



Carnegie Foundation for the Advancement of Teaching

ASSESSING TEACHING COLLABORATORY CONVENING

Meeting Summary

Report prepared by: Alicia Grunow, Jeannie Myung, Anthony S. Bryk and Paul LeMahieu

May 24-25, 2011

Carnegie Foundation for the Advancement of Teaching

Funded through a cooperative agreement with the Institute for Education Sciences

Assessing • Teaching •
Improving • Learning

Overview

On May 24th and 25th, 2011, the Carnegie Foundation for the Advancement of Teaching hosted a convening of 30 of the leading experts involved in the technical design of instruments and methodologies for assessing teaching (for a listing of participants, including biographies, see Appendix A). Broadly speaking, these participants represented those working on observational systems for examining teaching as well as those employing student achievement and related data to analyze the unique contributions (the so called “value-added”) of individual teachers to students’ outcomes.

Our objectives were to identify critical technical issues for inquiry and development as well as to explore the initiation of a learning community around these issues. The resulting 1.5 days offered engaging, cross-cutting conversation on the current terrain of teacher assessment in which participants identified areas of settled knowledge, needs and priorities for further inquiry, and ideas for the improvement of practice.

The convening enabled the development of three products, each of which is presented below:

- (1) A framework for organizing the teacher assessment space;
- (2) Version 1.0 of a research agenda; and
- (3) An initial list of activities for Carnegie to pursue.

In addition, the convening succeeded in spurring an interest in continued collaborative work amongst many of the participants. Many participants indicated interest in participating in future activities of the project.

Overview of the Convening Agenda

The convening agenda was organized into three parts (for the full agenda see Appendix B). The first morning was spent generating and prioritizing critical technical questions. The afternoon was dedicated to “deeper dive” conversations into a subset of questions identified as critical by the group. We also explored the different purposes the teacher assessment seeks to serve. The issue of how Carnegie and this group could have the most impact was explored on the final morning of the convening.

A couple of agenda design practices warrant some explanation. First, we invited two “practical implementers” of teacher assessment systems, one representing districts and the second state policy maker voices. Periodically, the participants turned to them for their reflections on the discussion. This served to deepen the conversation and ensure it was focused on the “on-the-ground needs” of those using these systems. Second, we also gave each participant three minutes (interspersed throughout the day) to share some interesting work they thought was particularly pertinent to the topic at hand. These presentations succeeded in providing us with a quick update on some of the cutting-edge

research occurring in the field. They also served to introduce participants' work to each other and engendered informal conversations between participants during breaks and mealtimes. Overall participants' evaluations of the event indicated that the structure of the two days was well suited to fostering productive conversations.

The Potential for Continued Work

The convening succeeded in establishing relationships to support potential future collaborative work. The group collectively recognized the enormity of this technical undertaking and the need for ongoing critical conversations across a diverse set of perspectives. They also validated the need for a neutral convening body to provide a safe space for sustained exploration and interaction around these challenging technical issues, away from the urgency and pressure of practice, political action, and academia. For our part, we also now have a network of experts that are willing to be called upon in future efforts.

A Framework for Organizing Teacher Assessment

During the convening, we introduced and tested a framework for organizing the complex terrain of teacher assessment based on the purpose of assessment and the assessment methodology. We tested this framework with the participants and found it to be a useful guiding framework. It served to both spur and anchor engaging conversation on the design and use of tools and protocols for assessing teacher quality. This framework is presented and described in more detail in what follows.

Purposes of Teacher Assessment Systems

Teacher assessment systems are developed to serve multiple purposes. Most visibly, districts are building teacher assessment systems to evaluate teachers and inform consequential decisions about their employment and compensation. Teacher assessment systems are also intended to support the improvement of teaching. Finally, teacher assessment can be the basis of district research on evaluating its efforts¹. A simple analysis of the uses of teacher assessment systems in the Race to the Top contexts illustrates the three different purposes the current explosions of data systems are intended to support (see Table 1).²

Table 1: Race to the Top States' Plans for Data Use

Evaluation	Improvement	Policy Research
<ul style="list-style-type: none">• Make decisions about promotion• Make decisions about tenure• Make decisions about compensation• Make decisions about teacher certification (re-certification)• Make decisions about teacher interventions• Make decisions about teacher dismissal	<ul style="list-style-type: none">• Provide teachers with relevant information as might guide their own efforts at professional growth and development• Provide differentiated or targeted professional development to teachers• Provide teachers with data in a timely manner so as to inform instruction	<ul style="list-style-type: none">• Evaluate LEA or school professional development systems• Examine/address issues of inequitable distribution of teachers• Evaluate pre-service programs• Help LEAs better understand issues around teacher retention• Allow policy makers to better understand the implications and consequences of various policy actions

¹ The field of healthcare has recognized the discreteness of the three purposes of measurement: evaluation, improvement, and research. See Solberg, Mosser, and McDonald (1997). The Three Faces of Performance Measurement: Improvement, Accountability, and Research. *Journal on Quality Improvement*, (23) 3, 135-145.

² From a summary document of Race to the Top States' plans for data use, distributed at the May 18th-19th, 2011 Department of Education Community of Practice meeting.

Each of these different purposes puts different demands on the technical and practical qualities of the assessment process. The ideal design for teacher assessment systems also will be differentiated by its intended use.

The System of Teacher Assessment: Intersecting Purposes and Measures

More typically, the field of teacher assessment is divided by the methodologies used to assess teachers. One approach to teacher evaluation is to focus on the primary outcomes of the system—namely student learning as demonstrated in some objective measures of academic achievement. Value-added methods apply sophisticated statistical procedures to measure teacher contributions to student learning gains to evaluate teacher effectiveness. A very different approach focuses on information about the classroom practices of teachers. These teacher evaluation tools, typically in the form observational protocols with attendant rubrics, focus on direct observations of teacher behaviors, classroom practices, student work, and the learning environment. States and districts also design the overall systems for information management and use, making decisions about how these sources of information will be combined, taken into consideration and used to support various purposes.

Each purpose and measurement strategy provides its own issues and challenges. When analyzed as an overall system, the unique purposes and measures intersect, creating a matrix for understanding the unique requirements of and technical needs within each cell. For a representation of this matrix, see Table 2.

This framework served two important functions at the meeting. First, it helped organize the generation of critical questions that needed to be answered in order for each tool to serve its intended purpose. Driving our conversations throughout the convening was: What would a system designer or implementer need to know and be able to do within this framework? Specifically, what questions would they need answered in order to design an effective teacher assessment system?

In addition to identifying knowledge needs that frame the research agenda, this framework highlighted the often overlooked role of measurement for improvement. In the current climate of teacher accountability, teacher assessment for evaluation has become the focus of public attention. Though practitioners and scholars alike frequently acknowledge the need for using measures of teaching to improve teaching, this potential is rarely explored in depth—often it is dwarfed by the urgent press to develop tools and protocols for accountability. We saw this convening of leading experts in teacher assessment as an opportunity to highlight this potential for applying teacher measurement to the improvement of teaching. This framework served to give improvement a place in the conversation.

Table 2: Framework for Organizing Teacher Assessment

		Purposes		
		Evaluation	Improvement	Research
Measures	Value-added			
	Observations			
	Overall System			

The framework seemed to resonate with the participants. Many stated that they found it very useful, and we received no challenge or push-back on it. We plan to continue to use this framework as a foundation for guiding our efforts moving forward.

A Preliminary Research Agenda (version 1.0)

Evaluation systems are currently being built and put into practice at a rapid pace in diverse contexts of use. Not surprisingly, this activity is generating, and will continue to generate, technical questions that will need to be answered if these efforts are to contribute to significant improvements in student learning. A primary focus of this convening was to begin to shape and synthesize these technical conversations. This convening resulted in surfacing an initial research agenda for building measurement systems for assessing teachers. The research agenda is made up of questions generated and prioritized by implementers of current evaluation systems and national experts engaged in developing and conducting research on such systems.

Prior to the convening, we participated in and observed a number of meetings with state and district officials currently engaged in implementing teacher evaluation systems³. We collected a broad array of technical questions that they consider critical to answer in order to implement feasible, valid, and reliable teacher assessment systems. We then synthesized these questions into an initial set of critical knowledge needs. Two weeks before the convening, we sent a brief survey to our participants asking participants to (a) select the questions they considered most important to address, and (b) suggest additional questions they considered to be critical to address for the design and implementation of reliable and valid teacher assessment systems. We sorted the resultant list into cells within the framework and populated the framework with these initial questions.

At the convening, we presented the framework described above, this time populated with the initial set of critical questions in the opening address to the participants. Participants were given an opportunity to generate additional critical questions that they thought the field needed to have answered and then sorted their questions into cells in the framework. Participants then voted for (a) which of the questions, if answered, would lead to the greatest improvement of the systems in use, and (b) which of the questions would most benefit from a conversation of the diverse expertise that was in attendance at the convening. This resulting list of questions sorted into cells in the framework and ranked by the number of votes received is presented in Table 3. We used the votes to inform the formation of small group discussion topics in order to delve more deeply into the questions deemed most critical to resolve.

Following up on this R&D agenda development, we plan to engage relevant stakeholder groups, from practitioners to policymakers, in vetting, validating and prioritizing the knowledge needs in this critical area of policy action.

³ Our participation in these meetings was funded outside of our cooperative agreement with IES.

	Evaluation	Improvement	Policy Research
Value-Added	<ul style="list-style-type: none"> • How do differences in test content and difficulty affect value-added? • How consistent are teacher performance results across alternative value-added models? • To what extent are an individual teacher's effects consistent across students or subjects? • Can we create value-added estimates when there's no pre-test or when not all students take a test? • Are value-added systems capable of measuring <i>improvement (growth)</i> in teacher performance to inform decisions predicated on such data? • How many performance levels can we statistically and meaningfully distinguish in value-added data? • What processes are necessary to roster students reliably? • How can we distinguish teacher effects from other factors contributing to differences in student learning gains (home conditions, school effects, classroom composition, peers, etc)? • What range of student outcome measures balance feasibility with sufficient coverage of the outcomes of interest? • Can we establish a statistically and conceptually defensible "minimum level of teacher performance?" • How should value-added systems handle students who do not spend a full year with one teacher? Have multiple teachers of record? Receive supplemental interventions? • What about possible confounding effects during the "summer periods?" • What is the optimal frequency of assessment for evaluation? • How much might individual student outliers influence value-added estimates? Are more robust methods needed? 	<ul style="list-style-type: none"> • Are value-added estimates useful for informing the improvement of teaching? How? • What is required of value-added systems in order to know if individual teachers are actually <i>improving</i> over time? Minimum data requirements? • Should we do value-added at individual or the school level? • What are mechanisms by which value-added could inform improvement? 	<ul style="list-style-type: none"> • How can we causally link district policy/practice to changes in teaching practice? • How would a district know if it's, on balance, actually improving practice through the introduction and implementation of a teacher assessment system?

	Evaluation	Improvement	Policy Research
Observations	<ul style="list-style-type: none"> • What are the capacity (training, ongoing retraining, audits, dynamic analysis, etc) requirements for inter-rater reliability? What does it take to maintain reliability of observations in practice? • How do we manage the cost and capacity issues involved with observational measures? • How many observations need to be conducted to derive a reliable and valid measure of instructional behavior? Does this vary as a function of grain-size and behavior prevalence? • What other kinds of data can be used to supplement observations? For example, logs, surveys, video, classroom artifacts. • Validity. What is effective teaching? Can it be measured by observation? • How do we avoid the dangers of local norming of observational measures? How do we train observers across different sites who are the best observers? • How do we handle the contextual nature of teaching: same practice may have different affects in different context or within different instructional model? • How can observational data or protocols be systematized to track teacher growth? • Should we broaden observational measures to include planning, parent conferences and other events outside classroom practices? • What should be done when different observers yield different assessments with observational systems? Should we assume that independent external evaluators are the “gold standard”? 	<ul style="list-style-type: none"> • How do we develop a culture around teacher evaluation that leads to a stance of improvement? • What are the mechanisms by which classroom observations can improve instructional quality (for example, learning from a measurement versus learning from a process of analysis)? • Can current observation systems reliably identify individual teacher’s improvement needs? • Can observational rubrics be generic or must they be subject matter or curriculum specific? • How can technology be used to reduce the cost and improve the effectiveness of observational teaching? • How specific (i.e., grade level, subject, instructional system) do assessments need be in order to inform teacher learning? • How do we best balance the length and complexity of observations against fairness concerns (which often urge that systems be made increasingly complex)? • What is the prescriptive detail needed (aka signaling capacity of rubrics) to inform improvement versus more reduced information that may be sufficient to predict future performance? • How often do teachers need feedback in order for it to be useful for improvement? • What kinds of feedback are most useful for improving teaching? For example, is measurement required or not? • What is required statistically of these systems in order to know if teachers are actually <i>improving</i> over time? • Do we have evidence of good use of these data for individual teacher improvement? For improvements by a school-based professional community? Possible misuses? • Does possible focus on improving observational indicators lead to improvement of teaching or could it undermine such improvement? 	

Overall Systems

- How much do our measures detect properties that are not attributed to individual teachers?
- How best to combine multiple measures of teacher effectiveness?
- How does alignment between measures shape the utility of whole evaluation systems?
- How correlated do we want value added and observational measures to be and why?
- How can we design a targeted, cost-effective system that recognizes diversity of teaching quality in the U.S.?
- How can assessment be implemented in a way that teachers see as helpful and part of their work?
- What are the actual costs to implement and sustain reliable teacher assessment systems? What are the components of these costs (monetary, in-kind, opportunity)?
- How best to communicate value-added results and their limitations to teachers, parents, public, etc? What are costs-benefits?
- How robust are these systems against cheating?
- What would building a system for practice improvement look like?
- How can we integrate into one coherent system for users?
- How can we directly link measures to professional development?
- How can other types of evidence be integrated into value-added and observations?
- Given that VA and observational measures are imperfect, how are they best used in decision making around high-stakes decisions such as retention, promotion, compensation.
- How much do our measures detect properties that are not attributed to individual teachers?
- What is the arc of development for teachers (in terms of career stage, subject, grade level)?
- To what extent do the major players understand how this works?
- How can we minimize the degree to which the measures of performance get distributed when the stakes are attached (not just about cheating)?
- How can we isolate the effects of policies to change instruction when the theory is that the policies will change school “culture”?
- What should happen when observation data and value-added measures do not agree?
- What happens when a good teacher has a bad year?
- What do we need to know about the relationship among knowledge, beliefs, dispositions, and perspectives in order to better promote change?
- What are the stakeholder groups that need to be involved in the improvement process?
- What are the pros and cons of tightly coupled vs. loose systems in terms of links between measures and actions.
- What are capacity needs of districts/states for decision making and data use?
- Might such systems undermine cooperative, collective action by school-based professional communities?
- Is “reliable and defensible” the right question when we are talking about human behaviors? (What is the role of transparency?)
- How can we best communicate results of VA or observations to teachers/parents?
- How can we assess knowledge (especially declarative knowledge)?
- How can greater coupling be achieved across leverage points?
- Can one system serve the needs of different uses, users, and purposes?
- Are we indifferent to “good teaching” or are there certain core values we want reflected in the system?
- Are there ways to weight factors associated with teacher effectiveness?



Assessing Teaching Collaboratory Convening
May 24-25, 2011

PARTICIPANT BIOGRAPHIES



Joseph A. Aguerrebere is President and CEO of the National Board for Professional Teaching Standards® (NBPTS) in Arlington, Virginia. NBPTS is an independent, nonprofit organization that advances the quality of teaching and learning by establishing rigorous standards for what accomplished teachers should know and be able to do. It provides a national voluntary system certifying teachers who meet these standards. Dr. Aguerrebere was born and raised in East Los Angeles. After graduating from Garfield High School (subject of the movie, *Stand and Deliver*), he attended the University of Southern California where he earned a BA in political science, and masters and doctorate degrees in educational administration. His education career includes serving as a teacher and administrator in elementary, middle, and high school settings in California. He later served as a professor of educational administration at California State University, Dominguez Hills in Los Angeles where he prepared educators to work successfully in urban settings. Prior to joining NBPTS, Dr. Aguerrebere was Deputy Director at the Ford Foundation in New York. He is recognized as a national expert on school reform, serves on numerous boards, and is an advisor to education organizations, journalists, and government.



Since founding Teachscape in 1999, **Mark Atkinson** has worked tirelessly to create an education company that is committed to improving the quality of instruction that U.S. children receive by improving the caliber of teaching they receive every day. Prior to Teachscape, Mark was a network news producer with extensive experience in reporting, producing and directing network news documentaries. He served as senior producer and manager of New Markets for CBS News Productions, where he developed new business opportunities focusing on multimedia production and the education market. He has directed and produced works with broadcast luminaries such as Mike Wallace and Peter Jennings, and has received numerous awards, including the Gold Baton, the highest honor of the Alfred I. duPont-Columbia University Awards for work associated with reporting the war in Bosnia; the Overseas Press Club Award; and an Emmy Award for reporting on the United Nations peacekeeping mission in Bosnia. Mark is a graduate of Yale University.



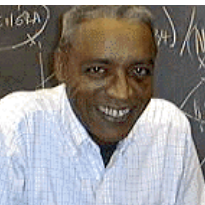
Courtney Bell is a Research Scientist in ETS's Understanding Teaching Quality Center. She completed her doctorate at Michigan State University in Curriculum, Teaching, and Educational Policy after earning her B.A. in Chemistry at Dartmouth College. Courtney's dissertation study, funded through a Spencer Foundation Dissertation Fellowship, won the AERA Division L Dissertation of the Year award. That longitudinal interview study investigated urban parents' selection of schools for their middle and high school children. A former high school science teacher, Courtney's work looks across actors in the educational system to better understand the intersections of policy and practice. Her current studies use mixed-methods to analyze teacher learning, the measurement of teaching, and the effects of racially desegregated schools on student learning. She is currently PI on two large validity studies of instruments designed to measure

Appendix A: Assessing Teaching Collaboratory Participant Bios

teaching. These studies are funded by the Gates, W.T. Grant and Spencer Foundations. As a part of her work on another study, Measures of Effective Teaching, she is working with colleagues at ETS to develop measures of teachers' content knowledge for teaching in ELA and mathematics. Courtney has published in scholarly journals including Educational Evaluation and Policy Analysis, Journal for Research in Mathematics Education, American Journal of Education, Journal of Education Policy, and Teachers College Record.



Damian Betebenner is a senior associate with the National Center for the Improvement of Educational Assessment (NCIEA). His work currently centers exclusively on the development, implementation, integration and reporting/communication of state level growth analyses. He is the architect of the Colorado Growth Model which, in addition to Colorado, has been adopted by or is in various stages of implementation in more than 20 other states. In 2010 the model received the National Council on Measurement in Education's annual award for Outstanding Dissemination of Educational Measurement Concepts to the Public. In addition, Dr. Betebenner is the project lead for interactive Colorado Growth Model data visualization software that was recognized by Adobe Software as a Max Award Finalist at its 2009 Adobe Max convention for innovative uses of Adobe technology. Dr. Betebenner holds a Ph.D. in Mathematics from the University of Wyoming and a Ph.D. in Educational Measurement from the University of Colorado, Boulder.



Lloyd Bond (Ph. D., 1976, Psychology/Psychometrics, The Johns Hopkins University) is currently a retired (2008) Senior Scholar with the Carnegie Foundation for the Advancement of Teaching in Stanford California and Emeritus Professor of Education at the University of North Carolina, Greensboro. From 2002 to 2008 he was a Senior Scholar at Carnegie working in the area of assessment across several Carnegie Foundation programs. Dr. Bond taught test theory and psychometrics at the University of Pittsburgh from 1976 to 1988, and at the University of North Carolina (Greensboro) from 1988 to 2002. Professor Bond has published widely in the area of assessment, measurement theory, and testing policy and has made fundamental contributions to the measurement of complex performance and to the literature on measures of quantitative reasoning. He has held editorial positions on the leading journals in educational and psychological measurement and serves on numerous commissions and panels devoted to testing and testing policy. He is currently a member of the Design and Analysis Committee of the National Assessment of Educational Progress (NAEP) and the Psychometric Panel of the College Board. A fellow of both the American Psychological Association and the American Educational Research Association (AERA), Professor Bond is the recipient of numerous honors and awards, including the Presidential Citation from AERA for Contributions to Educational Measurement. Dr. Bond was recently honored by having the new K-8 charter school in the Chicago neighborhood where he grew up named in his honor.



Cynthia Brunswick, Chicago New Teacher Center Director, leads the pioneering Chicago office of the New Teacher Center which was established in 2006 to bring the NTC's proven, comprehensive, mentor-based induction program to Chicago Public Schools. Under Dr. Brunswick's leadership, CNTC worked closely with CPS in 2009 to successfully scale up its induction program from 342 teachers in 92 schools to 1079 teachers in 380 schools. Additionally, she is spearheading the creation of a new induction program which support up to 60 new Chicago principals. Prior to joining CNTC, Dr. Brunswick was the Director of Literacy Professional Development for the Center for Urban School Improvement at the University of Chicago. There she led a program which designed and delivered professional development for principals and teacher leaders in a number of Chicago Public and Charter Schools. She began her career as a middle school and high school teacher on Chicago's Southside, and dedicated 10 years to teaching in one of Chicago's hardest to staff schools.

Appendix A: Assessing Teaching Collaboratory Participant Bios



Anthony S. Bryk is the ninth president of The Carnegie Foundation for the Advancement of Teaching. He held the Spencer Chair in Organizational Studies in the School of Education and the Graduate School of Business at Stanford University from 2004 until assuming Carnegie's presidency in September 2008. He came to Stanford from the University of Chicago where he was the Marshall Field IV Professor of Urban Education in the sociology department, and where he helped found the Center for Urban School Improvement, which supports reform efforts in the Chicago Public Schools. He also created the Consortium on Chicago School Research, a federation of research groups that have produced a range of studies to advance and assess urban school reform. His current research and practice interests focus on the organizational redesign of schools and school systems and the integration of technology into schooling to enhance teaching and learning.



Joanna Cannon is the Executive Director of Teacher and Principal Evaluation at the NYC Department of Education. In this role, Joanna is responsible for designing and implementing a new teacher and principal evaluation system for NYC's 80,000 teachers and 1,700 principals. Joanna also serves on the NY State Task Force for teacher and principal evaluation, which is responsible for creating recommendations and guidance for the recently adopted legislation governing educator evaluation in NY State. Joanna joined the NYCDOE in 2007 and previously served as the Deputy Executive Director of the NYCDOE's Research Office. Joanna holds a Ph.D. from Columbia University and completed a postdoctoral fellowship at the University of Chicago, where her research focused on lesson study, teacher cognition, and the design of mathematics curriculum and assessment.



Tom Corcoran directs the Consortium for Policy Research in Education (CPRE) at Teachers College, Columbia University, and is principal investigator of the Center on Continuous Instructional Improvement (CCII). Previously, Corcoran served as Policy Advisor for Education for New Jersey Governor Jim Florio, Director of School Improvement for Research for Better Schools, and Director of Evaluation and Chief of Staff of the New Jersey Department of Education. He has served as a consultant to urban school districts and national foundations on improving school effectiveness and equity. He served as a member of the National Research Council's K-8 Science Learning Study and its Steering Committee on Evaluating Options for Common Standards, and has been a visiting professor of education policy at the Woodrow Wilson School of International and Public Affairs at Princeton University since 1999.



Charlotte Danielson is an internationally-recognized expert in the area of teacher effectiveness, specializing in the design of teacher evaluation systems that, while ensuring teacher quality, also promote professional learning. She advises State Education Departments and National Ministries and Departments of Education, both in the United States and overseas. She is in demand as a keynote speaker at national and international conferences, and as a policy consultant to legislatures and administrative bodies. Ms. Danielson's many publications range from defining good teaching ("Enhancing Professional Practice: a framework for teaching," 2007), to organizing schools for student success ("Enhancing Student Achievement: a framework for school improvement," 2002), to teacher leadership ("Teacher Leadership that Strengthens the Profession," 2006), to professional conversations ("Talk about Teaching! Conducting Professional Conversations," 2009), to numerous practical instruments and training programs (both onsite and online) to assist practitioners in implementing her ideas.

Appendix A: Assessing Teaching Collaboratory Participant Bios



Dan Goldhaber is the Director of the Center for Education Data & Research (CEDR) and a Professor in Interdisciplinary Arts and Sciences at the University of Washington-Bothell. He is also an Affiliated Scholar at the Urban Institute, the co-editor of *Education Finance and Policy*, and a member of the Washington State Advisory Committee to the U.S. Commission on Civil Rights. Goldhaber previously served as an elected member of the Alexandria City School Board from 1997–2002, and as an Associate Editor of *Economics of Education Review*. Goldhaber's work focuses on issues of educational productivity and reform at the K–12 level, with a current focus on the broad array of human capital policies that influence the composition, distribution, and quality of teachers in the workforce. Topics of published work in this area include studies of the stability of value-added measures of teachers, the effects of teacher qualifications and quality on student achievement, and the impact of teacher pay structure and licensure on the teacher labor market. Previous work has covered topics such as the relative efficiency of public and private schools, and the effects of accountability systems and market competition on K–12 schooling. Goldhaber's research has been regularly published in leading peer-reviewed economic and education journals such as: *American Economic Review*, *Review of Economics and Statistics*, *Journal of Human Resources*, *Journal of Policy and Management*, *Journal of Urban Economics*, *Economics of Education Review*, *Education Finance and Policy*, *Industrial and Labor Relations Review*, and *Educational Evaluation and Policy Analysis*. The findings from these articles have been covered in more widely accessible media outlets such as National Public Radio, the *New York Times*, the *Washington Post*, *USA Today*, the *Wall Street Journal*, and *Education Week*. Goldhaber's research has been funded by the U.S. Department of Education, the Carnegie Corporation of New York, the Bill and Melinda Gates Foundation, the Ewing Marion Kauffman Foundation, the Joyce Foundation, the Smith Richardson Foundation, and numerous other private foundations. Goldhaber holds degrees from the University of Vermont (BA, Economics) and Cornell University (MS and PhD, Labor Economics).



Louis Gomez is the Helen Faison Professor of Urban Education and Sr. Scientist at the Learning Research and Development Center (LRDC) at The University of Pittsburgh. Professor Gomez is also currently serving as a Senior Partner at the Carnegie Foundation for the Advancement of Teaching in Palo Alto, CA. His scholarship focuses on understanding how to support organizational change in schools and other institutions. Along with his colleagues, Professor Gomez has been dedicated to collaborative research and development with urban communities to bring the current state-of-the-art in instruction and support for community formation to traditionally underserved schools. Most recently, Professor Gomez has turned his attention to problem solving research and development. This is *R&D organized around high-leverage problems* embedded in the day-to-day work of teaching and learning and the institutions in which these activities occur.

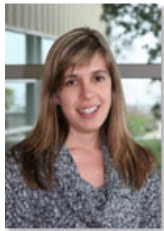


Sharon Greenberg is an Education Consultant. She is a co-founder of the Center for School Improvement at the University of Chicago (now the Urban Education Institute) where she served as director of research and helped to develop core programmatic initiatives in literacy, social services, and leadership. In Chicago she also served as a consultant to the Chicago Public Schools Chief Education Officer, and was responsible for the conceptualization and delivery of one strand of professional development for the district's Area Instructional Officers and Area Reading Coaches. More recent work in California includes charter school design. Recent consulting activities include that of senior researcher on the Performance-based Assessment of Literacy Coaching (PALC) conducted by Professors Pinnell, Fountas and Bryk, as well as senior researcher on an investigation of early career English Language Arts teachers in New York City middle schools conducted by Professors Grossman and Loeb. Presently Greenberg is a consultant to the Carnegie Foundation where she is involved in strategic planning and proposal development. Her substantive focus at the Foundation is building capacity in the areas of literacy and second language learning. Greenberg also serves as a consultant to the lead team of the BTEN Partnership. She has a Ph.D. from the University of Chicago and a B.A. and M.A.T. degree from Stanford University.

Appendix A: Assessing Teaching Collaboratory Participant Bios



Pam Grossman is the Nomellini-Olivier Professor of Education at the Stanford University School of Education. She completed her undergraduate degree in English at Yale University and her PhD from Stanford University. Her research interests include teacher education and professional education more broadly, teacher knowledge, and the teaching of English in secondary schools. She has been engaged with a five year study of pathways into teaching in New York City schools, focusing on the features of preparation that affect student achievement. Building on this work, she has investigated the classroom practices of middle-school English teachers that are associated with student achievement. She is a member of the National Academy of Education and currently serves as the Faculty Director of the new Center to Support Excellence in Teaching (CSET). A former high school English teacher, Grossman also teaches the prospective English teachers in Stanford's teacher education program.



Alicia Grunow is the Senior Associate for Learning Teaching at the Carnegie Foundation for the Advancement of Teaching. She leads the Foundation's program of work focused on the development of assessments and practices for the purposes of improving teaching. She also leads efforts to adapt tools from improvement research to support change efforts in education. Her background has afforded her with an unusual combination of practical experience in the development of teachers and technical skills in statistical analysis. For the past four years she has worked as an instructor in Stanford's Teacher Education Program (STEP), teaching classes on practices to support the academic achievement of English Language Learners. During that time she also worked as a research assistant on a variety of projects, conducting large-scale quantitative analyses. She earned her masters in economics and doctorate in educational administration and policy analysis at Stanford University in 2011. Grunow received her B.A. in Psychology from Reed College in 1999, completed the Bilingual and ESL Teachers Leadership Academy at Bank Street College in 2005. Before coming to Stanford, she taught for seven years in elementary school programs designed for English Language Learners in both Denver and New York City.



Douglas Harris is an economist and Associate Professor of Educational Policy and Public Affairs at the University of Wisconsin at Madison. His research explores the efficiency and equity of K-12 and higher education programs, especially teacher evaluation and accountability. He is the author of *Value-Added Measures in Education* (Harvard Education Press, 2011). In 2008, he chaired the 2008 National Conferences on Value-Added in Madison and Washington, DC. His research on value-added has been published in books journals (*Education Finance and Policy*, *Journal of Policy Analysis and Management*, and *Journal of Public Economics*). He is extending this work to a new project on measuring the performance of colleges and universities, and in ways that integrate value-added within a cost-effectiveness and productivity framework. He is also co-director of the Wisconsin Scholars Longitudinal Study (WSLS), analyzing a program that provides financial aid to randomly selected low-income college students. He is an affiliate of the Center for Analysis of Longitudinal Data in Education Research (CALDER) and his research has been supported with funding from the U.S. Department of Education and a variety of foundations: Carnegie Corporation, Gates, WT Grant, Joyce, Lumina, Smith Richardson, and Spencer.



Heather C. Hill is an associate professor at the Harvard Graduate School of Education. Her primary work focuses on developing new measures of mathematics teacher and teaching quality, and using these measures to inform current policies and instructional improvement efforts. Over the period 2000-2010, she and colleagues developed an assessment of teachers' mathematical knowledge for teaching (MKT) as well as an observational instrument to evaluate the mathematical quality of instruction (MQI) within classrooms. She is codirector of the National Center for Teacher Effectiveness and also principal investigator of a five-year study examining the effects of Marilyn Burns Math Solutions professional development on teaching and learning. Her other interests include

Appendix A: Assessing Teaching Collaboratory Participant Bios

instructional improvement efforts in mathematics and the role that language plays in the implementation of public policy. She has served as section chairs for the American Educational Research Association and Society for Research on Educational Effectiveness conferences, and on the editorial boards of *Journal of Research in Mathematics Education* and the *American Educational Research Journal*. She is the coauthor, with David K. Cohen, of *Learning policy: When state education reform works* (Yale Press, 2001).



Andrew Ho is a psychometrician working at the intersection of educational statistics and educational policies. His research informs and improves the development, use, and interpretation of large-scale educational accountability metrics. He has studied the consequences of "proficiency"-based accountability metrics, the validation of high stakes test score trends with low stakes comparisons, and the potential for alternative accountability structures—like "growth models" and "index systems"—to improve school- and classroom-level incentives. His current projects include articulating meaningful contrasts between accountability models for student growth and developing new achievement gap and growth metrics for cross-test comparison and validation. He has his Ph.D. in Educational Psychology and his M.S. in Statistics from Stanford University. Dr. Ho has been a postdoctoral fellow at the National Academy of Education and Spencer Foundation and a recipient of the Jason Millman Promising Measurement Scholar Award from the National Council on Measurement in Education.



Brian Jacob is the Walter H. Annenberg Professor of Education Policy, Professor of Economics, and Director of the Center on Local, State and Urban Policy (CLOSUP) at the Gerald R. Ford School of Public Policy. He is also a Faculty Research Fellow at the National Bureau of Economic Research and an Executive Committee Member of the National Poverty Center. He has previously served as a policy analyst in the NYC Mayor's Office and taught middle school in East Harlem. His primary fields of interest are labor economics, program evaluation, and the economics of education. His current research focuses on urban school reform and teacher labor markets. In recent work, he has examined school choice, education accountability programs, housing vouchers, and teacher labor markets.



Thomas Kane is Deputy Director for Research and Data within the College Ready team at the Bill and Melinda Gates Foundation as well as Professor of Education and Economics at the Harvard Graduate School of Education. His research has had an impact on a range of education policies affecting both the K-12 and postsecondary sectors, including the design of school accountability systems, charter school laws, teacher recruitment and retention, financial aid for college, college admissions and community colleges. From 1991 through 2000, he was an assistant and associate professor of public policy at Harvard's Kennedy School of Government. From 1995 to 1996, Kane was on leave from Harvard, serving as a senior economist within President Clinton's Council of Economic Advisers. Kane has also been a professor of public policy at UCLA and has held visiting fellowships at the Brookings Institution in Washington and the Hoover Institution at Stanford University.



Nicole B. Kersting is an Assistant Professor of Education (Teaching, Learning, and Socio-Cultural Studies) and a faculty member of the Interdisciplinary Graduate program in Statistics at the University of Arizona. Her research is focused on the measurement of different aspects of teacher quality: teacher knowledge, instructional quality and student learning. She developed and validated a novel approach to measure teacher knowledge in mathematics that is based on teachers' analyses of classroom video clips. She received a Masters in Linguistics from the Friedrich-Wilhelm University in Bonn, Germany, a Masters and a Ph.D. in quantitative research methodology from UCLA.

Appendix A: Assessing Teaching Collaboratory Participant Bios



Paul G. LeMahieu is the Senior Managing Partner for Design, Development, and Research at the Carnegie Foundation for the Advancement of Teaching and is graduate faculty in the College of Education at the University of Hawai'i – Mānoa. With degrees from Yale College (AB), Harvard University (EdM), and University of Pittsburgh (PhD), his scholarly interests focus on educational assessment and accountability as well as classroom learning and the professional development and policy environments that support it. From 2002 to 2010, LeMahieu was Director of Research and Evaluation for the National Writing Project at the University of California, Berkeley. Prior to that, LeMahieu served as Superintendent of Education for the State of Hawai'i, the chief educational and executive officer of the only state system in the United States that is a unitary school district, serving over 190,000 students with annual budgets totaling over \$1,800,000,000. LeMahieu has published extensively on issues as diverse as testing policy and practice; educational accountability; staff development; school effectiveness; nontraditional work roles for women; minority achievement issues; science education; and vocational education. He has received a number of major awards for his contributions to educational theory and practice from the American Educational Research Association, the Evaluation Research Society, the Buros Institute of Measurement, the National Association of Test Directors, and the Association for Supervision and Curriculum Development. He has been President of the National Association of Test Directors and Vice President of the American Educational Research Association. He served on the National Academy of Sciences' Board on International Comparative Studies in Education, and Mathematical Sciences Education Board. He is a Founding Director of the Center for the Study of Research on Expertise in Teaching and Learning, served on the National Board on Testing Policy, and the National Board on Professional Teaching Standards.



Susanna Loeb is a professor of education at Stanford University, faculty director of the Center for Education Policy Analysis, and a co-director of Policy Analysis for California Education (PACE). She specializes in the economics of education and the relationship between schools and federal, state and local policies. Her research addresses teacher policy, looking specifically at how teachers' preferences affect the distribution of teaching quality across schools, how pre-service coursework requirements affect the quality of teacher candidates, and how reforms affect teachers' career decisions. She also studies school leadership and school finance, for example looking at how the structure of state finance systems affects the level and distribution of resources across schools. Susanna is a senior fellow at the Stanford Institute for Economic Policy Research, a faculty research fellow at the National Bureau of Economic Research, a member of the Policy Council of the Association for Policy Analysis and Management, and Co-Editor of Educational Evaluation and Policy Analysis.



Daniel F. McCaffrey is a senior statistician at the RAND Corporation, where he holds the PNC Chair in Policy Analysis. He is a fellow of the American Statistical Association and is nationally recognized for his work on value-added modeling for estimating teacher performance. McCaffrey oversees RAND's efforts as part of the Gates Foundation's Measures of Effective Teaching study to develop and validate sophisticated metrics to assess and improve teacher performance. He is currently leading RAND's efforts on two additional studies comparing value-added measures to other measures of teaching, including classroom observations, and is a major partner in the National Center on Performance Incentives, which is conducting random control experiments to test the effects of using value-added to reward teachers with bonuses. He is co-principal investigator of a project funded by the Institute of Education Sciences (IES) that is developing alternative value-added models of teachers' effectiveness. McCaffrey is also the principal investigator of a National Institute on Drug Abuse-funded study, and recently worked on the design of an IES-funded random trial of the Cognitive Tutor Geometry curriculum. He led an evaluation of the Pennsylvania Value-Added Assessment Pilot Program (PVAAS) and was the lead statistician on two randomized field trials of school-based interventions: evaluations of the Project ALERT Plus middle and high school drug prevention program and the teen dating violence prevention curriculum, Break the Cycle.

Appendix A: Assessing Teaching Collaboratory Participant Bios

McCaffrey received his Ph.D. in statistics from North Carolina State University.



Robert Meyer, Research Professor and Senior Scientist at the Wisconsin Center for Education Research (WCER), is the Director of the Value-Added Research Center (VARC). Meyer is known for his research on value-added modeling and evaluation methods and is currently working on projects funded by the Institute of Educational Sciences (U.S. Department of Education), the Joyce Foundation, the Milwaukee Public Schools, the National Science Foundation, and the Wisconsin Department of Public Instruction. Over the last decade and a half, Meyer has worked closely with districts and states to develop and apply innovative statistical methods. He has conducted major statistical evaluations of programs and policies

such as SAGE (the Wisconsin class-size initiative), systemic reform in Texas, integrated versus traditional mathematics, and professional development and other math and science reforms in Cleveland and Riverside, California. Meyer has also worked with numerous districts to develop and implement value-added indicator and accountability systems, including the school report card implemented in the Milwaukee Public Schools in 2002.



Jeannie Myung is a Research Associate for the Learning Teaching program at the Carnegie Foundation. She received her Ph.D. in Administration and Policy Analysis at the Stanford University School of Education. Her dissertation focused on school district personnel practices around the selection and development of teacher leaders. Prior to graduate school, she was an elementary public school teacher in San Jose, California through Teach for America. Jeannie holds a B.A. in political science

from Yale College.



David Pearson is a faculty member in the programs in Language and Literacy and Cognition and Development at the Graduate School of Education at the University of California, Berkeley, where he served as Dean from 2001-2010. Current research projects include *Seeds of Science/Roots of Reading*--a Research and Development effort with colleagues at Lawrence Hall of Science in which reading, writing, and language as are employed as tools to foster the development of knowledge and inquiry in science--and the Strategic Education Research Partnership--a

collaboration between UC Berkeley, Stanford, and the SFUSD designed to embed research within the portfolio of school-based issues and priorities. Prior to coming to Berkeley in 2001, he served on the faculties of education at Michigan State, Illinois, and Minnesota. Awards include the 1989 Oscar Causey Award (NRC) for contributions to reading research, the 1990 William S. Gray Citation of Merit (IRA) for contributions to reading research and practice, the 2005 Albert J. Harris Award (IRA) for the year's best reading disability publication, and the 2003 Alan Purves Award (NCTE) for a publication impacting practice. In 2006 the University of Minnesota honored him with the Alumni Outstanding Achievement Award, and in 2010 AERA gave him Distinguished Contributions to Research in Education Award. He is the founding editor of the *Handbook of Reading Research* now in its fourth volume, he edited *Reading Research Quarterly* and the *Review of Research in Education*, and he has served on the Editorial Review Board for some 20 educational journals. Professor Pearson received his B.A. in History from the University of California at Berkeley, taught elementary school in California for several years, and went on to complete his Ph.D. in Reading Education at the University of Minnesota. He completed post-doctoral study at the University of Texas, Austin and Stanford University

Appendix A: Assessing Teaching Collaboratory Participant Bios



Raymond Pecheone is the founder and Executive Director of the Stanford Center for Assessment Learning and Equity (SCALE), a new center launched in 2009 that focuses on (a) the development of performance assessments for teachers and administrators at the school, district and state levels and (b) the development of a performance-based system for student assessment to support the development of the next generation of formative and summative assessments at the district, state and federal levels. SCALE provides comprehensive supports for standards based teaching and learning and is built around the development of interactive assessment and multimedia instructional tools to support college and career readiness. Prior to launching the SCALE center, Dr. Pecheone has held a variety of leadership roles in the Connecticut State Department of Education as the Bureau Chief for Curriculum, Research, Testing and Assessment, the Co-director of the first Assessment Development Lab for the National Board for Professional Teaching Standards (NBTS) and as a lead consultant to the Chief State School Officers (CCSSO) in the design and development of innovative assessments for teachers and administrators. Additionally, Dr. Pecheone co-founded the Interstate Teacher Assessment and Support Consortium (INTASC), which develops national standards and assessments for teachers and administrators and is housed at the Council of Chief State School Officers. During a sabbatical at Teachers College Columbia, he directed the redesign of the New York student assessment system including the NY Regents examination. Most recently, Dr. Pecheone is leading a national assessment for pre-service teaching which includes 22 states and 85 universities. Lastly he is directing a large scale student assessment research project in NYC to develop performance assessment systems for students that support deeper learning and are designed to prepare all students for college and career success.



Robert Pianta is the Dean of the Curry School of Education at the University of Virginia, the Novartis Professor of Education, Professor of Psychology, and the director of UVa's Center for Advanced Study of Teaching and Learning. Dean Pianta and his education research team have proven what it takes to build better teachers. With more than \$50 million in grant funding, his team has developed a system to both assess and improve a teacher's effectiveness in the classroom. The Classroom Assessment Scoring System or CLASS is an observational measure that has been tested and proven effective in several large national studies and is being utilized by every Head Start program in the country, touching 50,000 teachers and over a half million students. Having earned a B.S. and an M.A. in Special Education from the University of Connecticut and a Ph.D. in Psychology from the University of Minnesota, Bob began his career as a special education teacher. He joined the Curry faculty in 1986 and began his appointment as dean in 2007. Bob's work has been nationally recognized by Andrea Mitchell of NBC Nightly News, Jay Mathews of the Washington Post, and best-selling author Malcolm Gladwell, and he was asked to provide recommendations public education to the Obama Presidential Transition Team. In his recommendations he wrote, "Good teachers are key. If we want to improve our students' learning, we need to improve the quality of teachers and of teaching."



Gay Su Pinnell is a professor in the School of Teaching and Learning at The Ohio State University. She has extensive experience in classroom teaching and field-based research, and in developing comprehensive approaches to literacy education. She has received the International Reading Association's Albert J. Harris Award for research and the Charles A. Dana Foundation Award for her contributions to the field of education. She is also a member of the Reading Hall of Fame. Together with Irene Fountas she has authored numerous books, videos, and websites with Heinemann that are considered standards in the field of literacy instruction and staff development. Their latest innovations are The Fountas & Pinnell Leveled Literacy Intervention and The Fountas & Pinnell Benchmark Assessment System, a comprehensive assessment system for grades K-8. Fountas and Pinnell together present workshops nationwide on a variety of literacy-instruction topics through Heinemann Professional Development.

Appendix A: Assessing Teaching Collaboratory Participant Bios



Sean Reardon is associate professor of education and (by courtesy) sociology at Stanford University, specializing in research on the effects of educational policy on educational and social inequality, on the causes, patterns, trends, and consequences of social and educational inequality, and in applied statistical methods for educational research. His primary research examines the relative contribution of family, school, and neighborhood environments to racial/ethnic and socioeconomic achievement disparities. In addition, he develops methods of measuring social and educational inequality (including the measurement of segregation and achievement gaps) and methods of causal inference in educational and social science research. He teaches graduate courses in applied statistical methods, with a particular emphasis on the application of experimental and quasi-experimental methods to the investigation of issues of educational policy and practice. Sean received his doctorate in education in 1997 from Harvard University. He has been a recipient of a William T. Grant Foundation Scholar Award, a Carnegie Scholar Award, and a National Academy of Education Postdoctoral Fellowship.



William L. Sanders was formerly a senior research fellow with the University of North Carolina system and is senior manager of value-added assessment and research for SAS Institute Inc. in Cary, N.C. He assumed the SAS position in June of 2000, upon retiring after more than 34 years as professor and director of the University of Tennessee's Value-Added Research and Assessment Center. Sanders has served as an advisor to policy makers at the federal level; he has worked with many states and districts interested in developing a value-added component to leverage their testing data into more precise and reliable information for better decision making. In addition to his assignment as director of the Value-Added Research and Assessment Center at the University of Tennessee, Sanders had leadership responsibilities for the Statistical and Computing Services Unit and served as an adjunct professor in the department of statistics within the College of Business Administration. Outside the area of education, he has been a statistical consultant to the agricultural, manufacturing, engineering and development industries. He has served as a statistical consultant to numerous regional research projects involving researchers from many universities and disciplines. Sanders was the Jason Millman Memorial Lecturer at the National Evaluation Institute in San Jose, Calif., (July 2000). He received his bachelor's degree and doctorate from the University of Tennessee, Knoxville



James W. Stigler is Professor of Psychology at UCLA and a Senior Partner at the Carnegie Foundation for the Advancement of Teaching. He was Director of the TIMSS video studies, and founder and CEO of LessonLab. He has authored numerous articles and books, including *The Teaching Gap* (with James Hiebert, Free Press, 1999/2009) and *The Learning Gap* (with Harold Stevenson, Simon & Schuster, 1992). He received his A.B. from Brown University, a Masters in Education from the University of Pennsylvania, and a Ph.D. in Developmental Psychology from the University of Michigan. He has received numerous awards for his research, including a Guggenheim Fellowship and the QuEST award from the American Federation of Teachers. Stigler is best known for his observational studies of mathematics and science teaching, and has pioneered the use of multimedia technology for the study of classroom instruction.

Appendix B: Assessing Teaching Collaboratory Agenda



Assessing Teaching Collaboratory Convening May 24-25, 2011

AGENDA

Tuesday 5/24

8:00-8:30am Breakfast

8:30-9am Welcome, Background and Setting the Stage

Objectives

- Identify critical technical issues (knowledge gaps and implementation gaps) for inquiry/development in building measurement systems for assessing teachers (v 1.0 of research agenda)
- Prioritize knowledge and implementation gaps with respect to their potential for improving practice and appropriateness for solution through design-engineering processes.
- Build relationships for and interest in forming a learning and improvement community around a subset of these issues

9-9:45am Generating Additional Questions

9:45-10am Individual Presentations

10-10:20am Break

10:20-11am Prioritizing a Research Agenda

11-11:15am Individual Presentations

11:15-12:15pm Taking Stock: Whole Group Conversation

12:15-1pm Lunch

1-1:15pm Individual Presentations

1:15-3pm Technical Conversations

3-3:15pm Break

3:15-3:30pm Individual Presentations

3:30-5:15pm Broad Systems Questions

Reception and Dinner to follow

Wednesday 5/25

8-8:30am Breakfast

8:30-9:45am Summary and Feedback

9:45-10:15am Interests and Potential for Working Together

10:15-10:30am Break

10:30-11:15am Advising the Carnegie Foundation on Actions

11:15-11:30am Closing

11:30am-12:30 Lunch