

# Evaluating Teaching

## 16:300:650:01

### 3 credits

Instructor: Drew Gitomer	drew.gitomer@gse.rutgers.edu
Phone Number: 732-932-7496, ext 8126	10 Seminary Pl Rm 9
Office Hours: by appointment	Prerequisites or other limitations: none
Mode of Instruction: <input type="checkbox"/> Lecture <input checked="" type="checkbox"/> Seminar <input type="checkbox"/> Hybrid <input type="checkbox"/> Online <input type="checkbox"/> Other	Permission required: <input checked="" type="checkbox"/> No <input type="checkbox"/> Yes  Directions about where to get permission numbers: from the instructor

#### **Learning goals**

1. To understand the policy issues related to the evaluation of teaching
2. To learn about and gain experience in the use of different methods for measuring teaching quality
3. To understand the different measurement approaches that are used to evaluate teaching quality, including the limitations of various methods
4. To understand and be able to analyze and critique teaching evaluation systems.

#### **Course catalogue description**

Evaluating teaching quality is a cornerstone of current educational reform efforts. There are multiple motivations for improved evaluation practices, ranging from professional development to making decisions about employment and compensation. Over the last two decades, new models of evaluation that include value-added modeling based on student test scores, classroom observations, student perception surveys, teacher portfolios, classroom assignments, and new measures of teacher knowledge have been introduced.

This course will explore the policy context, methods, and research around the evaluation of teaching quality. We will study the assumptions, values, promises, and challenges of using different methods to make inferences about the quality of teaching and the effectiveness of particular teachers. The course will also examine current policy initiatives (e.g., Race to the Top) and implementations at state and district levels that are designed to improve teacher evaluation systems.

The course will also provide direct experience with a number of evaluation tools in order to better understand the nature of these tools, the insights about teaching that they can provide, and the challenges to using them effectively.

### **Class materials:**

Students are asked to purchase the following books:

Harris, D. (2011). *Value-added methods in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.

Ladson-Billings, G. (2009). *The dreamkeepers: Successful teachers of African American children*. San Francisco, CA: Jossey-Bass Publishers. (students should read this book prior to the first class)

In addition, students will be asked to read a set of papers from the Reading List that will be posted on eCollege and/or are available through web links provided in the citation.

### **Reading List**

#### **Week 1 (September 9): Why the push for better teacher evaluation?**

This week we will encounter some of the drivers that have led to a focus on teacher evaluation as a mechanism for educational improvement. We will review traditional approaches to teacher evaluation. We will also begin to consider the theories of action that guide the current focus on teacher evaluation.

- Toch, T., & Rothman, R. (2008). *Rush to judgment: Teacher evaluation in public education*. Washington, DC: Education Sector. Retrieved from [http://www.educationsector.org/usr\\_doc/RushToJudgment\\_ES\\_Jan08.pdf](http://www.educationsector.org/usr_doc/RushToJudgment_ES_Jan08.pdf)
- Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research*, 73(1), 89–122. Retrieved from <http://www.jstor.org/stable/pdfplus/3516044.pdf>
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. New York, NY: The New Teacher Project. Retrieved from <http://widgeteffect.org/downloads/TheWidgetEffect.pdf>
- Wright, S., Horn, S., & Sanders, W. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11, 57–67. Retrieved from [http://www.sas.com/govedu/edu/teacher\\_eval.pdf](http://www.sas.com/govedu/edu/teacher_eval.pdf)

## **Week 2 (September 16): Conceptions of teacher quality**

We will explore various conceptions of teacher quality and draw a distinction between teacher quality and teaching quality. We will examine the connections and disjunctions between these various conceptions and their implications for measuring teaching practice.

Ball, D. L., & Hill, H. C. (2009). Measuring teaching quality in practice. In D. H. Gitomer (Ed.), *Measurement issues and the assessment for teacher quality* (pp. 80–98). Thousand Oaks, CA: Sage.

Cohen, D. (2010). Teacher quality: An American educational dilemma. In M. M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 375–402). San Francisco, CA: John Wiley & Sons.

\*Hanushek, E. A. (2002). Teacher quality. In L. T. Izumi & W. M. Evers (Eds.), *Teacher quality* (pp. 1–13). Stanford, CA: Hoover Institution Press. Retrieved from <http://hanushek.stanford.edu/sites/default/files/publications/Hanushek%202002%20Teacher%20Quality.pdf>

[Kennedy, M. M. \(2010\). Attribution error and the quest for teacher quality. \*Educational Researcher\*, 39\(8\), 591–598. Retrieved from](http://edr.sagepub.com/content/39/8/591.full.pdf)

### **Week 3 (September 23): Value-added methods – what are they?**

This week we will examine the fundamental assumptions and techniques of value-added models for evaluating teachers. While we will not go into all of the finer technical details, we will learn the basic approaches, nomenclature, and assumptions that underlie models so that students can become literate readers of studies using these techniques.

Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service. Retrieved from <http://www.ets.org/Media/Research/pdf/PICVAM.pdf>

Briggs, D. (2008, November). *The goals and uses of value-added models*. Paper presented at the workshop of the Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability, National Research Council, Washington, DC.

Harris, D. (2011). *Value-added methods in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.

### **Week 4 (September 30): Applications of value-added methods to evaluate teachers**

We will explore some studies that apply value-added methods to the evaluation of teachers. We will examine how researchers test a range of models with different assumptions and variables included in order to make inferences about teacher effectiveness.

Goldhaber, D. D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. *Educational*

*Evaluation and Policy Analysis*, 22(2), 129–145. Retrieved from <http://epa.sagepub.com/content/22/2/129.full.pdf>

Gordon, R., Kane, T. J., & Staiger, D. O. (2006). *Identifying effective teachers using performance on the job* (Hamilton Project Discussion Paper 2006-01). Washington, DC: Brookings Institution. Retrieved from [http://www.brookings.edu/views/Papers/200604hamilton\\_1.pdf](http://www.brookings.edu/views/Papers/200604hamilton_1.pdf)

Hanushek, E. A. (2010, December). *The economic value of higher teacher quality* (NBER Working Paper No. 16606). Cambridge, MA: National Bureau of Economic Research. Retrieved from [http://www.nber.org/papers/w16606.pdf?new\\_window=1](http://www.nber.org/papers/w16606.pdf?new_window=1)

### **Week 5 (October 7): Evaluating value-added methods**

Researchers have widely disparate views about the merits of using value-added methods to evaluate teachers. We will examine the rationales for different positions as well as some of the technical literature that helps us understand the nature of support for the underlying assumptions of value-added methods. How reasonable is it to make claims that a particular teacher “caused” student learning?

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., . . . Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers* (EPI Briefing Paper 278). Washington, DC: Economic Policy Institute. Retrieved from <http://www.epi.org/page/-/pdf/bp278.pdf>

Glazerman, S., Loeb, S., Goldhaber, D., Staiger, D., Raudenbush, S., & Whitehurst, G. (2010). *Evaluating teachers: The important role of value-added*. Washington, DC: Brown Center on Education. Retrieved from [http://www.brookings.edu/~media/research/files/reports/2010/11/17%20evaluating%20teachers/1117\\_evaluating\\_teachers.pdf](http://www.brookings.edu/~media/research/files/reports/2010/11/17%20evaluating%20teachers/1117_evaluating_teachers.pdf)

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794–831. Retrieved from <http://aer.sagepub.com/content/48/3/794.full.pdf>

\*Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy*, 4(4), 492–519. Retrieved from <http://cepa.stanford.edu/sites/default/files/reardon%20raudenbush%20EFP%20VAM%20paper%20resubmission.pdf>

## **Week 6 (October 14): Classroom observation – introducing systems**

Note: During the next three weeks, the class will gain experience in using observation protocols to score video samples of actual classroom practice. This week we will begin to observe teaching practice through video and understand how observation protocols are developed and how they differ from each other.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: ASCD.

Grossman, P. *Protocol for Language Arts Teaching Observation (PLATO)*. (current version with permission)

Hill, H. *Mathematical Quality of Instruction (MQI)*. (current version with permission)

Pianta et al. *Classroom Assessment Scoring System*. (pilot version with permission)

Stodolsky, S. S. (1984). Teacher evaluation: The limits of looking. *Educational Researcher*, 13(9), 11–18.

Walberg, H. J. (1991). Productive teaching and instruction: Assessing the knowledge base. In H. C. Waxman & H. J. Walberg (Eds.), *Effective teaching: Current research* (pp. 33–62). Berkeley, CA: McCutchan.

## **Week 7 (October 21): Classroom observation**

My colleagues and I have been involved in a number of studies of classroom observation. We will be reading these to understand issues associated with measuring teaching quality through observation, including implications for evaluation and professional development.

We will continue to use different protocols to better understand the challenges and potential of using classroom observation to understand teaching.

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

Casabianca, J. M., McCaffrey, D. F., Gitomer, D. H., Bell, C. A., & Hamre, B. K. (in press). Effect of observation mode on measures of secondary mathematics teaching. *Educational and Psychological Measurement*.

Gitomer, D. H., & Bell, C. A. (2013). Evaluating teaching and teachers. In K. F. Geisinger (Ed.), *APA handbook of testing and assessment in psychology* (Vol. 3, pp. 415–444). Washington, DC: American Psychological Association.

Gitomer, D. H., Bell, C. A., Qi, Y., McCaffrey, D. F., Hamre, B. K., & Pianta, R. C. (in press). The instructional challenge in improving teaching quality: Lessons From a Classroom Observation Protocol. *Teachers College Record*.

### **Week 8 (October 28): The validity of observation scores**

How does one establish the validity of using observations to make inferences about teacher quality? During this week we will first discuss what validity means in this context and then examine a number of recent studies that attempt to develop validity evidence for different observation approaches. We will learn about both the methods and the validity evidence for classroom observations.

We will continue to have experiences in using observation protocols.

Bill & Melinda Gates Foundation. (2011). *Learning about teaching: Initial findings from the Measures of Effective Teaching project*. Retrieved from [http://www.metproject.org/downloads/Preliminary\\_Findings-Research\\_Paper.pdf](http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf).

Grossman, P., Loeb, S., Cohen, J., Hammerness, K., Wyckoff, J., Boyd, D., & Lankford, H. (2010, May). *Measure for measure: The relationship between measures of instructional practice in middle school English language arts and*

*teachers' value-added scores* (NBER Working Paper No. 16015). Cambridge, MA: National Bureau of Economic Research. Retrieved from [http://www.nber.org/papers/w16015.pdf?new\\_window=1](http://www.nber.org/papers/w16015.pdf?new_window=1)

Hamre, B. K., Pianta, R. C., Mashburn, A. J., & Downer, J. T. (2007). *Building a science of classrooms: Application of the CLASS framework in over 4,000 U. S. early childhood and elementary classrooms*. Charlottesville, VA: University of Virginia. Retrieved from <http://fcd-us.org/sites/default/files/BuildingAScienceOfClassroomsPiantaHamre.pdf>

Stodolsky, S. S. (1990). Classroom observation. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary school teachers* (pp. 175–190). Newbury Park, CA: Sage.

## **Week 9 (November 4): Classroom artifacts**

Another aspect of classroom practice that we can look at to understand teacher quality is the nature of assignments and assessments that teachers ask students to do. This week we will learn about how researchers have used artifacts to study classroom practice, how practice is sampled, and what studies have been found about the quality of teaching using such instruments.

We will also gain experience in using an assignment protocol to score classroom assignments and student work.

Borko, H., Stecher, B., & Kuffner, K. (2007). *Using artifacts to characterize reform-oriented instruction: The Scoop Notebook and rating guide* (CSE Technical Report 707). Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing (CRESST)/UCLA. Retrieved from <http://www.eric.ed.gov/PDFS/ED495853.pdf>

Matsumura, L. C., Garnier, H., Slater, S. C., & Boston, M. (2008). Toward measuring instructional interactions at-scale. *Educational Assessment*, 13(4), 267–300.

Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago, IL: Consortium on Chicago School Research. Retrieved from <http://ccsr.uchicago.edu/sites/default/files/publications/p0a02.pdf>

Wenzel, S., Nagaoka, J. K., Morris, L., Billings, S., & Fendt, C. (2002). *Documentation of the 1996–2002 Chicago Annenberg Research Project strand on authentic intellectual demand exhibited in assignments and student work: A technical process manual*. Chicago, IL: Consortium on Chicago School Research. Retrieved from <http://ccsr.uchicago.edu/sites/default/files/publications/p67.pdf>

## **Week 10 (November 11): Content knowledge for teaching**

What is the nature of knowledge that teachers need to have in order to help students learn? Measures of subject-matter knowledge have long been used for the initial certification of teachers using tests like *Praxis*. This week we will examine these

traditional measures and contrast them with new models of knowledge assessments that pay much greater attention to an idea broadly defined as pedagogical content knowledge. We will examine studies of content knowledge for teaching and also experience some of these assessments.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389–407. Retrieved from <http://jte.sagepub.com/content/59/5/389.full.pdf>

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition & Instruction*, 26(4), 430–511. Retrieved from <http://www.tandfonline.com/doi/pdf/10.1080/07370000802177235>

Gitomer, D. H., Howell, H., Phelps, G., Weren, B., & Croft, A. (draft). Evidence on the validity of content knowledge for teaching assessments. To appear in K. Kerr, R. Pianta, & T. Kane (Eds.), *Measures of effective teaching research volume*.

### **Week 11 (November 18): Emerging teacher evaluation systems**

States and districts across the country are now developing teacher evaluation systems that use a combination of measures we have explored thus far. This week we will explore several models that are being implemented and how scores from different measures are being combined to make judgments about teachers. We will also examine findings related to developing compensation systems based on measures of teaching quality.

Bill and Melinda Gates Foundation. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains* (MET Project Research Paper). Seattle, WA: Author. Retrieved from [http://www.metproject.org/downloads/MET\\_Gathering\\_Feedback\\_Research\\_Paper.pdf](http://www.metproject.org/downloads/MET_Gathering_Feedback_Research_Paper.pdf)

Chait, R., & Miller, R. (2009). *Paying teachers for results: A summary of research to inform the design of pay-for-performance programs for high-poverty schools*. Washington, DC: Center for American Progress. Retrieved from [http://www.americanprogress.org/issues/2009/05/pdf/performance\\_pay.pdf](http://www.americanprogress.org/issues/2009/05/pdf/performance_pay.pdf)

Headen, S. (2011). *Inside IMPACT: D.C.'s model teacher evaluation system* (Education Sector Reports). Washington, DC: Education Sector. Retrieved from [http://www.educationsector.org/sites/default/files/publications/IMPACT\\_Report\\_RELEASE.pdf](http://www.educationsector.org/sites/default/files/publications/IMPACT_Report_RELEASE.pdf)

Springer, M. G., Ballou, D., Hamilton, L., Le, V., Lockwood, J. R., McCaffrey, D., . . . Stecher, B. (2010). *Teacher pay for performance: Experimental evidence from the Project on Incentives in Teaching*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University. Retrieved from [http://www.rand.org/content/dam/rand/pubs/reprints/2010/RAND\\_RP1416.pdf](http://www.rand.org/content/dam/rand/pubs/reprints/2010/RAND_RP1416.pdf)

Steele, J. L., Hamilton, L. S., & Stecher, B. M. (2010). *Incorporating student performance measures into teacher evaluation systems* (RAND Corporation Technical Report Series). Santa Monica, CA: RAND Corporation. Retrieved from [http://www.rand.org/pubs/technical\\_reports/2010/RAND\\_TR917.pdf](http://www.rand.org/pubs/technical_reports/2010/RAND_TR917.pdf)

## **Week 12 (November 25): National Board for Professional Teaching Standards (NBPTS)**

The National Board for Professional Teaching Standards was established 20 years ago to create an advanced certification system for highly accomplished teachers. It was a voluntary system and attempted to emulate the same type of board certification as existed in other professions. Becoming board certified required teachers to participate in a year-long comprehensive assessment that included videos, student work, and a great deal of teacher reflection on their work. This week we will examine both the characteristics of the NBPTS assessments as well as research findings on the validity of the NBPTS assessments.

Cantrell, S., Fullerton, J., Kane, T. J., & Staiger, D. O. (2008). *National Board certification and teacher effectiveness: Evidence from a random assignment*

*experiment* (NBER Working Paper No. 14608). Cambridge, MA: National Bureau of Economic Research. Retrieved from [http://www.nber.org/papers/w14608.pdf?new\\_window=1](http://www.nber.org/papers/w14608.pdf?new_window=1)

Gitomer, D. H. (2008). Reliability and NBPTS assessments. In L. Ingvarson & J. Hattie (Eds.), *Assessing teachers for professional certification: The first decade of the National Board for Professional Teaching Standards* (pp. 231–253). New York, NY: Elsevier.

National Research Council. (2008). *Assessing accomplished teaching: Advanced-level certification programs*. Committee on Evaluation of Teacher Certification by the National Board for Professional Teaching Standards. M. D. Hakel, J. A. Koenig, & S. W. Elliott (Eds.). Washington, DC: National Academies Press.

Pearlman, M. (2008). The design architecture of NBPTS certification assessments. In L. Ingvarson & J. Hattie (Eds.), *Assessing teachers for professional certification: The first decade of the National Board for Professional Teaching Standards* (pp. 55–91). New York, NY: Elsevier.

The following are documents from the NBPTS system:

National Board for Professional Teaching Standards. (2003). *NBPTS adolescence and young adulthood English language arts standards, 2<sup>nd</sup> edition*. Retrieved from <http://www.nbpts.org/sites/default/files/documents/certificates/nbpts-certificate-aya-ela-standards.pdf>

National Board for Professional Teaching Standards. (2010). *Mathematics standards*. Retrieved from <http://www.nbpts.org/sites/default/files/documents/certificates/nbpts-certificate-ea-math-standards.pdf>

National Board for Professional Teaching Standards. (2012). *Middle childhood generalist standards, 3<sup>rd</sup> edition*. Retrieved from <http://www.nbpts.org/sites/default/files/documents/certificates/nbpts-certificate-mc-gen-standards.pdf>

**Week 13 (December 2): Educational accountability**

What does it mean for an educational system to be accountable? This week we examine a number of perspectives on accountability, including those that view accountability as a shared responsibility of multiple stakeholders in the system. We will consider *who* is accountable for *what* and to *whom*.

Gitomer, D. H. (2008). Crisp measurement and messy context: A clash of assumptions and metaphors – Synthesis of Section III. In D. H. Gitomer (Ed.), *Measurement issues and the assessment for teacher quality* (pp. 223–233). Thousand Oaks, CA: Sage.

Jacob, B. A. (2011). Do principals fire the worst teachers? *Educational Evaluation and Policy Analysis*, 33(4), 403–434. Retrieved from <http://epa.sagepub.com/content/33/4/403.full.pdf>

Ladd, H. F. (2007, November). *Holding schools accountable revisited*. Paper presented at APPAM Fall Research Conference, Washington, DC.

Ladson-Billings, G. (2008). Opportunity to teach: Teacher quality in context. In D. H. Gitomer (Ed.), *Measurement issues and the assessment for teacher quality* (pp. 206–222) Thousand Oaks, CA: Sage.

Stecher, B. M., Vernez, G., & Steinberg, P. (2010). *Reauthorizing No Child Left Behind: Facts and recommendations* (RAND Corporation Report Series). Santa Monica, CA: RAND. Retrieved from [http://www.rand.org/pubs/monographs/2010/RAND\\_MG977.pdf](http://www.rand.org/pubs/monographs/2010/RAND_MG977.pdf)

### **Week 14 (December 9): Group presentations**

## **Grading and Activities**

Students will be evaluated on their participation in class (30%) and 3 assignments.

Assignment 1 (20%) – Students will be asked to provide a synthetic and critical summary of a set of papers provided by the instructor that addresses, from different perspectives, a key issue focused on the evaluation of teaching.

Assignment 2 (20%) – Students will be asked to select a particular instrument that is used in teaching evaluation and provide an analysis that considers its conceptual basis, core values and assumptions, and validity evidence.

Assignment 3 (30%) – Students (in small groups) will examine a current statewide or district-wide evaluation system and provide a critical analysis of the characteristics, theory of action, strengths, weaknesses, challenges, and potential risks associated with the evaluation system. Analysis will be submitted in written format and each group will present their findings to the class.

## **Description of Activities**

Most classes will include an orienting lecture by the instructor followed by a discussion of the readings. Effective discussions are a shared responsibility for all members of the class. Students are expected to read the materials thoroughly, to consider orienting questions provided with the readings, to contribute thoughtful questions, and to participate fully in the discussions. Students will also be asked to establish connections between the readings and contemporary assessment issues that arise from their own experiences, current policy issues, and/or other research studies in an online discussion following each class.

A number of the articles are fairly technical in nature and are identified with an asterisk (\*). I would like you to try to read and have some familiarity with these papers even though it is likely you will not fully understand them. We will review these papers in detail during class.

In several classes, students will learn about specific tools for evaluating classroom observations and classroom assignments. We will study the protocols that guide the scoring of teaching and use them to make judgments about real samples from classrooms.

