

## **Chapter 21**

# **Assessment in the Real World: The Case of New York City**

ELIZABETH TALEPOROS

This chapter is devoted to a story from a city whose districts have faced and continue to grapple with huge challenges in helping teachers enable students from diverse backgrounds to achieve at high levels. My story is about one of the largest school systems in the United States and how it weathered storms of assessment and accountability. It begins in the hot political climate of the late 1960s.

New York City communities were clamoring for more control over their local schools. Turmoil over local control of schools in East Harlem and Ocean-Hill Brownsville sparked legislative action that resulted in the creation of thirty-two community school districts, each reflecting the unique local needs, interests, and cultural concerns of the neighborhoods that comprised them. The once all-powerful Central City Board of Education was replaced with local control, giving the districts the power to implement curriculum as it reflected local culture, and make decisions about how schools were to deliver education to neighborhood students.

The Central Board remained with a much narrower set of responsibilities, primarily focused on policy and accountability. It was responsible for setting standards, providing guidelines and support for instruction, and providing the public with test scores to reflect how successful schools were in meeting the standards. As part of the 1969 Decentralization Law, the Board was to hire a chancellor, whose responsibilities included monitoring achievement of schools in the newly created community school districts. The law required the publication of school rank by reading achievement.

Although prior city-wide testing existed, this requirement gave birth to the high-stakes city-wide testing program. Its primary mandate was reading achievement. The mathematics community advocated for city-wide assessment of math-

ematics also, and mathematics was added to the city-wide testing program. As it evolved, great controversy ensued about what grades and types of students were to be included, and the instructional implications of a common testing program without a common curriculum.

Later, the state testing program was implemented at grades 3, 6, 8, and high school. Its primary mandate was to assure minimal competency by the time students were ready to graduate. At high school, certain students also took a more challenging Regents examination, reflecting what has been characterized as a bifurcated system, with different academic curricula and assessments being applied to students in different tracks.

The city-wide testing program, mostly focused on elementary and middle school grades, used only norm-referenced tests. Each year, schools were ranked by the percentage of students scoring at or above grade level on these tests. Progress was tracked by this measure alone. Reading and mathematics achievement were thus monitored by the Central Board, and by real estate agents, public policy analysts, newspaper reporters, and all types of constituents from government and public advocacy groups.

The different approaches of the city and state programs led to great confusion, primarily about emphasis and curriculum. Teachers were responding to their local boards' requirements, and two sets of assessment mandates. One focused on achievement over a low bar of minimal competency; the other focused on achievement at or above a national median, where by definition half of the scores were below average. These differences perplexed teachers whose work was being assessed by different kinds of tests, with the success being evaluated with entirely different types of metrics.

In the 1990s, the city and the state began to come closer together, in the spirit of the then-burgeoning movement for school reform and high-stakes high-standards accountability. In the city, a group of people began to re-examine the city-wide program. Could we alter it, so that testing was a worthwhile experience for students and supported good instruction at the same time? We began to conceive of a mathematics performance assessment, where students were required to solve problems and explain how they got their answers, for the 75,000 students in the city. Certainly, we were concerned about a number of issues associated with scoring: getting reliable and valid scores, and providing professional development so that teachers could score the tests. Despite these concerns, we became convinced that this process was a way to turn instruction from the drill and kill approach to that of problem solving, reasoning, and communication.

We began by simply posing three problems to all seventh graders, asking them to solve the problems, describe their approach, and communicate how they

reached their conclusions. We developed a general 6-point rubric and classified scores as high, medium, or low proficiency. Each score point was described in the general rubric, with specific examples cited for the given problem. Every middle school mathematics teacher was trained in scoring in which each potential score point was illustrated by a few papers gathered during pilot testing a variety of ways students could achieve the score.

We made a policy decision to give the classification “low” to responses that gave the correct answer alone, with no explanation. This focus on different ways to solve problems, with communication about the method chosen and the steps taken, had an enormous impact on teachers. The aspects of performance that were valued changed. Students had to focus on their reasoning and communication skills.

Later, the program expanded to include fifth as well as seventh grade. We developed a method to scale the Performance Assessment in Mathematics (PAM) scores with the regular multiple-choice norm-referenced scores. To build on the design established in mathematics, the city developed a reading test for grades 4 and 6 to make the assessment program comprehensive and inclusive of depth of thinking dimensions as well as the traditional sampling from the breadth of the content domains.

The state program was then modeled in the same fashion for grades 4 and 8, eventually displacing the city program and expanding the topics tested. The reading test expanded to include greater emphasis on listening skills and writing. The mathematics test expanded to include more problems, with a significant increase in the challenge they provided to students. Now, with the No Child Left Behind requirements, the state program holds great promise for teaching and learning in New York. The evolution that produced this balanced and deeper approach to assessment was slow and incremental. But its impact on instruction has been profound, and it is hoped that teachers and students will continue to benefit in meeting the continuing challenges.

Our story does not end here. The structure of the educational system in New York has changed again. There is now a newly reorganized Department of Education, which, some say, has taken more centralized authority and responsibility. The political pressures remain, and have in some ways even increased. But fortunately, the city’s educational system is still struggling to balance the external political agenda with the work of a group of talented educators whose passion for standards-based curriculum, instructional approaches, and assessment is indeed to be recognized and applauded. The story of New York City is, and will undoubtedly remain, the continuing story of a work in progress.