

Assessment Matters:

*Constructing Model State Systems
to Replace Testing Overkill*



FairTest

National Center for Fair & Open Testing

Assessment Matters:

Constructing Model State Systems to Replace Testing Overkill

By Monty Neill
Executive Director, FairTest

A Report by the National Center for Fair & Open Testing

October 2016

Thank You

FairTest expresses our deep appreciation to people who offered their valuable time for interviews and discussions as we prepared this report. They have all contributed greatly to our understanding.

Avram Barlowe, Joe Battaglia, Richard Chang, Sarah Chang, Ann Cook, Shawna Coppola, Kristina Danahy, Kathy D'Andrea, Robin Coyne, Dan French, Ayla Gavins, Matthew Glanville, Paul Leather, Kate Lucas, Dennis Littky, Deborah Meier, Mission Hill School teachers, Rob Riordan, Rollinsford Grade School teachers, Lynne Stewart, Hanna Vaandering, and Elliot Washor.

FairTest also thanks our funders who supported this project: Bay and Paul Foundations, Open Society Foundations, New World Foundation, Schott Foundation, National Education Association and numerous state and local affiliates, and many generous individuals.

FairTest staff who contributed to this report: Lisa Guisbond, David Mirabella, Bob Schaeffer.

Cover photograph credit: New York Performance Standards Consortium students at work. Photo by Roy Reid.

FairTest

P.O. Box 300204

Boston, MA 02130

www.fairtest.org

fairtest@fairtest.org

617-477-9792

Assessment Matters:

Constructing Model State Systems to Replace Testing Overkill

Table of Contents

Part III: Performance Assessment Examples	4
Preface to Part III.....	4
Rollinsford Grade School.....	5
New York Performance Standards Consortium	8
Learning Record.....	11
Work Sampling System.....	14
Big Picture Learning.....	16
International Baccalaureate	18
Resources	20

Note: The report this section is excerpted from is available online at:
<http://www.fairtest.org/assessment-matters-constructing-model-state-system>

© FairTest 2016. Contents may be reproduced provided credit to FairTest is provided and no income is derived from the reproduction.

Assessment Matters:

Constructing Model State Systems to Replace Testing Overkill

Part III

Performance Assessment Examples

The U.S. has fine schools that make great use of performance assessment. They do so despite the damage wrought by No Child Left Behind with its insistence on dominating education with standardized tests.

Less common are systems or networks that exemplify high-quality assessment and demonstrate it can be done on a large scale. In some cases, these networks have established comparability of the meaning of achievement levels across schools.

Part III presents examples:

- Rollinsford Grade School exemplifies high-quality teaching, learning and assessing.
- The New York Performance Standards Consortium is a subsystem of 38 public high schools that uses performance-based assessments.
- The Learning Record is a portfolio-like tool that was used in a growing number of schools prior to NCLB.
- The Big Picture Company is a network of mostly public schools which prizes performance assessment, though its schools vary in their assessment practices.
- The Work Sampling System combines portfolios, checklists and short summaries of elementary students' progress across academic and other domains.
- The International Baccalaureate is a worldwide system of schools that uses a variety of forms of performance assessments.

We could have included others. At the school level, for example, there is Boston's wonderful Mission Hill School, which has regularly opened its doors to us. We decided the excellent videos about MHS would be a superior alternative to a short write-up, as is the case for other schools. Systems that extensively use performance assessments also flourish in other countries. Linda Darling-Hammond's research provides examples, as does FairTest's fact sheet on multiple measures. (For more details, see Resources at end of this Part.)

Rollinsford Grade School

“The kids blow my mind every day. The thinking they share with me or others, the kindness they show each other, and how they know to ask for and get help.”

– Principal Kate Lucas

This small-town New Hampshire public school has used performance assessments as part of a remarkable transition to an inquiry-based instructional approach, in which the learning process is as important if not more important than the product.

RGS has organized its work around four pillars, which guide students toward becoming:

- Collaborative and compassionate members of our global society;
- Lifelong learners;
- Architects of their personal wellness; and
- Critical thinkers and problem solvers.



Photograph by Rollinsford Grade School

Rather than rely on externally developed performance tasks, or even teacher-made tasks administered to the students as tests, it prioritizes teacher-guided, student-focused assessments that evolve out of the curriculum. That is, within the curriculum, students have substantial choice in identifying questions to explore. The resulting products, from books read and written about to science and social studies research, provide some evidence of student progress, difficulties

and attainments. Other evidence comes from ongoing observation of and conversations with students (also known as “kid watching”).

FairTest staff visited Rollinsford in May 2016, visiting classrooms and talking to staff and students. That day, grades 1-2 and 5-6 were sharing samples of student work with one another and grades 3-4. The following week, all grades shared in an open house for the town. Students chose the work to present, sometimes completed and sometimes in progress. Many presented the books they had read, focusing on one they had written about and sharing a list of others. While there was overlap, most students had read some books that few or no other students had.

One group of grade 5-6 students investigated Southeast Asian river dolphins. They researched biology and ecology, wrote a report, prepared visual presentations, and sold tie-dyed T-shirts to raise money to save these endangered animals. They happily talked about their work and findings. On the wall outside a grade 4-5 classroom were student questions for a social studies investigation that was just beginning into civil rights and civil liberties. One student had asked, how can we stop the KKK and protect civil liberties? These are the sorts of “questions, problems and project-based learning (QPP)” around which the school’s teaching and learning is organized.

Another girl’s showcase was about the importance of “first impressions” and “type A” personalities. She reflected on her first days as a fifth grader in a grade 5/6 multi-age classroom and how her “bold” contributions from the start left a bad taste in the mouths of her sixth-grade female classmates. She embarked on a close reflection on her personality and how she might change the impression she makes on others. Ms. Lucas (2016) said, “Her greatest realization was that because she had learned to allow others to contribute first and disagree by saying things like ‘I see that perspective and I’m curious how it might change if we think about it this way,’ she has made and kept friends.”

Rollinsford’s approach to instruction and assessment affects both academic content and students’ self-awareness. The purpose of the school is to help children grow into good adults, not to implement an assessment system or measure “competencies.” The qualities described in the epigram above – thoughtfulness, kindness and knowing how to ask for and get help – are what counts.

When students graduate from Rollinsford, they cross the Maine state line to attend the Marshfield middle and high schools. RGS students do quite well in this highly regarded but traditional school – though their good grades are, to RGS staff, secondary to their broader goals or “pillars.” Ms. Lucas noted that Marshfield teachers said RGS students “ask a lot of questions,” another sign of an actively engaged student.

NH has adopted a “competency” system in which students advance and graduate by demonstrating sufficient knowledge and skills in subject areas. Rollinsford staff designed their own competencies four years ago. These share some alignment with state competencies and other subject-area national standards. RGS reports student progress in light of the competencies on a trimester basis.

However, Rollinsford Grade School (RGS) has so far chosen to limit its participation in the NH PACE performance assessment pilot (see Part II). New Hampshire districts/schools that develop and implement the state performance tasks are Tier 1 schools. Rollinsford participates in a second group, Tier 2, which engages in discussion and planning, often in preparation to become a Tier 1 school.

Principal Kate Lucas was pleased they participated in Tier 2: “We found the conversations helped to solidify our philosophies and approach. They forced us to be certain in our position

and have substantial support/evidence of student success. We also very much liked learning with the other Districts.”

RGS has not joined Tier I for several reasons. First, teachers would have to take a large amount of time from school days to design the common tasks used across districts. Second, they conclude that administering performance tests is not the same as assessing student work that arises from the curriculum. They fear that focusing on such test tasks would undermine their own approaches to teaching and learning. Ms. Lucas said, “It would have been extremely difficult to embed the assessment within our school culture given the PACE performance task ‘constraints’ (rules, procedures, expectations, etc.)”

The downside is that they continue to administer the SBAC exam in grades 3-6, which is not compatible with RGS educational practice but can be administered without pulling teachers from their classrooms for task development. They largely ignore SBAC. RGS will continue to participate in Tier II and the conversations with other schools, and consider whether conditions will facilitate their participation in Tier I and the full PACE project. (See further discussion of this issue in Part II.)

Resources

Lukas, K. 2016. Personal communication, email, August 1.

Interested readers can explore Rollinsford’s website for discussion of their understanding of inquiry-based learning, what they mean by levels of competency in the subject areas, samples of student projects, and more. <https://sites.google.com/a/rollinsford.k12.nh.us/rollinsford-grade-school/>

FairTest’s Monty Neill interviewed Principal Kate Lucas, literacy specialist Shawna Coppola, and consultant Lynne Stewart (from the Center for Collaborative Education), and talked with staff and students who were engaged in a showcase of their work, on May 12, 2016.

New York Performance Standards Consortium

Performance-based assessment works well for all students, but its success with the most vulnerable students is what makes the outcomes of the New York Performance Standards Consortium so impressive. The Consortium now includes 38 public, non-charter high schools, 36 in New York City.

The Consortium’s assessments are created by teachers and rooted in in-depth, project-based curricula and teaching. Its 2015 report, *Education for the 21st Century*, demonstrates that Consortium schools significantly outperform those in other New York City public schools while serving a similar population. In particular, more students from all demographic groups graduate, go to college and stay in college.



NY Performance Standards Consortium biology students at work. Photo by Roy Reid.

The Assessments

To “demonstrate college and career readiness and to qualify for graduation,” all Consortium schools require students to complete four Performance-Based Assessment Tasks (PBATs): an analytic literature essay, a social studies research paper, a student-designed science experiment, and higher-level mathematics problems that have real-world applications. They include both written and oral components.

The Consortium has permission from the state Department of Education to administer only one of the state graduation tests, English Language Arts. The PBATs, generally completed in 11th and 12th grades, replace the Regents exams in other subjects and for school accountability.

Education for the 21st Century explains that the PBATs “emerge from class readings and discussion. In some classes, the tasks are crafted by the teacher and in other instances by the student.” For example, in social studies, each student must write and then orally defend a research-based analytic paper on questions that have grown out of a history, government, or economics class. The Consortium’s data report includes samples of the wide range of social studies interests addressed by the students, as well as similar

samples for the other required tasks. In the oral defense for each PBAT, the student responds to questions from a panel of teachers and outside experts.

As Urban Academy history teacher Avram Barlowe (2016) explains, the PBATs require students to learn perseverance, how to assess and apply evidence, and explain their thinking in these assessments in written and oral forms. They “demand that students learn, through practice, how to read, write, calculate, observe and research in a critical manner.” A DVD series, *Teacher to Teacher*, shows how teachers and students build their courses to attain these ends.

All the PBATs and oral defenses completed for the common graduation requirement are evaluated using Consortium-wide scoring guides (“rubrics”). The report includes rubrics for the four subjects. These well-developed assessment standards, written and revised as needed by Consortium teachers, allow accurate evaluations of student work across schools. Samples of the work are blindly re-scored to evaluate both reliability of scoring and the challenge level of teacher assignments. Samples of student work (“exemplars”) that have gone through a series of moderation studies help both scorers and students to think about high-quality work.

Each school maintains collections of work that chronicle a student's growth. The college persistence data show that the extensive reading, writing and long-term planning required for the performance assessments prepare students well for higher education.

The Results

Consortium schools follow the same admissions process as other non-exam New York City high schools. The student population of the Consortium’s New York City schools mirrors the city’s student body (only two schools are outside NYC). These schools have nearly identical shares of blacks, Latinos, English language learners and students with disabilities. Students enter Consortium high schools with lower ELA and math average scores than citywide averages.

The Consortium dropout rate is half that of NYC public schools. Graduation rates for all categories of students are higher than for the rest of NYC and nearly double the city’s rates for ELLs and students with disabilities. The Consortium also tracks student persistence in college. The report demonstrates that rates of enrollment in third semester exceed the national average. (The Consortium is updating this information, for release in fall 2016).

Resources and References

The Consortium website is at <http://performanceassessment.org>

NY Performance Standards Consortium. N.D. *Education for the 21st Century: Data Report on the New York Performance Standards Consortium*.

http://performanceassessment.org/articles/DataReport_NY_PSC.pdf

Teacher to Teacher, a series of videos and books on the Consortium.

<http://www.teacherscollegepress.com/teachertoteacher.html>

Barlowe, A. 2016. "The New York Performance Standards Consortium," workshop presentation at Save Our Schools Conference, July 9, Washington, DC.

See also the Webinar on Performance Assessment, sponsored by the Forum on Educational Accountability (FEA). It features Ann Cook of the Consortium, Sally Thomas of the Learning Record, and Monty Neill from FairTest. <http://fairtest.org/view-new-webinar-authentic-performance-assessment> . The webinar is a good place to start; you can obtain just the slides as well, no voice (though listening and watching provides more depth).

The Learning Record

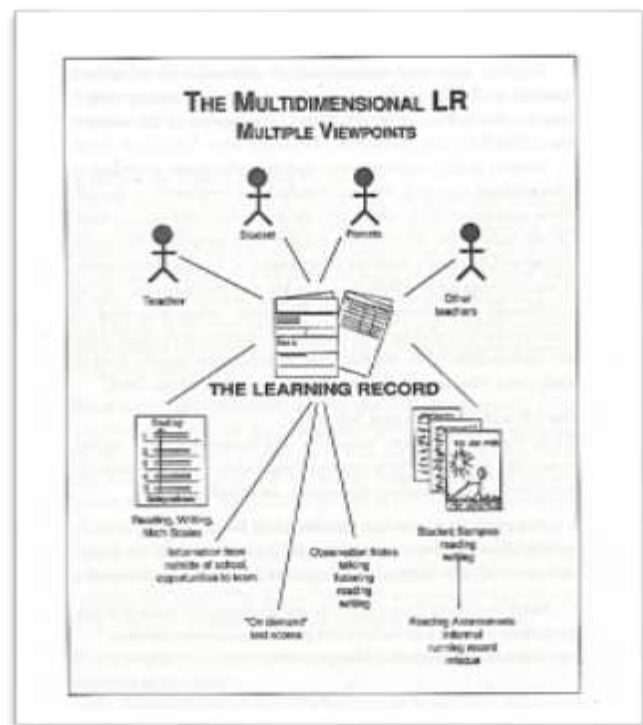
The *Learning Record* (LR) is a system of literacy and mathematics assessment, K-12, maintained and monitored by classroom teachers to document student progress toward agreed upon goals and standards in reading/language arts and math. Parents and students also contribute evidence to the Record. It is a superb means for gathering information to assist individual students and teachers. Its re-scoring (“moderation”) procedures demonstrate high inter-rater reliability between classroom teachers and independent re-scorers. Thus, it could be used for public reporting.

The LR was developed in England as the *Primary Language Record* for use with multi-lingual, multi-cultural young children in reading, writing, speaking and listening. Its structure provides a consistent framework for teachers to gather and evaluate evidence of student learning. It was adapted in the U.S. and expanded to include higher grades and mathematics. It grew to 25 U.S. schools in 2001, including many Bureau of Indian Affairs schools, before being largely swept aside by NCLB testing requirements.

In the LR, each teacher documents and evaluates each student’s progress over the course of the year. It begins with the teacher summarizing two conversations, with the student and the parent(s), each focusing on the child as a learner. Such information has been documented as improving teachers’ understanding, reducing referrals to special education, and strengthening parent relationships to the school. (Darling-Hammond, Anness & Falk, 1995, Ch. 6).

Each Record includes “observation notes” by the teacher of the student as a speaker and listener, reader, writer, and mathematician. It includes writing samples and reports on at least three books the student read and her understanding of them. Math covers four core domains. For each piece of documentation, the teacher describes the context of the observation or sample (e.g., individual or group work; whether the reading was literature or social studies, whether the book was new or familiar).

Teachers write extensive summary descriptions of the student’s status and progress three-quarters of the way through the year. These are reviewed and discussed with parents. Toward



the end of the year, the record is prepared for the coming year's teacher, adding any significant new information.

In reading and writing, each student's progress is placed numerically on a developmental scale. Separate reading scales cover grades K-3, 4-8 and 9-12. The scales are similar to many others that describe the process of learning to read and write. In K-3, teachers observe and document such things as students' use of letter-sound correspondence. They consider how children apply their prior life experience and how they build on their familiarity with the conventions of print as ways to understand what they read. Teachers use tools such as Running Records and Miscue Inventories as well as informal observations to determine which skills and strategies students are using and which ones they need to learn. In Grades 4-8, teachers look for signs that students are learning to read widely, for recreation as well as literature and informational text in all subject areas. In Grades 9-12, students provide evidence they can read critically across the curriculum.

Use of the LR scales have been validated (Hallam, 2001; Thomas, 2001). Moderation processes established good inter-rater agreement (above .75) between the classroom teachers and the independent re-scoring. This shows that with a good structure, diverse sources of information can be organized and evaluated to provide accurate, comparable evidence of student progress. Results can be aggregated and used to describe group attainment and progress. That is, if each originating teacher's judgment is solid, as supported by a review of 3 to 5 randomly sampled Records, then the aggregate information about classrooms and schools can be considered sound.

Hallam's (2001) validity study also compared LR scale scores with standardized test results. While the LR and norm-referenced tests such as the CTBS and Stanford Achievement Test are fundamentally different, they measure some of the same skills. Thus, the expectation is for positive but not very high correlations, which is what Hallam found in her investigation of 3009 individual LRs. Correlations ranged from .48 to .65. Hallam also found strong support among teachers for the LR's positive effects on professional development.

Practice and research in reading and writing demonstrate the LR is a reliable, valid, comparable and educationally sound method of evaluating individual progress and status using multiple sources of evidence, and aggregating that information to provide public information about schools. Math was a later development for the LR and thus the scales developed for the subject were not evaluated in depth. As with the LR in reading and writing, teachers provide varied specific evidence of learning that could be scored on LR with common rubrics, to provide reliable and valid data.

Resources and References

For more detailed information about the LR, see <http://www.fairtest.org/learning-record>. It includes links to articles, reports, sample recording forms, reading and writing scales, professional development, etc. For example, a sample recording form for young children is at [http://www.fairtest.org/sites/default/files/LR-%20reporting%20form%20-%20Elementary_Eng%20\(1\).pdf](http://www.fairtest.org/sites/default/files/LR-%20reporting%20form%20-%20Elementary_Eng%20(1).pdf). This page also links to further discussions of the validity and reliability of the LR.

FairTest thanks Mary Barr, Myra Barrs and Sally Thomas for their contributions to our knowledge, shared over many years.

Articles and Books

Barr, M., Craig, D., Syverson, M, and Fiset, D. 1999. *Assessing Literacy with the Learning Record: A Handbook for Teachers, Grades K-6*. Portsmouth, NH: Heinemann.

Barr, M., and Syverson, M. 1999. *Assessing Literacy with the Learning Record: A Handbook for Teachers, Grades 6-12*. Portsmouth, NH: Heinemann.

Darling-Hammond, L., Ancess, J., and Falk, B. 1995. *Authentic Assessment in Action*. New York: Teachers College Press. See esp. Ch. 5, “The Primary Language Record at P.S. 261.

Hallam, P.J. 2001. “Findings on Literacy Learning, Environment and Learning Record™ Validity, 2001 at Combined Learning Record™ Sites.” Center for Language in Learning. <http://www.fairtest.org/sites/default/files/LR%20validity%2099-01all%20scores.pdf>.

Thomas, S. 2001. “Learning Record Shows Promise For Accountability Uses.” *FairTest Examiner*, Fall. <http://www.fairtest.org/learning-record-shows-promise-accountability-uses>

Work Sampling System

The Work Sampling System (WSS) was developed by Samuel Meisels, one of the nation's foremost authorities on the assessment of young children, and his colleagues. Teachers observe children, collect and evaluate samples of student work in structured portfolios, use checklists, and prepare summary reports (shared with parents) three times per year. Originally for pre-school through grade 3, it now goes extends through grade 6. WSS has been demonstrated to have strong validity and reliability as well as parental acceptance.

The system is based on seven domains or categories, each with performance indicators:

- Personal and Social Development focuses on self-identity, the self as a learner, and social development.
- Language and Literacy is based on the theory that students learn to read and write the way they learn to speak, naturally and slowly.
- Mathematical Thinking focuses on children's approaches to mathematical thinking and problem solving.
- Scientific Thinking emphasizes the processes of scientific investigation, because process skills are embedded in and fundamental to all science instruction and content.
- Social Studies includes understanding from personal experience and learning about the experiences of others.
- Arts focus on how using and appreciating the arts enables children to demonstrate what they know and to expand their thinking.
- Physical Development includes developing fine and gross motor skills and competence to understand and manage personal health and safety.

Pearson now owns the WSS. Some users charge that the "grade-specific guidelines" for student achievement now based on the Common Core State Standards are developmentally inappropriate. However, they add, the tools for observing, describing and summarizing remain useful. Because checklists produce numbers, some observers fear that the numbers would be too seductive. That could undermine WSS' observational and narrative uses as well as parent interpretations. Investigations, including by FairTest staff, found that experienced teachers wrote highly individualized summaries of each child. Each seemed to be a clear snapshot, not a generic summary of attainments. WSS thus seems to effectively combine personalization and systematization, though the question of parental interpretation remains open.

In studies conducted when WSS was owned by Dr. Meisels, researchers found strong evidence of validity, reliability and parental approval (FairTest *Examiner*). One compared WSS reports by 17 teachers with their students' results on the individually administered Woodcock-Johnson (WJ) battery for literacy and math. They found correlations between WSS and the WJ comprehensive scores in reading, writing and math to range mostly from .50 and .75. These are high enough to support a claim of teacher assessment accuracy and not so high as to suggest that the WSS only measures what the WJ measures. The WSS also was found to be an accurate means of identifying children in need of special services, based on correlations with the WJ.

A second study found that parents in Pittsburgh viewed WSS positively. They believed its use benefited their children whether they were high or low achievers. Most of the families were low-income African Americans. The more parents knew about WSS, the greater their satisfaction. The researchers concluded, “[T]his study demonstrates that when schools using a systematic, curriculum-embedded performance assessment make an effort to keep parents informed about the assessment, and when consistent informal communications between parents and teachers takes place, parental reactions to performance assessment can be very positive.”

Another study compared children who had been enrolled in WSS with a demographically matched sample of children who had not experienced WSS and with all other children in the district. The WSS students showed substantially higher gains over time on conventional, group-administered achievement tests.

If the WSS is accurate and parents respect it, could its results, such as teacher-generated summary reports, be aggregated for use in reporting student learning at the school level? The WSS was not designed for this purpose. Prof. Meisels expressed concern about the potentially corrupting effects on teacher judgments if they know the results could be used to evaluate schools or make high-stakes decisions about students.

The accuracy of this curriculum-embedded assessment adds to the evidence that it is possible to construct public reporting systems based substantially on teacher judgment. If “accountability” consequences are genuinely helpful rather than punitive, the legitimate concerns about corrupting effects could be resolved.

Resources and References

The WSS is now owned by Pearson; information on it is available at <http://www.pearsonclinical.com/childhood/products/100000755/the-work-sampling-system-5th-edition.html>.

FairTest *Examiner* articles on WSS are available at <http://www.fairtest.org/work-sampling-system> and <http://www.fairtest.org/trusting-teacher-judgment>.

Big Picture Learning

Big Picture Learning (BPL) is a network of 65 schools across the U.S., with as many in a handful other countries. While a few of BPL's U.S. members are charters, most are regular public schools, and most are high schools. What distinguishes them are factors such as:

- grouping students in “advisories” of no more than 15, which provides a sort of home group for students throughout their school years;
- engaging in out-of-school internships. High school students are often at internships two days a week, learning with a mentor from the local community;
- project-based learning that bridges the real world of internships with academic learning that mixes teacher- and student-led studies, which enable pupils to build on their interests; and
- performance assessments, including student presentations and discussions of their work (“mini-portfolios”) several times per year.



Big Picture Learning students with advisor. Photo from Big Picture Learning.

The performance assessments are not standardized across BPL, nor do they use a common scoring guide. Their students are subject to federal and state testing requirements, which have in some cases undermined portfolio and performance practices.

One school reporting harmful consequences is the founding BPL school, Met High School in Providence, R.I. Met leaders told FairTest that preparing for

state exams, particularly the math test, took time from further developing common assessment practices in the school. Fortunately, Rhode Island has dropped its pending high-stakes high school exit exam, which should alleviate pressure. The Met seeks to return to developing its own assessments.

Met students produce extended biographies, often 75-100 pages, and senior thesis projects, as well as complete internships and coursework. Many have earned college credits, as is true across BPL schools. Met plans a 2016-17 pilot in which students, in collaboration with teachers, post their portfolios to a website, enabling more shared study of student products. The school has its own rubrics, crafted by the teachers, for evaluating projects and internships. Met curriculum director Joe Battaglia said they are used first for planning and guidance and second

for assessing student work. Met and other BPL schools are also grappling with how to document, assess and give credit for the large amounts of learning beyond what is initially planned when internships and courses are designed.

Met forms its own district, drawing students from across the state, with about 70% qualifying for free- and reduced-price lunch. Thirty percent of their students enter with lower than grade 4 reading levels. A detailed investigation of RI high schools found the Met at the top in 80 out of 81 categories, reported Dennis Littky, founding co-director of the Met and BPL.

A study of BPL graduates from across the network reported strong results. A 2015 report concludes, “Findings show that the Big Picture Learning model is highly effective at fostering positive relationships, helping students discover and pursue their interests, and raising high school graduation and college entrance rates” (Arnold, *et al.*, 2015). The schools averaged a 92% graduation rate, of whom over 90% started at college or other post-secondary institutions. Of those, from the classes of 2006 and 2007, more than two thirds had earned a Bachelor’s (35% and 24%), Associate’s or other credential, or were still enrolled in 2012. This compares well with the national average, particularly of students from lower-income strata. (For example, the four-year college graduation rate of the lowest income quintile is 10.6%.)

Most of BPL students “come from communities with high levels of academic under-achievement, geographic transition, and high school dropouts.” Much of BPL’s success is due to the deep relationships built in the schools, especially between advisors and students, and the resulting strong sense of school community that sustains the youth. Most remain in some contact with advisors after graduating. Parental engagement in BPL schools also is far higher than in most U.S. high schools. However, the report concluded that many students’ preparation for college work in math and science was weak, an issue BPL is grappling with.

Resources and References

Arnold, K.D., Soto, E.B., Wartman, K.L., Methven, L., and Brown, P.G. 2015. *Post-secondary Outcomes of Innovative High Schools: The Big Picture Longitudinal Study*. Boston College, March 3. Available at

<https://d3jc3ahdjad7x7.cloudfront.net/9hloszW4FyNM5EdJWri39BVKbVpArurU9gAFe3FmKmcuICyK.pdf>

Big Picture Learning website, <http://www.bigpicture.org/>

FairTest visited Big Picture Learning on May 10, 2015, talking with co-founder Dennis Littky, Joe Battaglia and other staff. FairTest also talked with BPL co-founder Elliot Washor.

International Baccalaureate

“The International Baccalaureate® aims to develop inquiring, knowledgeable and caring young people who help to create a better and more peaceful world through intercultural understanding and respect.”

Founded in 1968, International Baccalaureate (IB) is a worldwide system of thousands of schools known for strong academics and in-depth learning to prepare students for college. IB contains four programs: Primary Years, Middle Years (ages 11-16), Diploma (final two years of high school), and Career-related, which includes some Diploma courses and additional work aimed at career preparation. IB periodically reviews its member schools to re-approve them. Slightly more than half of IB Diploma candidates are from the U.S. (76,000).

IB assessments vary significantly by program level. In the Primary program, there is formative but no external, summative assessment. All students complete a personal project at the end of the program.

The Middle program assessments are primarily local. Students complete a project of personal interest to them in each of the final two Middle years. Typically, these are presented in a “showcase.” Until 2015-16, IB provided voluntary moderation (re-scoring) for samples of student work submitted by participating schools. That year, IB initiated an optional exam on computer. Schools now choose between the portfolio and the test. Only a few have signed up for the test at this point, including some who see it as necessary to meet government requirements.

The Diploma program requires students to pass courses in six areas: literature, foreign language, humanities (which includes history), science, math and a student option, such as art. These are offered at “Standard” and “Higher” levels, with students required to complete higher level work in at least three areas.

Students take a combination of internal and external exams in each subject. For art, IB uses portfolios. The external tests, taken on paper and marked on a 1-7 scale, carry more weight, typically around 80%. They are centrally scored by international teams of teachers and retired teachers. This requires IB to address language and cultural differences among students and graders. Matthew Glanville, head of assessment, says it is also difficult to score consistently across subject areas, but they are generally satisfied with their accuracy.

In 2015, IB registered 580,000 students to take exams and granted about 56,000 diplomas. The exams are primarily responses to essay prompts, but include short and longer structured problems. Multiple-choice items are rarely used because IB expects students to demonstrate deeper knowledge and higher order thinking.

Each Diploma student also must pass a Theory of Knowledge course which includes an extended, research-based essay the student chooses with teacher support. The essays are

graded by IB rather than the school. Glanville says that students often complain about this project, then report it was their most valuable preparation for college.

IB staff emphasize the importance of teacher assessing. It provides guides for teachers to use in developing and marking tests, projects and other classwork.

Boston's public Josiah Quincy Upper School offers the Middle and Diploma IB programs, covering grades 6-10 and 11-12. The student body is 92% low income, 50% Asian, 30% Black, 15% Hispanic and 5% White; its location on the edge of Chinatown explains in part the disproportionate Asian participation. Twenty-two percent have disabilities and most of them are fully included in the IB program. Not all students who graduate complete the IB program, but two-thirds of the Upper students are on track to do so. Quincy joined the Middle program exam, but only for grade 10, the final Middle year. The school also participates in voluntary moderation activities in grades 6-10.

Josiah Quincy staff pointed out that the Diploma program focuses heavily on preparing for the exams, while the Middle program is more holistic. Even with the exam focus, an internal survey found students prioritize learning for its own sake over grades, and they prize self-direction and taking ownership of their learning. Staff also said that moving toward a portfolio approach might be a good idea.

In conclusion, IB assessment, particularly in the Diploma program, provides limited student choice, though students say they greatly value those options. IB shows that a wide range of exams can be fairly scored. Their exams can do a far better job of evaluating in-depth knowledge than current tests common in the U.S. In a "system of systems" that states could establish under ESSA, IB or similar approaches are a valid option. The question states will face is how to establish comparability between IB results, those of current state tests, and other locally determined assessments.

Resources and References

International Baccalaureate website, <http://www.ibo.org/>.

Interview with Matthew Glanville, Head of Assessment Principles and Practice, June 14, 2016.

Visit to Josiah Quincy Upper School, Boston, MA, June 23, 2016. Discussions with Richard Chang, co-principal; Robin Coyne, Guidance Counselor; and Sarah Chang and Kristina Danahy, IB co-Coordiators.

Additional Resources on Performance Assessments

These are in addition to references included in each section.

FairTest fact sheets:

A Better Way to Evaluate Schools. 2010. <http://fairtest.org/fact-sheet-better-way-evaluate-schools-pdf>

Multiple Measures: A Definition and Examples from the U.S. and Other Nations. <http://www.fairtest.org/fact-sheet-multiple-measures-definition-and-exampl>.

Books and Reports:

Annotated Bibliography: Performance Assessment. N.D. FairTest.

<http://www.fairtest.org/annotated-bibliography-performance-assessment>. (These books and reports are primarily from the 1990s; many remain very useful.)

Darling-Hammond, L. 2010. *The Flat World and Education*. New York: Teachers College Press.

Meier, D. 2002. *The Power of Their Ideas*. Boston: Beacon Press.

Wood, G., Darling-Hammond, L., Neill, M., Roschewski, P. 2007, May. Refocusing Accountability: Using Local Performance Assessments to Enhance Teaching and Learning for Higher Order Skills. <http://www.fairtest.org/refocusing-accountability-using-local-performance->

Videos:

Beyond Measure. <http://beyondmeasurefilm.com/>. By Vicki Abeles, with a companion book.

FairTest: You Can't Judge Learning with a Standardized Test. Four-minute video available at <https://www.youtube.com/watch?v=WkJ1st6vDyY>.

Good Morning Mission Hill. <http://goodmorningmissionhill.com>. Tom and Amy Valens. See also a series of shorts based on the film's footage, *A Year at Mission Hill*.

<http://www.ayearatmissionhill.com>. Video #9, "Seeing the Learning," focuses on assessment.

Most Likely to Succeed. <http://www.mltsfilm.org/>. Produced by Ted Dintersmith, Greg Whitely and Adam Leibowitz.

Schools That Change Communities.

<http://www.docmakeronline.com/schoolsthatchangecommunities.html>. Bob Gliner.