

**Testimony Regarding**

**S.B. 24: An Act Concerning Educational Competitiveness (Teacher Evaluation Provisions)**

Robert Cotto, Jr., Ed.M.

Education Committee

February 22, 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.<sup>i</sup> 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 . . ."<sup>iii</sup> 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,<sup>iii</sup> but are not limited to: district-wide and department exams, tests, quizzes, papers and projects.<sup>iv</sup> 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.<sup>v</sup> 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.<sup>vi</sup> Over the last three years, 8<sup>th</sup> grade students in Connecticut have also seen increases in the number of instructional hours for math and reading, while other subjects have fluctuated.<sup>vii</sup> 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 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.<sup>viii</sup> Achievement indicators such as standardized test scores are summary measures of school, home, personal, and community factors.<sup>ix</sup> Indeed, standardized test scores tend to reflect a combination of, “what students bring to school,” and what they learn in the classroom.<sup>x</sup>

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.<sup>xi</sup> 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.<sup>xii</sup> 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.”<sup>xiii</sup> 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.<sup>xiv</sup> 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<sup>xv</sup>, 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.<sup>xvi</sup>

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.**<sup>xvii</sup>

---

<sup>i</sup> 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.”

<sup>ii</sup> See SB 24, Sec. 30 (d)

<sup>iii</sup> 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](http://www.montgomeryschoolsmd.org/departments/development/documents/TeacherPGS_handbook.pdf).

<sup>iv</sup> 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.

---

<sup>v</sup> 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](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.

<sup>vi</sup> 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](http://sdeportal.ct.gov/Cedar/WEB/ct_report/DTHome.aspx)>

<sup>vii</sup> 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](http://sdeportal.ct.gov/Cedar/WEB/ct_report/DTHome.aspx)>

<sup>viii</sup> E.g. Is there a culture of exclusion or inclusion? Is the culture positive or punitive? Is it parent-friendly or parent-intimidating?

<sup>ix</sup> 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.

<sup>x</sup> Popham, W. J. (1999). Why standardized tests don't measure educational quality. *Educational Leadership*, 56(6), 8– 15.

<sup>xi</sup> 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>.

<sup>xii</sup> 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.

<sup>xiii</sup> See Economic Policy Institute, Problems with the Use of Student Test Scores to Evaluate Teachers (Aug. 29, 2010)

<sup>xiv</sup> 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.

<sup>xv</sup> 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.”

<sup>xvi</sup> 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](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.”).

<sup>xvii</sup> 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](http://www.newhavenindependent.org/archives/upload/2011/09/NHPS_Talent_Update_110912.pdf).



## **Addition through Subtraction: Are Rising Test Scores in Connecticut School Districts Related to the Exclusion of Students with Disabilities?**

### **Executive Summary**

**January 2012**

According to a report by the National Center for Education Statistics, results on the Connecticut Mastery Test (CMT) showed more positive changes than on the National Assessment of Educational Progress (NAEP) between 2007 and 2009 in 4<sup>th</sup> grade reading and math.<sup>1</sup> The report noted that this discrepancy could be due to a change in state standards, focus in the curriculum or instruction, or the student populations taking either test. Indeed, the regulations for the testing of students with disabilities in Connecticut changed between 2007 and 2009. After 2009, the percentage of students at or above the proficient level on the CMT in math and reading appeared to increase more quickly than in past years. This paper examines whether the increase in the percentage of students meeting the proficient level or above on the CMT was related to the exclusion of students with disabilities.

This report describes changes in student participation and the percentage of students at or above the proficient level on the Connecticut Mastery Test. The description is based on a close analysis of declining participation rates of students in the state from the years 2005-2006 through 2010-2011. By examining the relationship between changes in participation and resulting percentage of students at or above the proficient level for the state in math and reading, we hope to build awareness and encourage discussion about the role and limits of CMT data to evaluate public education. Finally, we offer revised percentages of students at or above the proficient level for the state and each district in order to demonstrate the disparate impact from changing participation rates on the standard CMT.

This analysis focuses on the connection between declining participation rates and increasing percentages of students at or above the proficient level in the state for several reasons. In 2009, students with disabilities that districts determined would not have passed the regular CMT in math and/or reading were offered a modified assessment (MAS). The evidence shows that when these students were excluded from the standard CMT calculations, the percentage of students at or above the proficient level appeared to rise more rapidly in math and reading.

This is concerning for several reasons. As a matter of policy, No Child Left Behind and Race to the Top rely on various data from standardized tests to assess educational improvement from year to year. In practice, policymakers in Connecticut are using distorted standardized test data to assess schools in the state. This study raises a number of questions about the unintended consequences of the MAS and the uses of standardized test scores to evaluate the quality of public education. We also offer preliminary recommendations to policymakers.

### **Findings**

Our analysis of Connecticut Mastery Test participation and proficient-level data yields the following key findings:

- Using statistical analysis to identify any correlation between the percent of students participating on the standard CMT and the percent of students at or above the proficient level, we found evidence of a *strong association between the declining percentage of all students taking the standard Connecticut Mastery Test and the increasing percent of students at or above the proficient level.*<sup>11</sup> In other words, across school districts, as more students with disabilities

were excluded from score results, a greater percentage of students were reported as reaching the proficient level.

- When we include students with disabilities that took the modified assessment (MAS) in reading and math in 2009-2011 in the total sample of 4<sup>th</sup> and 8<sup>th</sup> grade test-takers in each year, then the revised percentages of students at or above the proficient level would be *two to three percentage points lower than the state reported*.<sup>iii</sup> Notably, even after revision, there do appear to be positive changes in percent at or above the proficient level from year to year, though these changes are not as large as they appear prior to revision.<sup>iv</sup> In other words, Connecticut experienced higher percentages of students at or above proficient in the 4<sup>th</sup> and 8<sup>th</sup> grade between 2008 and 2011, but not to the degree that the state initially reported.
- Statewide improvements in standard Connecticut Mastery Test (CMT) scores reported by the Connecticut State Department of Education (SDE) between 2008 and 2009 -- the period of the largest reported gains -- were largely the result of the exclusion of students with disabilities from these standard test results, rather than overall improvements. For example, 84% of the reported improvement in 4<sup>th</sup> grade math proficiency between 2008 and 2009 and 69% of the improvement in 8<sup>th</sup> grade reading proficiency could be attributed to the exclusion of these students. Much of the reported improvements in later years could also be attributed to this exclusion, though there were some modest overall gains as well.
- The percentage of students with disabilities who were assigned to take the MAS rather than the standard CMT varied substantially across school districts, from 0% to 12.8%. Most districts had some participation on the modified assessment that affected their test score data.

Additional findings include:

- Students with disabilities participating in the MAS in 4<sup>th</sup> grade reading and math made up 4.5% and 3.4% respectively of the state's total population (n=41,266) in 2011.<sup>v</sup> In 2011, only 52.5% of all 4<sup>th</sup> grade students with disabilities in the state took the standard CMT in reading.<sup>vi</sup> In contrast, 99% of all non-disabled 4<sup>th</sup> grade students took the standard CMT in reading.<sup>vii</sup>
- School districts with larger declines in participation because of placing students with disabilities on the modified assessment (MAS) were more likely to experience an increase in the percent at or above the proficient level on the standard CMT in math and reading, but not writing.

## Recommendations

As first steps, we recommend that the state clearly explain the impact of the modified assessment system and the exclusion of students with disabilities on the results of the standard CMT and carefully monitor the modified assessment system (MAS). Additionally we recommend that policymakers use a variety of indicators -- both quantitative and qualitative -- to assess the progress of public education, and reconsider and modify policies that assign "rewards" and "punishments" based on standardized test scores.

---

<sup>i</sup> Bandeira de Mello, V. (2011) "Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005-2009". (NCES 2011-458). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC; Government Printing Office: page 25-26. Web. 6 Oct. 2011.

<<http://nces.ed.gov/nationsreportcard/pdf/studies/2011458.pdf>>

<sup>ii</sup> We calculated Pearson's *r* to identify any association between percent at or above proficient and the percent participation on the standard CMT in math, reading, and writing between the years 2006 and 2011. We compared these statistics in the same grade over the five years to simulate the method used by the No Child Behind Act. Precisely, there was a strong inverse, or negative correlation, between percent at or above proficient and the percent participation during these years. This data set included several instances when participation went "up" and the percent at or above proficient went "down."

---

<sup>iii</sup> See Appendix E for a recalculated percentage of students at or above the proficient level on the standard CMT in math and reading from 2009 - 2011 in 4<sup>th</sup> and 8<sup>th</sup> grade for the state. Note that the recalculations for 2009 are based on an estimated number of students taking the modified assessment on the MAS pilot. The state did not report the 2009 pilot MAS data.

<sup>iv</sup> See Appendix H and I for year over year percentage of students at or above the proficient level in selected subgroups.

<sup>v</sup> eMetric. *Data Interaction for Connecticut Mastery Test, 4<sup>th</sup> Generation*. "2011 Connecticut Mastery Test Participation Report-Grade 4." CT Department of Education, 2011. Web. 1 August 2011.

<sup>vi</sup> eMetric. *Data Interaction for Connecticut Mastery Test, 4<sup>th</sup> Generation*. "2011 Connecticut Mastery Test-Grade 4-State-Subgroups by Special Education." CT Department of Education, 2011. Web. 1 August 2011.

<sup>vii</sup> eMetric. *Data Interaction for Connecticut Mastery Test, 4<sup>th</sup> Generation*. Subgroup Report, "2011 Connecticut Mastery Test-Grade 4-State-Subgroups by Special Education." CT Department of Education, 2011. Web. 1 August 2011.