Testimony Regarding
S.B. 24: An Act Concerning Educational Competitiveness (Teacher Evaluation Provisions)
Robert Cotto, Jr., Ed.M.
Education Committee
February 21, 2012

Senator Stillman, Representative Fleischmann, and distinguished Members of the Education Committee:

I am testifying today on behalf of Connecticut Voices for Children, a research-based public education and advocacy organization that works statewide to promote the well-being of Connecticut's children, youth, and families. **Connecticut Voices for Children advocates for a balanced and broad approach to public schools that supports all children's learning and well-being**.

The Governor's Bill on education, S.B. 24, proposes to use the recently adopted teacher evaluation guidelines for the purposes of promotion, certification, and other benefits for teachers. CT Voices for Children urges caution to avoid unintended, negative consequences of this evaluation system, on vulnerable students such as students with disabilities, language minority students, and low-income students.

Our schools should continue to provide **a broad educational program that serves all children's learning and development.** The goals may include, but are not limited to, academic skills, critical thinking, the arts and literature, preparation for skilled work, social skills and work ethic, citizenship, and emotional health.[i] A well-crafted teacher evaluation process is a part of ensuring that public schools meet these goals. The teacher evaluation provisions of this bill may fall short of ensuring these broad goals if they rely too heavily on results from the CMT and/or CAPT.

The bill establishes four designators for the purposes of teacher evaluation: below standard, developing, proficient and exemplary. Based on guidelines developed by PEAC, the designators are dependent on the results of standardized tests such as the CMT and CAPT, as well as observations and other indicators. Furthermore, the bill defines "effective practice" as, "a rating of developing, proficient or exemplary under the evaluation system …."[ii] Because up to 30% of these teacher evaluations must be based on standardized test scores, "effective practice" will in significant part be determined by teachers' ability to get students to improve on standardized tests of basic skills.

We welcome the inclusion of broad evidence of student learning in these evaluations and the use of observations to judge teacher skill. There are a number of other types of authentic student assessments, in addition to test results, that should be included in teacher evaluation systems. Such assessments that could be part of a teacher portfolio of student work might include,[iii] but are not limited to: district-wide and department exams, tests, quizzes, papers and projects.[iv] Unlike standardized test results, which can only offer a small and static snapshot of student performance, the types of assessments mentioned above offer a more dynamic and complete picture that can more accurately capture student growth throughout a school year. The CMT and CAPT only exist for four subjects, but alternative authentic assessments can also apply to the broad goals and subject areas that schools offer and should be responsible for providing all students.

Managing teachers by using student test scores can create incentives to narrow educational goals, or "teach to the test". There is evidence that Connecticut's schools, being responsible for only CMT and CAPT scores, have shifted resources and instruction to subject areas that are part of the CMT at the expense of a rich, broad curriculum. From 2002-2009, Connecticut's second graders have had on average fewer hours of instruction in Computer Education, Health, Language Arts, and Social Studies and more hours in mathematics.[v] Between 2002 and 2011, fifth graders have gradually experienced fewer hours of social studies and health, while having more hours of instruction in math-a tested subject.[vi] Over the last three years, 8th grade students in Connecticut have also seen increases in the number of instructional hours for math and reading, while other subjects have fluctuated.[vii] Given the responsibility to raise math, reading, and writing test scores, the state's schools and districts have shifted the focus of the instruction, on average, to these tested subjects.

It is also is important to note that whatever growth is charted must be **balanced by information about the context of the department, school and district where it occurs.** The culture of these overlapping jurisdictions is vital in fostering the success of all students, so any evaluation must also consider the environment in which learning is taking place. [viii] Achievement indicators such as standardized test scores are summary measures of school, home, personal, and community factors.[ix] Indeed, standardized test scores tend to reflect a combination of, "what students bring to school," and what they learn in the classroom.[x]

Test information, skillfully interpreted, can often help teachers, administrators, parents, and students to identify the strengths and weaknesses of students in basic English literacy skills. Nevertheless, the CMT and CAPT were not designed to judge teacher skill or knowledge. Since students are not distributed randomly or equitably within classrooms, between schools, and across districts, using standardized test scores as an indicator for teacher evaluation could penalize those teachers who work with the least prepared students in the most challenging classroom conditions.[xi] This is particularly true for tests that only measure absolute, or standards-based levels, rather than indicators of relative improvement over time.

Along this same line, since most standardized testing currently focuses on verbal/reading skills and mathematics, it is important to be careful not to unintentionally negatively affect students whose first language is not English or who do not speak English at home, as well as the teachers of these students. In one study of a Texas school with a majority of Latino students and families, teachers indicated that the introduction of a new test-based management system contributed to lost instructional time, undermined student-teacher relationships, narrowed curriculum, and led to less skillful teaching.[xii] Given the growth of the Latino children population in Connecticut, a population that is both bilingual and bicultural, an evaluation system for teachers based on English literacy tests may disproportionately affect this community.

A further important concern is to ensure that standardized test results are transparent and accurate. Recently, CT Voices for Children examined the exclusion of students with disabilities from the standard CMT in a report entitled "Addition through Subtraction: Are Rising Test Scores in Connecticut School Districts Related to the Exclusion of Students with Disabilities?," (Executive Summary attached to this testimony.) This report concluded that the exclusion of thousands of students with disabilities from reported Connecticut Mastery Test results distorted reported trends in test scores. To ensure more valid, or "apples to apples" comparisons of trends over time, Connecticut Voices for Children recommended that state officials clarify the impact of the exclusion of students with disabilities when reporting on changes in CMT scores over time. Further, we suggested that policymakers:

- Use a variety of indicators, not just standardized test scores, to evaluate improvements in public education; and
- Reconsider policies that assign rewards and punishments based on these test scores.

Given these and other concerns, the Board on Testing and Assessment of the National Research Council of the National Academy of Sciences concluded that estimates of teacher effectiveness based on growth in student test scores, "should not be used to make operational decisions because such estimates are far too unstable to be considered fair or reliable."[xiii] A study commissioned by the U.S. Department of Education also concluded that policymakers must carefully consider using test-score data to make high-stakes decisions about educators because the statistical evidence was prone to a substantial amount of error.[xiv] In short, even powerful statistical models of test-score growth often misidentified skillful teachers as ineffective.

Connecticut Voices for Children urges policy makers to carefully consider the appropriate role of test scores in overall teacher evaluation[xv], and balance this with other indicators of teacher performance, such as in-class observations and a portfolio of work, in order to minimize the potential for unreliable results.[xvi]

The provisions of this bill related to teacher evaluation, certification, and promotion raise several questions:

- Is there enough evidence at this time to institute the use of standardized test scores as a basis for teacher evaluation, certification, and promotion?
- How will the state ensure that the most vulnerable students will not experience narrowed curriculum, exclusion, and/or lower-quality instruction because of teacher evaluations based, in part, on student test scores?
- How will the challenges of working with students with disabilities, low-income students, and language minority students, be included as mitigating factors in the evaluation process?

Ultimately, the goal of teacher evaluations should be to **ensure that students are receiving the highest level of instruction possible, to support learning new knowledge and skills and develop students' talents, abilities, and potential.** Connecticut should proceed cautiously in its implementation of such a system as proposed in this bill; acting with care will help to ensure that the chosen process is crafted to reflect the values of our community. Evaluation should connect to preparation and teacher development. The state should look to the wealth of existing research to implement a system of teacher evaluation that relies on a balanced approach, including in-class observation and a portfolio of student work, to **cultivate skilled teachers who will continue to grow and improve over time.**[xvii]

---

[i] Rothstein, Jacobsen, and Wilder. *Grading Education: Getting Accountability Right.* Economic Policy Institute; Washington, D.C. Teachers College Press; New York, NY: 2008. *See* Chapter 2 on "Weighting the public education."

[ii] *See* SB 24, Sec. 30 (d)

[iii] Many of these suggestions come from the PAR system as implemented in Rockville, Maryland, available at http://www.montgomeryschoolsmd.org/departments/ development/ documents/TeacherPGS_handbook.pdf.

[iv] Such assessments are more likely to reflect student achievement within the context of the relevant curriculum set by the district. They also avoid the oft-cited pitfall of teachers "teaching to the test," where teachers narrow what is studied to only include subjects likely to be evaluated on standardized tests.

[v] Connecticut State Department of Education. *Connecticut Education Data and Research (CEDaR) Data Tables.* "Hours of Instruction by Subject Area-Grade 2." CT Department of Education, 2011. Web. 1 Dec. 2011. <http://sdeportal.ct.gov/Cedar/WEB/ct_report/DTHome.aspx> There was also a small increase in hours of instruction of instruction for foreign language.

[vi] Connecticut State Department of Education. *Connecticut Education Data and Research (CEDaR) Data Tables.* "Hours of Instruction by Subject Area-Grade 5." CT Department of Education, 2011. Web. 1 Feb. 2012. <http://sdeportal.ct.gov/Cedar/WEB/ct_report/DTHome.aspx>

[vii] Connecticut State Department of Education. *Connecticut Education Data and Research (CEDaR) Data Tables.* "Hours of Instruction by Subject Area-Grade 8." CT Department of Education, 2011. Web. 1 Dec. 2011. <http://sdeportal.ct.gov/Cedar/WEB/ct_report/DTHome.aspx>

[viii] *E.g.* Is there a culture of exclusion or inclusion? Is the culture positive or punitive? Is it parent-friendly or parent-intimidating?

[ix] Harris, Angel L. *Kids Don't Want To Fail: Oppositional Culture and the Black-White Gap.* Harvard University Press; Cambridge, MA; 2011." *Kindle* edition. Harris location 2144 of 3588. *Kindle* Edition.

[x] Popham, W. J. (1999). Why standardized tests don't measure educational quality. Educational Leadership, 56(6), 8– 15.

[xi] Pamela D. Tucker and James H. Stronge, *Linking Teacher Evaluation and Student Learning*, Association for Supervision and Curriculum Development, 2005 at 98. *See also* Peter Z. Schochet and Hanley S. Chian, *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains*, U.S. Dept. of Education, *available at* http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf.

[xii] Sloan. Kris. "Playing to the Logic of the Texas Accountability System: How Focusing on "Ratings"-Not Children-Undermines Quality and Equity." *Leaving Children Behind: How "Texas Style" Accountability Fails Latino Youth.* Valenzuela et. al. Albany, NY: State University of New York Press, 2005. Print.
See page 160-171.

[xiii] *See* Economic Policy Institute, Problems with the Use of Student Test Scores to Evaluate Teachers (Aug. 29, 2010)

[xiv] Schochet, Peter Z. and Hanley S. Chiang (2010). *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains* (NCEE 2010-4004). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

[xv] Rothstein, Jesse. "Teacher Quality In Educational Production: Tracking, Decay, and Student Achievement*" *The Quarterly Journal of Economics.* Cambridge, MA: 2010 February, 175-214. Retrieved 13 Jan. 2012. Rothstein concludes on page 211, "First, and most clearly, the stakes attached to VAM- based measures should be relatively small."

[xvi] See Daniel F. McCaffrey, et. al., *Evaluating Value-Added Models for Teacher Accountability* (2003) RAND Corporation, *available at* http://www.rand.org/pubs/monographs/2004/RAND_MG158.pdf. *See also* Pamela D. Tucker and James H. Stronge, *Linking Teacher Evaluation and Student Learning*, Association for Supervision and Curriculum Development, 2005 ("We maintain that measures of student learning are vitally important to judging the effectiveness of teachers and schools, but should never usurp professional judgment that integrates knowledge of other factors that affect instruction, such as the lack of resources, overcrowding, and community poverty. Teaching and learning are far too complex to be reduced to a single test result or even a battery of tests. Tests, however, can serve as indicators of other problems in specific classrooms or schools that need to be addressed through staff development, teacher mentoring, greater resources, or reorganization of time and curriculum.").

[xvii] According to the New Haven Teacher Evaluation Process Update, given on September 12, 2011, "The evaluation and development process can, should and must continue to strengthen … there are real opportunities to improve and build on the process," *available at* http://www.newhavenindependent.org/archives/upload/2011/09/NHPS_Talent_ Update_110912.pdf.