

March 26, 2012

Misconceptions and Realities about Teacher and Principal Evaluation

An Open Letter of Concern to

*Mayor Rahm Emanuel,
Chicago Public Schools CEO Jean-Claude Brizard,
and the Chicago School Board*

Regarding Chicago's Implementation of Legislation for the Evaluation of Teachers and Principals

Chicago Public Schools (CPS) plans to implement dramatic changes in the 2012-2013 school year. As university professors and researchers who specialize in educational research, we recognize that change is an essential component of school improvement. We are very concerned, however, at a continuing pattern of changes imposed rapidly without high-quality evidentiary support.

The new evaluation system for teachers and principals centers on misconceptions about student growth, with potentially negative impact on the education of Chicago's children. We believe it is our ethical obligation to raise awareness about how the proposed changes not only lack a sound research basis, but in some instances, have already proven to be harmful.

In this letter, we describe our concerns and relevant research as we make two recommendations for moving forward:

1. **Pilot and adjust the evaluation system before implementing it on a large scale.**
2. **Minimize the percentage that student growth counts in teacher or principal evaluation.**

We also urge consulting on the above steps with the professors and researchers among us who bring both scholarly and practical expertise on these issues.

Background

In January 2010, the Illinois State Legislature—in an effort to secure federal Race to the Top funds—approved an amendment to the Illinois School Code known as the Performance Evaluation Review Act (PERA), which requires districts to include “student growth” as a significant portion of teacher and principal evaluation. While most of the state does not have to implement a new evaluation plan for teachers until 2016, CPS was able to get written into the law an early implementation date of September 2012 for at least 300 schools.

The proposed rules associated with PERA will not be finalized until April 2012 at the earliest. Nevertheless, CPS is moving ahead with teacher and principal evaluation plans based on the proposals. The suggested rules define “significant” use of student growth as at least 25% of a principal's or teacher's evaluation in the first two years of implementation, and 30% after that, with the possibility of making student growth count for as much as 50%.

The PERA law mandates that multiple measures of student growth be used in teacher evaluation. The proposed rules identify three types of measures: standardized tests administered beyond Illinois (Type I), assessments approved for use districtwide (Type II), and classroom assessments aligned to curriculum (Type III). Under the proposed rules, every teacher's student growth will be determined through the use of at least one Type III assessment, which means that two Type IIIs would be used if no Type I or II is appropriate.

In what follows, we draw on research to describe three significant concerns with this plan.

Concern #1: CPS is not ready to implement a teacher-evaluation system that is based on significant use of "student growth."

For Type I or Type II assessments, CPS must identify the assessments to be used, decide how to measure student growth on those assessments, and translate student growth into teacher-evaluation ratings. They must determine how certain student characteristics such as placement in special education, limited English-language proficiency, and residence in low-income households will be taken into consideration. They have to make sure that the necessary technology is available and usable, guarantee that they can correctly match teachers to their actual students, and determine that the tests are aligned to the new Common Core State Standards (CCSS). In addition, teachers, principals, and other school administrators have to be trained on the use of student assessments for teacher evaluation. This training is on top of training already planned about CCSS and the Charlotte Danielson Framework for Teaching, used for the "teacher practice" part of evaluation.

For most teachers, a Type I or II assessment does not exist for their subject or grade level, so most teachers will need a Type III assessment. While work is being done nationally to develop what are commonly called assessments for "non-tested" subjects, this work is in its infancy. CPS must identify at least one Type III assessment for every grade and every subject, determine how student growth will be measured on these assessments, and translate the student growth from these different assessments into teacher-evaluation ratings in an equitable manner.

If CPS insists on implementing a teacher-evaluation system that incorporates student growth in September 2012, we can expect to see a widely flawed system that overwhelms principals and teachers and causes students to suffer.

Concern #2: Educational research and researchers strongly caution against teacher-evaluation approaches that use Value-Added Models (VAMs).

Chicago already uses a VAM statistical model to determine which schools are put on probation, closed, or turned around. For the new teacher-evaluation system, student growth on Type I or Type II assessments will be measured with VAMs or similar models. Yet, ten prominent researchers of assessment, teaching, and learning recently wrote an open letter that included some of the following concerns about using student test scores to evaluate educators¹:

- a. *Value-added models (VAMs) of teacher effectiveness do not produce stable ratings of teachers. For example, different statistical models (all based on reasonable*

¹ Baker, E., et al. (2011). *Correspondence to the New York State Board of Regents*. Retrieved October 16, 2011 from http://www.washingtonpost.com/blogs/answer-sheet/post/the-letter-from-assessment-experts-the-ny-regentsignored/2011/05/21/AFJHIA9G_blog.html.

assumptions) can yield different effectiveness scores.² Researchers have found that how a teacher is rated changes from class to class, from year to year, and even from test to test.³

- b. There is no evidence that evaluation systems that incorporate student test scores produce gains in student achievement. In order to determine if there is a relationship, researchers recommend small-scale pilot testing of such systems. Student test scores have not been found to be a strong predictor of the quality of teaching as measured by other instruments or approaches.⁴
- c. Assessments designed to evaluate student learning are not necessarily valid for measuring teacher effectiveness or student learning growth.⁵ Using them to measure the latter is akin to using a meter stick to weigh a person: you might be able to develop a formula that links height and weight, but there will be plenty of error in your calculations.

Concern #3: Students will be adversely affected by the implementation of this new teacher-evaluation system.

When a teacher's livelihood is directly impacted by his or her students' scores on an end-of-year examination, test scores take front and center. The nurturing relationship between teacher and student changes for the worse, including in the following ways:

- a. With a focus on end-of-year testing, there inevitably will be a narrowing of the curriculum as teachers focus more on test preparation and skill-and-drill teaching.⁶ Enrichment activities in the arts, music, civics, and other non-tested areas will diminish.
- b. Teachers will subtly but surely be incentivized to avoid students with health issues, students with disabilities, students who are English Language Learners, or students suffering from emotional issues. Research has shown that no model yet developed can adequately account for all of these ongoing factors.⁷
- c. The dynamic between students and teacher will change. Instead of "teacher and student versus the exam," it will be "teacher versus students' performance on the exam."

² Papay, J. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163-193.

³ McCaffrey, D., et al. (2004). *Evaluating value-added models of teacher accountability*. Santa Monica, CA: Rand Corporation.

⁴ See Burris, C., & Welner, K. (2011). Conversations with Arne Duncan: Offering advice on educator evaluations. *Phi Delta Kappan*, 93(2), 38-41.

⁵ Goe, L., & Holdheide, L. (2011). *Measuring teachers' contributions to student learning growth for nontested grades and subjects*. Retrieved February 2, 2012 from <http://www.tqsource.org/publications/MeasuringTeachersContributions.pdf>.

⁶ Committee on Incentives and Test-Based Accountability in Education of the National Research Council. (2011). *Incentives and Test-Based Accountability in Education*. Washington, DC: National Academies Press.

⁷ Baker, E., et al (2010). *Problems with the use of test scores to evaluate teachers*. Washington, DC: Economic Policy Institute. Retrieved October 16, 2011 from http://epi.3cdn.net/b9667271ee6c154195_t9m6iij8k.pdf; Newton, X., et al. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy and Analysis Archives*. Retrieved October 16, 2011 from <http://epaa.asu.edu/ojs/article/view/810/858>; Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, 4(4), 537-571.

- d. *Collaboration among teachers will be replaced by competition. With a “value-added” system, a 5th grade teacher has little incentive to make sure that his or her incoming students score well on the 4th grade exams, because incoming students with high scores would make his or her job more challenging.*
- e. *When competition replaces collaboration, every student loses.*

Our Recommendations

1. Pilot and adjust the evaluation system before implementing it on a large scale.

Any annual evaluation system should be piloted and adjusted as necessary based on field feedback before being put in place citywide. In other words, Chicago should pilot models and then use measures of student learning to evaluate the model. Delaware spent years piloting and fine-tuning their system before putting it in place formally statewide. Conversely, Tennessee’s teacher-evaluation system made headlines when its hurried implementation led to unintended negative consequences.

2. Minimize the percentage that student growth counts in teacher or principal evaluation.

Until student-growth measures are found to be valid and reliable sources of information on teacher or principal performance, they should not play a major role in summative ratings. Teacher-practice instruments, such as the Charlotte Danielson Framework, focus on what a teacher does and how practice can be strengthened. Students benefit when objective feedback is part of their teachers’ experience. Similar principal frameworks serve the same purpose.

We, Chicago-area university professors and researchers who specialize in educational research, conclude that hurried implementation of teacher evaluation using student growth will result in inaccurate assessments of our teachers, a demoralized profession, and decreased learning among and harm to the children in our care. It is wasteful of increasingly limited resources to implement systemwide a program that has not yet been field-tested. Our students are more than the sum of their test scores, and an overemphasis on test scores will *not* result in increased learning, increased well-being, and greater success. According to a nine-year study by the National Research Council⁸, the past decade’s emphasis on testing has yielded little learning progress, especially considering the cost to our taxpayers.

We support accountability and high standards. We want what is best for our students. We believe, however, that an unproven and potentially harmful evaluation system is not the path to lasting school improvement. We must not lose sight of what matters the most—the academic, social, and emotional growth and well-being of Chicago’s children.⁹

* * *

⁸ Committee on Incentives and Test-Based Accountability in Education of the National Research Council. (2011). *Incentives and Test-Based Accountability in Education*. Washington, DC: National Academies Press.

⁹ Note: This letter was adapted from the letter written by Sean C. Feeney, Ph.D. and Carol C. Burris, Ed.D., which was signed by more than 1400 New York principals in opposition to New York’s evaluation plan. <http://www.newyorkprincipals.org>.

Signed by 88 educational researchers across Chicagoland, as of March 26, 2012. University affiliations are listed for identification purposes only.

1. (Primary Contact) Kevin Kumashiro, University of Illinois at Chicago, kevink@uic.edu, 312-996-8530
2. Ann Aviles de Bradley, Northeastern Illinois University
3. William Ayers, University of Illinois at Chicago
4. Martha Biondi, Northwestern University
5. Leslie Rebecca Bloom, Roosevelt University
6. Robert Anthony Bruno, University of Illinois at Urbana-Champaign
7. Brian Charles Charest, University of Illinois at Chicago
8. Amina Chaudhri, Northeastern Illinois University
9. Ronald E. Chennault, DePaul University
10. Sumi Cho, DePaul University
11. Katherine Copenhaver, Roosevelt University
12. Gabriel Cortez, Northeastern Illinois University
13. Todd DeStigter, University of Illinois at Chicago
14. Renee Dolezal, University of Illinois at Chicago
15. Sarah Donovan, University of Illinois at Chicago
16. Aisha El-Amin, University of Illinois at Chicago
17. Stephanie Farmer, Roosevelt University
18. Rocío Ferreira, DePaul University
19. Joby Gardner, DePaul University
20. Erik Gellman, Roosevelt University
21. Judith Gouwens, Roosevelt University
22. Eric Gutstein, University of Illinois at Chicago
23. Horace R. Hall, DePaul University
24. Cecily Relucio Hensler, University of Chicago
25. Peter B. Hilton, Saint Xavier University
26. Lauren Hoffman, Lewis University
27. Marvin Hoffman, University of Chicago
28. Nicole Holland, Northeastern Illinois University
29. Amy Feiker Hollenbeck, DePaul University
30. Stacey Horn, University of Illinois at Chicago
31. Diane Horwitz, DePaul University
32. Marie Tejero Hughes, University of Illinois at Chicago
33. Seema Iman, National Louis University
34. Valerie C. Johnson, DePaul University
35. Susan Katz, Roosevelt University
36. Bill Kennedy, University of Chicago
37. Jung Kim, Lewis University
38. Michael Klonsky, DePaul University
39. Pamela J. Konkol, Concordia University Chicago
40. Emily E. LaBarbera-Twarog, University of Illinois at Urbana-Champaign
41. Crystal Laura, Chicago State University
42. Pauline Lipman, University of Illinois at Chicago
43. Alberto Lopez, Northeastern Illinois University
44. Norma Lopez-Reyna, University of Illinois at Chicago
45. Antonina Lukenchuk, National Louis University
46. Christina L. Madda, Northeastern Illinois University
47. Eleni Makris, Northeastern Illinois University
48. Christine Malcom, Roosevelt University

49. Kathleen McNerney, Saint Xavier University
50. Elizabeth Meadows, Roosevelt University
51. Erica R. Meiners, Northeastern Illinois University
52. Marlene V. Meisels, Concordia University Chicago
53. Gregory Michie, Concordia University Chicago
54. Daniel Miltner, University of Illinois at Chicago
55. Tom Moher, University of Illinois at Chicago
56. Carol Myford, University of Illinois at Chicago
57. Isabel Nuñez, Concordia University Chicago
58. Tammy Oberg De La Garza, Roosevelt University
59. Esther Ohito, University of Chicago
60. Tema Okun, National Louis University
61. Irma Olmedo, University of Illinois at Chicago
62. Bradley Porfilio, Lewis University
63. Amira Proweller, DePaul University
64. Isaura B. Pulido, Northeastern Illinois University
65. Therese Quinn, School of the Art Institute of Chicago
66. Eileen Quinn Knight, Saint Xavier University
67. Josh Radinsky, University of Illinois at Chicago
68. Arthi Rao, University of Illinois at Chicago
69. Dale Ray, University of Chicago
70. Sarah Maria Rutter, University of Illinois at Chicago
71. Karyn Sandlos, School of the Art Institute of Chicago
72. William H. Schubert, University of Illinois at Chicago
73. Brian D. Schultz, Northeastern Illinois University
74. Amy Shuffleton, University of Wisconsin at Whitewater
75. Noah W. Sobe, Loyola University Chicago
76. Sonia Soltero, DePaul University
77. Gerri Spinella, National Louis University
78. David Stovall, University of Illinois at Chicago
79. Simeon Stumme, Concordia University Chicago
80. Tom Thomas, Roosevelt University
81. Richard M. Uttich, Roosevelt University
82. Robert Wagreeich, University of Illinois at Chicago
83. Frederico Waitoller, University of Illinois at Chicago
84. Norman Weston, National Louis University
85. Daniel White, Roosevelt University
86. Jeff Winter, National Louis University
87. Chyrese S. Wolf, Chicago State University
88. Kate Zilla, National Louis University

This letter can be downloaded at <http://www.createchicago.blogspot.com>.