILS
Race	SASS, NTPS, ECLS, HSLS, CCD, ICILS
Age	SASS, NTPS, ECLS, HSLS, CCD, ICILS
Teacher Certification and Knowledge: Rationale/Logic — A teacher's knowledge and preparation experience influence his or her interactions in classrooms with students, thereby contributing to the processes and outcomes of teaching practice.	
Degree	SASS, NTPS, PISA, ECLS, HSLS, MGLS, CCD
Certificate	SASS, NTPS, PISA, ECLS, HSLS, MGLS, CCD
Name of university	SASS, NTPS, PISA, ECLS, HSLS, MGLS, CCD
Graduate/undergraduate courses on teaching	SASS, NTPS
Taken tests (Praxis I/II)	SASS, NTPS
Teacher induction program	SASS, NTPS
Alternate certification program	SASS, NTPS
Preparedness to use computer	SASS, NTPS
Variety of instructional methods	SASS, NTPS
Use of data	SASS, NTPS
Certified by the National Board for Professional Teaching Standards	SASS, NTPS
Highly qualified teacher (HQT) according to state's requirements	SASS, NTPS

Table A-1: Indicator Type, Associated Constructs, Specific Items, and Rationale (Cont.)

Indicators	Data Sources
Teacher Beliefs:	
Rationale/Logic — Teachers' beliefs about their jobs, their schools, students, and the specific content they teach shape how they interact with students as well as the school community, thereby influencing both process and outcome indicators.	
1. Job Satisfaction	
Feelings about profession	SASS, NTPS, NAEP, PISA, ECLS
2. School Environment	
Teacher influence	SASS, NTPS, ECLS
School support	SASS, NTPS, ECLS
School climate	ECLS
Belief about other teachers/principals	MGLS
Instructional leadership	MGLS
3. Beliefs about Students	
Teachers' expectations for student achievement	TIMSS
Teaching limitations	HSLs
Opinions about student learning	HSLs
Family background and student learning	HSLs
4. Beliefs about Curriculum and Instruction	
Material, curriculum	TIMSS
Different assessment strategies	TIMSS
Influence over teaching techniques	ECLS
Teacher In-Service Experience:	
Rationale/Logic — Supports for teacher learning directly enhance the human resources directed to instructional quality and interactions with students.	
1. Professional Development Experience	
Participation in professional development activities	SASS, NTPS, ICILS
Master/mentor teacher assigned by school	SASS, NTPS
Hours spent in workshops/seminars	ECLS
Sharing of teaching experience, improving topic	PIRLS
2. Job Evaluation	
Frequency of informal evaluation	SASS, NTPS
Salary and Benefits:	
Rationale/Logic — Compensation-related items provide one form of incentive that motivates teachers' work.	
Base salary/total salary	CCD
Outside sources of income	SASS, NTPS
Retirement pension check paid from a teacher retirement system	SASS, NTPS
Reimbursement for travel expense	SASS, NTPS
Hours paid to deliver instruction	SASS, NTPS
Hours required to work to receive base pay	SASS, NTPS
Additional compensation from school for extracurricular activities	SASS, NTPS
Additional compensation based on student performance	SASS, NTPS

Table A-1: Indicator Type, Associated Constructs, Specific Items, and Rationale (Cont.)

Indicators	Data Sources
School Characteristics: Rationale/Logic — School characteristics identify features of the context that influence aspects of teachers' work, including motivation, capacity, supports, and interactions with students.	
Public/private school	SASS, NTPS
School system offers tenure	SASS, NTPS
School capacity	PISA, ECLS
Teacher observation	PISA
Teacher collaboration	PISA
Teacher engagement	PISA
Assistance from paid aides	ECLS
Physical education involvement	ECLS
Aspects of remedial help in school	HSLs
Class size	MGLS
Student ever threatened to injure/physically attack	SASS, NTPS
Student Characteristics: Rationale/Logic — Characteristics of students exert individual and collective influence on interactions among students and with teachers.	
Students in individualized education program (IEP)	SASS, NTPS
Limited English proficiency or English language learners (ELLs)	SASS, NTPS
Percentage of students not prepared for material	HSLs
Average number of students taught	SASS, NTPS
Process (Denotes what is actually done in providing instruction)	
Teacher Practices: Rationale/Logic — These practices occur inside classrooms between and among teachers, students, and content.	
Discussion of student performance/learning	NAEP
Homework assignment frequency	TIMSS
Time spent in library	PIRLS
Approach for students who fall behind	PIRLS
Time spent on different activities by children	ECLS
Emphasis on information and communication technologies (ICT)-based capabilities in students	ICILS
Emphasis on tools, such as digital learning games, graphing, or drawing software	ICILS
Different collaboration methods	NAEP
Usage of computers for English/Language Arts	SASS, NTPS, ECLS
Specific practices for reading and math	PIRLS, TIMSS, MGLS
Pedagogical approach to reading or writing	NAEP
Methods of student assessment	NAEP, ECLS
Emphasis on cognitive processes/social skills	NAEP

Table A-1: Indicator Type, Associated Constructs, Specific Items, and Rationale (Cont.)

Indicators	Data Sources
Teaching Assignment: Rationale/Logic — These measures supply broad characterization of interactions between teachers and students in classrooms.	
Hours spent on teaching	SASS, NTPS, PIRLS, TIMSS
Activity spent most time teaching in	ECLS
Outcome (Denotes the effect of education on the status of students and population)	
Student Engagement: Rationale/Logic — This group reflects the student outcome measures contained in the questionnaires that may be influenced by various structural and process measures in a school/community.	
Attendance	MGLS
Absenteeism	SASS, NTPS
Dropouts	SASS, NTPS

Appendix B: Types of Indicators on Federal Surveys

Table B-1: Structure, Process, and Outcome Indicators on Federal Surveys (Number and Percent)

Structure	No. of Indicators	Process	No. of Indicators	Outcome	No. of Indicators
Teacher Characteristics:	8 (10.1%)	Teaching Practice:	13 (16.5%)	Student Engagement:	3 (3.8%)
1. Employment Status	3 (3.8%)				
2. Associational Membership	1 (1.3%)				
3. Teacher Demographics	4 (5.1%)				
Teacher Certification and Knowledge:	12 (15.2%)	Teaching Assignment:	2 (2.5 %)		
Teacher Beliefs:	13 (16.5%)				
1. Job Satisfaction	1 (1.3%)				
2. School Environment	5 (6.3%)				
3. Beliefs about Students	4 (5.1%)				
4. Beliefs about Practice	3 (3.8%)				
Teacher In-Service Experience:	5 (6.3%)				
1. Professional Development Experience	4 (5.1 %)				
2. Job Evaluation	1 (1.5 %)				
Salary and Benefits:	8 (10.1%)				
School Characteristics:	11 (13.9 %)				
Student Characteristics:	4 (5.1 %)				
Total:	61 (77.2 %)		15 (18.9%)		3 (3.8 %)

Note: Each indicator item was counted once, including cases where the same or similar indicator appeared in more than one of the surveys included in the sample. To calculate percentages, each item was divided by the total number of items, which was 79.

Appendix C: Source List for Educational Surveys Used in this Report

The following ongoing surveys and questionnaires were used in creating this report.

National (retrieved from NCES website)

- Common Core of Data Teacher Compensation Survey: School-Level Public-Use Data File [Measurement instrument], 2007. https://nces.ed.gov/ccd/pdf/tcs_sch_2006071a_gen.pdf
- Early Childhood Longitudinal Study: Fifth Grade Teacher Questionnaire: Kindergarten Class of 1998-1999 [Measurement instrument], Spring 2004. <https://nces.ed.gov/ecls/pdf/fifthgrade/teacherquexform.pdf>
- High School Longitudinal Study [Measurement instrument], 2009. https://nces.ed.gov/surveys/hsls09/pdf/2009q_teacher.pdf
- National Assessment of Educational Progress: Reading Teacher Questionnaire Grade 8 [Measurement instrument], 2011. https://nces.ed.gov/nationsreportcard/subject/tdw/pdf/instruments/BQ11_Teacher_Read_G08.pdf
- Middle Grade Longitudinal Study: Mathematics Teacher Survey [Measurement instrument], 2017. <https://surveys.nces.ed.gov/MGLS/Images/MathematicsTeacherSurvey.pdf>

- National Teacher and Principal Survey: Teacher Questionnaire [Measurement instrument], 2016. https://nces.ed.gov/surveys/ntps/pdf/1516/Teacher_Questionnaire_2015-16.pdf
- Schools and Staffing Survey: Teacher Questionnaire, 2012. <https://nces.ed.gov/surveys/sass/pdf/1112/SASS4A.pdf>

International

- International Computer and Information Literacy Study (International Association for the Evaluation of Educational Achievement). Teacher Questionnaire for the Main Study, 2018. https://nces.ed.gov/surveys/icils/pdf/icils2018ms_teacher_questionnaire_en_us.pdf
- Programme for International Student Assessment: General Teacher Questionnaire – English [Measurement instrument], 2015. https://nces.ed.gov/surveys/pisa/pisa2015/questionnaires/General_Teacher_Q_English.html
- Progress in International Reading Literacy Study: Teacher Questionnaire Grade 4 [Measurement instrument], 2016. https://nces.ed.gov/surveys/pirls/pdf/P16_TQ_final.pdf
- Trends in International Mathematics and Science Study: Teacher Questionnaire Mathematics Grade 8 [Measurement instrument], 2015. https://nces.ed.gov/timss/pdf/2015_8th_grade_Teacher_Questionnaire_Math.pdf

About the Authors

Gary Sykes is a Principal Research Scientist in ETS's Student and Teacher Research Group, where he is responsible for investigating a wide range of measures, instruments, and procedures used to assess and improve teaching quality. His interests focus particularly on issues of equity in teaching; policies related to teachers, teaching, and teacher education; and efforts to enhance the teaching profession.

Courtney Bell, Principal Research Scientist in ETS's Center for Global Assessment, is a former high school science teacher and teacher educator. Her teaching in rural North Carolina convinced her to study the intersections of research, policy, and practice. Her studies use mixed methods to analyze teaching, teacher education, and the validity of teaching quality measures, especially observational measures. Recent studies investigate how administrators learn to use a

high-stakes observation protocol, how to understand the validity of observational measures, how measures of teaching compare across countries, and the ways in which observation protocols capture high-quality teaching for students with special needs.

Bhavya Shukla, Senior Research Assistant at ETS, holds a master's degree in Education and Social Policy from New York University. She has primarily worked in the education sector in India on school leadership development programs and a randomized control study on measures for developing intrinsic teacher motivation for secondary school teachers. Her interests lie particularly in equity in K-12 education. Some of her more recent work includes projects on developing an indicator system to evaluate teaching quality in the United States, a special education teacher framework, and global assessment of classroom teaching and learning.

Endnotes

Studies that estimate the relative importance of school-related resources to student achievement report that teacher effects are centrally important. See Daniel Aaronson, Lisa Barrow, and William Sander, "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics* 25 (2007): 95–135, <https://faculty.smu.edu/millimet/classes/eco7321/papers/aaronson%20et%20al.pdf>; Raj Chetty, John N. Friedman, and Jonah E. Rockoff, *Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood* (NBER Working Paper No. w19424), (Cambridge, MA: National Bureau of Economic Research, 2013), <https://www.nber.org/papers/w19424>; Barbara Nye, Spyros Konstantopoulos, and Larry V. Hedges, "How Large Are Teacher Effects?" *Educational Evaluation and Policy Analysis* 26 (2004): 237–257; Steven G. Rivkin, Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica* 73 (2005): 417–458, <https://econ.ucsb.edu/~jon/Econ230C/HanushekRivkin.pdf>; and Jonah E. Rockoff, "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," *American Economic Review* 94 (2004): 247–252, <https://doi.org/10.1257/0002828041302244>.

Gauged against the other in-school investments researchers have studied, such as reducing class sizes, teachers constitute the single largest within-school influence on student achievement. See Rivkin, Hanushek, and Kain, "Teachers, Schools."

Studies have demonstrated the cumulative impact of teachers extending over multiple years. Teacher effects cumulate over the years. Positive teacher effects in later years cannot compensate for weak teacher effects in the early years of schooling. See William L. Sanders and June C. Rivers, *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement* (Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center, 1996), <https://www.beteronderwijsnederland.nl/files/cumulative%20and%20residual%20effects%20of%20teachers.pdf>; William L. Sanders and Sandra P. Horn, "Research Findings from the Tennessee Value-Added Assessment System (TVAAS) Database: Implications for Educational Evaluation and Research," *Journal of Personnel Evaluation in Education* 12, no. 3 (1998): 247–256; and Se Woong W. Lee, "Pulling Back the Curtain: Revealing the Cumulative Importance of High-Performing, Highly Qualified Teachers on Students' Educational Outcomes," *Educational Evaluation and Policy Analysis* 40, no. 3 (2018): 359–381, <https://doi.org/10.3102/0162373718769379>.

See for example, Dan Goldhaber, Lesley Lavery, and Roddy Theobald, "Uneven Playing Field? Assessing the Teacher Quality Gap between Advantaged and Disadvantaged Students," *Educational Researcher* 44, no. 5 (2015): 293–307, <https://doi.org/10.3102/0013189X15592622>.

It suggests regarding both teacher experience and the value-added model (VAM) that there is an uneven distribution of teaching in Washington state.

See Hamilton Lankford, Susanna Loeb, and James Wyckoff, "Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis," *Educational Evaluation and Policy Analysis*, 2002, <http://epa.sagepub.com/content/24/1/37>. This paper documents the uneven distribution of teacher experience and qualifications in New York City.

See Dan Goldhaber, Vanessa Quince, and Roddy Theobald, "Has It Always Been This Way? Tracing the Evolution of Teacher Quality Gaps in U.S. Public Schools," *American Educational Research Journal* 55, no. 1 (2018): 171–201, <https://doi.org/10.3102/0002831217733445>. It documents the persistent gaps in teacher experience, test scores on certification exams, and VAM in Washington and North Carolina.

A wide range of studies have measured interactions between teachers and students as a critical input for student achievement. Various measures have been used but for two illustrative studies. See Bridget K. Hamre, Robert C. Pianta, Jason T. Downer, Jamie DeCoster, Andrew J. Mashburn, Stephanie M. Jones, Joshua L. Brown, Elise Cappella, Marc Atkins, Susan E. Rivers, Marc A. Brackett, and Aki Hamagani, "Teaching through Interactions. Testing a Developmental Framework of Teacher Effectiveness in over 4,000 Classrooms," *Elementary School Journal* 113, no. 4 (2013): 461–487, https://pdxscholar.library.pdx.edu/cgi/viewcontent.cgi?article=1005&context=psy_fac; and Lindsey Clare Matsumura, Helen E. Garnier, Sharon Cadman Slater, and Melissa D. Boston, "Toward Measuring Instructional Interactions At Scale," *Educational Assessment* 13, no. 4 (2008): 267–300, <https://doi.org/10.1080/10627190802602541>.

A number of recent reports on observational measures of classroom interactions demonstrate that the same teacher may get different observation scores depending on the race, gender, and prior academic achievement of the students in that specific classroom. See Shanyce Campbell and Matthew Ronfeldt, "Observational Evaluation of Teachers: Measuring More than We Bargained for?" *American Educational Research Journal* 55, no. 6 (2018): 1233–1267, <https://doi.org/10.3102/0002831218776216>.

The point made here is that teaching is not what teachers do but what they do with their students in interactions around the content of instruction as influenced by a range of contextual factors. For treatment of this perspective on teaching, see David K. Cohen and Deborah Loewenberg Ball, *Instruction, Capacity, and Improvement* (CPRE Report Series RR-43), (Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania, 1999), http://www.cpre.org/sites/default/files/researchreport/783_rr43.pdf; David Cohen, Stephen W.

Raudenbush, and Deborah Loewenberg Ball, "Resources, Instruction, and Research," *Educational Evaluation and Policy Analysis* 25, no. 2 (2003): 119–142, <https://doi.org/10.3102/01623737025002119>.

An expanding body of research has explored the processes through which teachers are distributed across states, districts, schools, and within schools. A main conclusion from this work is that students from low-income and minority backgrounds are least likely to be taught by qualified teachers. For a concise review of these studies, see Gary Sykes and Kacy Martin, "Equitable Access to Capable Teachers: The States Respond," *Educational Policy Archives* 27, no. 39 (2019): 1–47, <https://epaa.asu.edu/ojs/article/view/3743>.

Studies providing evidence of the effects of teacher credentials are mixed. Some measures of qualifications or credentials appear more strongly related to student outcomes than others, although this relationship is often conditioned on the subject matter being measured (i.e., stronger relationships with achievement in mathematics than in reading). See Dan D. Goldhaber and Dominic J. Brewer, "Does Teacher Certification Matter? High School Certification Status and Student Achievement," *Educational Evaluation and Policy Analysis* 36, no. 3 (2000): 129–145, <https://doi.org/10.3102/01623737022002129>.

Years of experience matter to some degree. See Tara Kini and Anne Podolsky, *Does Teaching Experience Increase Teacher Effectiveness? A Review of the Research* (Palo Alto, CA: Learning Policy Institute), 2016, https://learningpolicyinstitute.org/sites/default/files/product-files/Teaching_Experience_Report_June_2016.pdf, but master's degrees for the most part do not. See Dan D. Goldhaber and Dominic J. Brewer, "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity," *Journal of Human Resources* 32, no. 3 (1997): 505–523, https://www.jstor.org/stable/146181?seq=1#metadata_info_tab_contents; and Dan D. Goldhaber and Dominic J. Brewer, "Does Teacher Certification Matter? High School Certification Status and Student Achievement," *Educational Evaluation and Policy Analysis* 36, no. 3 (2000): 129–145, <https://doi.org/10.3102/01623737022002129>.

Overall, though, the relationship among credentials and value-added measures of student achievement is moderate. Some studies consequently discount the relationship. See Dan Goldhaber, "Everyone's Doing It, But What Does Teacher Testing Tell Us about Teacher Effectiveness?" *Journal of Human Resources* 42, no. 4 (2007): 765–794, <https://files.eric.ed.gov/full-text/ED509664.pdf>; and Thomas J. Kane, Jonah Rockoff, and D. O. Staiger, "What Does Certification Tell Us about Teacher Effectiveness? Evidence from New York City," *Economics of Education Review* 27 (2007): 615–631, <https://doi.org/10.3386/w12155>.

Others conclude that credentials do matter, at least to some degree. In particular, when credentials are combined rather than being measured singly, the effect is strengthened. See Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor, "Teacher Credentials and Student Achievement: Longitudinal Analysis with Student Fixed Effects," *Economics of Education Review* 26, no. 6 (2007): 673–682, <https://doi.org/10.1016/j.econedurev.2007.10.002>; and Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor, *Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects*, Working Paper 11 (Washington, DC: Center for Analysis of Longitudinal Data in Education Research, 2007): <https://doi.org/10.1353/jhr.2010.0023>. For example, novice teachers with mediocre test scores and substandard credentials tend to produce less achievement than their contrasting counterparts.

One important source of information about gaps in achievement among student groups are the NAEP reports that have been issued over the years. While some gaps in achievement have closed, others have not. Significant gaps remain between minority and white students, and students from low-income households vs. more affluent students. See Sean F. Reardon, "The Widening Academic Achievement Gap between the Rich and Poor: New Evidence and Possible Explanations," in *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, eds. Greg J. Duncan and Richard J. Murnane (New York: Russell Sage, 2011), 91–116; and Jaekyung Lee and Todd Reeves, "Revisiting the Impact of NCLB High-Stakes School Accountability, Capacity, and Resources: State NAEP 1990–2009 Reading and Math Achievement Gaps and Trends," *Educational Evaluation and Policy Analysis* 34, no. 2 (2012): 209–231, <https://doi.org/10.3102/0162373711431604>.

There has been research in the pre-K context that systematically explores instruction across Head Start classrooms. Compare with Tran D. Keys, George Farkas, Margaret R. Burchinal, Greg J. Duncan, Deborah L. Vandell, Weilin Li, Erik A. Ruzek, and Carollee Howes, "Preschool Center Quality and School Readiness: Quality Effects and Variation by Demographic and Child Characteristics," *Child Development* 84, no. 4 (2013): 1171–1190, <https://doi.org/10.1111/cdev.12048>. But it is unclear how such findings map onto the K-12 context given the differences in state requirements for attendance and the common sorting of students to K-12 schools and classrooms.

The observation scores that are a part of many teacher evaluation systems could, in theory, be used to understand the distribution of instruction. However, they have been roundly criticized as inaccurate. See Daniel Weisberg, Susan Sexton, Jennifer Mulhern, and David Keeling, *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, 2009, <https://tntp.org/publications/view/the-widget-effect-failure-to-act-on-differences-in-teacher-effectiveness>.

Researchers are still sorting out whether and how to make sense of them as a trustworthy measure of instruction. See Matthew A. Kraft and Allison F. Gilmour, "Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness," *Educational Researcher* 46, no. 5 (2017): 234–249, <https://doi.org/10.3102/0013189X17718797>; and Chingos, Lindquist, and Whitehurst, *Evaluating Teachers*, 2014.

For discussion of criteria for indicators, see Richard J. Murnane, "Improving Education Indicators and Economic Indicators: The Same Problems?" *Educational Evaluation and Policy Analysis* 9, no. 2 (1987): 101–116, <https://doi.org/10.3102/01623737009002101>.

For discussion of this model, see Avedis Donabedian, "The Quality of Care. How Can It Be Assessed?" *Journal of the American Medical Association* 260, no. 12 (1988): 1743–1748.

Donabedian, "Quality of Care," 1747.

Whether teacher surveys are the best means for developing indicators of teaching is an important question that we do not address here. Should policymakers seek to develop indicators of teaching, the methods for measurement and data collection is a critical question. Our analysis here provides a status report on current efforts, recognizing that future developments may include other measurement strategies.