

Teacher Work Group on Assessment Recommendations, Spring 2013

***Principal Writers:* Gerardine Carroll, The Center School, and Liza Campbell, Nova**

***Contributors:* Peter Henry, Seattle Public Schools; Chelsea Palmer, The Center School; Michaela Peterson, The Center School; Mallory Clarke, Garfield High School; Adam Gish, Garfield High School**

Introduction--Context

In 2008, the Seattle Public Schools contracted with Northwest Evaluation Association (NWEA) to administer the Measurement of Academic Progress (MAP) to all students in grades K-9 in order to inform teachers about student growth over a year's period and to adjust their teaching practices accordingly.

In Fall 2012, the staff at Garfield High School refused to administer the MAP exam. They cited a variety of concerns including:

- Superintendent Maria Goodloe-Johnson agreed to the purchase of the test when she served on the board of NWEA, an act criticized by the states auditor's office
- Seattle Public Schools (SPS) wastes precious instructional time as well as teacher, administrator and school resources administering MAP, including hours of computer access that should be used by students for research and other academic purposes
- SPS uses MAP as a part of teacher evaluations and as a means to determine student eligibility for advanced learning, practices that NWEA itself has cautioned against because the test is not designed for these purposes
- MAP's margin of error for high schoolers exceeds the expected growth for students, thus making the test invalid
- The district's demands for MAP administration—two tests two to three times yearly—deprives students of instruction time, the impact of which is felt most harshly by those students who struggle the most
- MAP is not designed to meet the needs of students with Individual Education Plans (IEPs)

Garfield's refusal to administer MAP has spread to other schools throughout Seattle, and the resistance to the testing movement has rippled across the nation and even to other countries. Why? Teachers recoil at the false notions that standardized tests are legitimate single measures of student academic achievement, that standardized tests should take precedence over instructional time, that standardized tests effectively assess teacher quality—in short, that standardized testing-based education is effective education. It is not, and we as teachers stand firm in our refusal to embrace ineffective and harmful practices. While the district has admitted, in the face of great pressure, that the MAP test does have flaws and that it will make changes to how MAP is administered, these changes still ignore some of the most fundamental flaws in the test. The changes seem to be based on making sure MAP is still given instead of making sure that it provides useful information about our students. This spring:

- Those in Algebra class will no longer take the MAP's general math test; instead they will take the algebra MAP. While this seems to make some sense when first viewed, it ignores the state's End of Course exam, whose administration window coincides in part with MAP's window. Two tests will be used to measure student progress in the same subject, using instructional time as assessment time instead. If the goal in testing is to measure student knowledge, one standardized test is more than enough, and only using such multiple-choice based tests with little depth of knowledge in their questions is certainly not sufficient.
- In reading, only those 9th graders who did not meet standard on the 8th grade state test (MSP, or Measurement of Student Progress) will be required to take the MAP reading. Logic tells us that students who have performed poorly on one reading test will likely perform poorly on another. Many of these students already receive support through English Language Learning and Individual Educational Plan classes as well as through programs such as Read Right and Read 180. Even if the Spring MAP results are accurate, do they yield information that is worth the teacher time, student effort, and school resources needed to help students already identified as struggling?

District Task Force on Assessment

In January, in response to the MAP boycott, Superintendent Jose Banda announced the formation of a task force whose purpose was to be “a productive way for educators and our community to come together to discuss concerns and find solutions that best meet the needs of all our students . . . The new task force will have the opportunity to explore and review the strengths and limitations of the MAP assessment, and will consider potential alternatives to future district testing programs” (Banda).

While we applaud Superintendent Banda's willingness to examine the issues surrounding MAP, we question the effectiveness of the task force-, despite the good intentions of its individual members, for the following reasons:

- Out of thirty members, five classroom teachers serve on the committee. While we realize that the voices of many stakeholders must be heard in order to synthesize different viewpoints, the severe limiting of teachers' perspectives ignores the reality of testing logistics to students' school day, of the uses of data in the classroom, and of the impact of results on instructional practices.
- On a related note, those individuals included in the task force were in no way democratically chosen, nor is there a sense that the intent to represent a cross-section of the city's teachers was considered.
- Part of the second meeting was spent reviewing “how MAP is currently used in schools,” including reports by Chris Carter, Principal, Mercer Middle School, and Jeff Clark, Principal, Denny International Middle School (*Classified*). While we respect the dedication schools have in using data to inform practice, we must point out that the data used must be meaningful and reliable. If the data are not, then the resulting changes to instruction are not either. Reports and reviews such as those given at the first meeting assume the MAP data are, in fact, true indicators of student progress towards state and district standards. Research on NWEA's own website proves this assumption incorrect: an individual student's

expected growth on MAP is less than the margin of error. When the test developer's studies question the reliability of its own data, we should not use the test to make decisions about instructional practices, student achievement, and teacher evaluation.

- Additionally, there are many administrators across the city who *do not use* MAP data to inform practice. Their voices, however, were not included in the conversation on “how MAP is currently used in schools.”
- Those of us who are concerned about the use of the MAP here in Seattle, and the overuse of standardized tests here in general, fear that the district's task force may be intended as a rubber stamp for those policy recommendations that the district was planning to make from the beginning. This fear is strengthened by the meeting notes, which indicate a lack of opportunity for open discussion. No specific discussion of MAP's shortcomings took place.

For all of these reasons, and because we felt the voices of concerned teachers would not be clearly heard through the district's task force, a group of teachers decided to hold sessions, open to all certificated teachers across the district, to discuss the use of assessments in Seattle. **The Teacher Work Group on Assessment** consisted of over twenty teachers, some connected to the Scrap-the-MAP Committee and some not, interested in discussing how our schools use assessments. Our goals were simple but ambitious:

- Organize in grade-level bands
- Tie assessments to classroom work and to standards
- Base assessments on best practices
- Make recommendations to the district

Research on Standardized Testing

Our first step was to delve into research on the use of standardized and high-stakes testing. A quick overview of this research shows that standardized testing:

- Narrows curriculum both within a subject and across the entire scholastic curriculum by de-emphasizing untested subjects
- Decreases rigor by emphasizing memory recall and test-taking skills over critical and creative thinking
- Exacerbates inequities for students of color/poverty
- Is often used for the purpose of implementing policies such as holding back elementary students and tracking students, which are shown to be detrimental
- Negatively affects students' self-perception as competent learners
- Narrows debate on what's considered important in education; ignores larger issues such as poverty, class size, funding equity (Williams)

What Works? A Very Brief Look at Finland and Singapore

Clearly educators must acknowledge the complexity of education. Unfortunately, one reality confronts us all: anything that is complex cannot be measured with simple assessments.

Internationally, others have acknowledged this simple truth. Finland’s school system is all the rage in educational circles, and this past fall Seattle hosted Pasi Sahlberg, a director general at the Finnish Ministry of Education and Culture, to speak about how a country that 40 years ago clutched the bottom rungs of educational achievement now stands at the top of the school ladder internationally. Sahlberg questions the proliferation of high stakes testing in other countries, including the U.S., saying Finland’s high performance on the Program on International Student Assessment (PISA) and its lack of standardized tests through its system “suggests that frequent student standardized testing is not a necessary condition for improving the quality of education as has been insisted upon by many advocates of competition-based public sector policies” (Sahlberg 66)

Singapore’s former Education Minister, Tharman Shanmugaratnam, has embraced quality and complex education as well, noting that the information age in which we live demands a move away from “rote learning, repetitive tests and a ‘one size fits all’ type of instruction” and towards “engaged learning, discovery through experiences, differentiated teaching, the learning of life-long skills, and the building of character, so that students can... develop the attributes, mindsets, character and values for future success” (Darling-Hammond). This kind of education can never be measured by a twice-yearly, computer-administered test.

Teacher Work Group on Assessment—Our Recommendations

Using research from a variety of sources and our own expertise as professional educators, the Teacher Work Group on Assessment compiled a list of Markers of Quality Assessments. These markers have guided our work, and we feel that all assessments—those developed or used by teachers in the classroom as well as those developed or adopted for district, state, or national use—should be evaluated with them. We grouped our criteria into three categories related to process, validity, and creation/review. Our research-based criteria are listed below.

- I. Assessments should incorporate a variety of measures, possibly gathered into a body of evidence that demonstrates abilities. These measures, taken together, should
 - ✓ Include classroom work
 - ✓ Allow teacher and student choice
 - ✓ Integrate with curriculum
 - ✓ Demonstrate student growth as well as standards achievement
 - ✓ Be free of gender, class, and racial bias

- II. Valid assessments
 - ✓ Reflect actual knowledge and learning, not test taking skills
 - ✓ Are educational in and of themselves
 - ✓ Are differentiated to meet students’ needs
 - ✓ Allow opportunity to go back and improve
 - ✓ Have tasks that reflect real world thinking and abilities

- III. The creation and review of assessments should
- ✓ Include community input
 - ✓ Undergo regular evaluation and revision by educators
 - ✓ Be graded by teachers collaboratively

When we looked at the current assessments used district-wide by the Seattle Public Schools and critiqued them based off of our list of criteria, we found most to be very inadequate. None of these assessments met all of the criteria, and few even met the majority. Those that may meet several seem to be:

- Columbia TC Quick Reading Assessment, used only for early grades to assess reading levels.
- The OSPI-developed assessments (formerly Classroom Based Assessments for Social Studies and Classroom Based Performance Assessments for the Arts, Fitness, and Health)
- International Baccalaureate (IB) assessments, which a small percentage of high school students across the district participate in
- The Direct Reading Assessment (DRA)

There are other assessments that possibly meet the criteria but are used sporadically by individual teachers in the district, by the district in the past but currently discontinued, or under special circumstances. They include:

- Direct Writing Assessment (DWA)
- The OSPI Collection of Evidence (COE) tasks, offered as an alternative only for those students who have not been successful on state-mandated graduation exams such as the math EOCs and the HSPE.

This list comes with caveats. First, our use of the words *may* and *seem* in the introductory sentence is purposeful. Our committee does not have the research capabilities to complete a full analysis of these; in addition, no single assessment has been used by enough teachers on the committee to fully evaluate it and its implications. Second, no *single* assessment can be sufficient to measure our students' skills or growth. After all, our first criterion clearly states, "Assessments should incorporate a variety of measures, possibly gathered into a body of evidence that demonstrates abilities." In fact, with the boycott of the MAP came the frequent question, "What will you use to replace it?" An underlying assumption of this question is that one test can be replaced with another. It cannot. Individual assessments on the list above may serve as one of the measures, but we must move away from viewing any test or activity as being The One.

Thus one of the Teacher Work Group's major recommendations is that any assessment system adopted by the Seattle Public Schools includes classroom-embedded tasks that are integrated into the curriculum. We brainstormed a list within each subject area of what such tasks might look like. This list is neither perfect nor exhaustive. Instead we present it to show examples of the way in which teachers can be included in the process of developing assessments that are embedded into a rich classroom environment. In this way, assessments contribute positively to the educational environment rather than detract from it.

Science

- Model a complex ecosystem
 - Trace a particular element (nitrogen, carbon) through biotic and abiotic factors in the system
 - Analyze the population dynamics within the system
 - Discuss possible consequences to changes in populations or changes in abiotic factors in the system
 - Model possible consequences of the introduction of a new species to the system
 - Discuss ways in which the model is limited as a representation of a larger system

- Design an experiment using organisms
 - Identify a question or problem, and engineer a testable hypothesis
 - Design an experiment with appropriate variables to test this hypothesis
 - Execute the experiment, gather data, and write an analysis and conclusion based on results
 - Discuss bioethical principles regarding using live organisms and appropriate testing techniques
 - Apply the information gained to a new problem or question arising from the experiment

- World Health Project
 - Identify world health topic, clarifying relative scientific information (immunology, parasitology, genetics, etc.)
 - Access scientific communities to build understanding of underlying issues with the health issue
 - Explore socio-political issues surrounding world health topic (economics, culture, geography, history, etc.)
 - Engineer possible solutions to health issue with a focus on problem solving skills
 - Explore bioethical practices in terms of health issue and possible solutions
 - Assess the scope and limitations of these engineered solutions

Math

- Use understanding of various aspects of a circle including radius, diameter, circumference, area, etc, to model the spread of disease, movement of hurricane, sound from a concert, etc.
- Use trigonometric ratios to model real-world navigational dilemmas, solve problems
- Use algebraic understanding to represent various growth/decay models and apply them to various changing factors
- Develop various proposals for ideal next steps after examining novel financial circumstances
- Explore and come up with own questions about distance, rate, time, etc.

Language Arts

- Annotate texts to demonstrate comprehension and analyze author's style choices
- Write literary essay on character, motif, theme, or style
- Defend interpretation of text orally
- Create short story, poem, screenplay, or other artistic writing
- Stage interpretations of text, using tableau, skits, or video
- Review a dramatic performance, movie, or other performance

Social Studies

- Write analysis of historical event through social studies lenses
- Debate current issue with clear reference to text
- Evaluate through public exhibition (e.g., letter to editor, testimony to governmental body, public service announcement) current legislative issue
- Evaluate a theoretical geographic area for its economic uses and their political ramifications

All of the kinds of assessments above fall under the general category of “performance tasks.” In addition, they meet many of the Teacher Work Group’s criteria of quality assessment, such as:

- ✓ Include classroom work
- ✓ Allow teacher and student choice
- ✓ Integrate with curriculum
- ✓ Reflect actual knowledge and learning, not test taking skills
- ✓ Are educational in and of themselves
- ✓ Have tasks that reflect real world thinking and abilities

In addition, with rigorous construction and evaluation of the assessments themselves, they could easily meet the rest:

- ✓ Demonstrate student growth as well as standards achievement
- ✓ Be free of gender, class, and racial bias
- ✓ Are differentiated to meet students’ needs
- ✓ Allow opportunity to go back and improve
- ✓ Include community input
- ✓ Undergo regular evaluation and revision by educators
- ✓ Be graded by teachers collaboratively

Linda Darling-Hammond, a leading researcher on assessment who was tapped to address the SPS Task Force on Assessment, supports these kinds of performance tasks for many of the same reasons as the Teacher Work Group:

One reason that performance assessments embedded in classroom instruction may help support stronger learning for students is that they ensure that students are undertaking intellectually challenging tasks. If teachers use these kinds of assignments consistently, with feedback and opportunities to revise to meet high standards, the level of rigor in the

classroom increases. In addition, these assessments can provide information to teachers regarding how students think and try to solve problems. This feedback allows teachers to diagnose students' strengths as well as gaps in understanding. Because performance assessment tasks often yield multiple scores in different domains of performance, reflecting students' areas of strength and weakness, they can also help teachers identify what kind of help students need, so they can tailor instruction accordingly (Darling-Hammond 29).

Implications of Recommendations for the Seattle Public Schools

- The district must immediately halt the administration of the MAP test and the use of the invalid data to sort students and evaluate teachers
- A “variety of measures” approach to assessing students’ achievement and growth should be used; the district must not haphazardly replace the MAP test with a different, single test
- The details of this “variety of measures” approach must be developed collaboratively with teachers and the community

This last point is a most important one and raises the question: how should such an approach be developed? While we do not have details, we do have general principles and guidelines:

- Groups of teachers—at grade levels, at individual schools or regions of schools, or by departments—must meet regularly to develop, critique, adjust, and evaluate assessments that clearly line up with standards
- These same groups must meet regularly to evaluate student work created through these assessments and to make regular changes to the assessment system
- These groups can use for their work current professional development time, including early release days and the two state mandated PD days
- The money currently used for assessments that do not meet the Criteria for Quality Assessment, such as MAP, should be put towards the research and development of assessments that do
- The district and teacher representatives can develop new ways to use contracted time for this work, including adjusting the school day hours, using substitutes judiciously, and creating common planning time during the day
- Teachers can create rubrics as well as suggested unit activities that are standards- and research-based. Both can be made available to all staff through the district’s Fusion pages. They should be used, revisited, revised, and used again in a cycle of redevelopment

The members of The Teacher Work Group on Assessment acknowledge that our proposals are not simple ones nor are they easy to implement. They represent, however, a system that respects students as learners, teachers as professionals, and our schools as places of true learning.

Works Cited

- Banda, Jose. "Message from the Superintendent." Letter. 23 Jan. 2013. Seattle Public Schools., n.d. Web. 1 Mar. 2013.
- Classified Joint Task Force—Meeting #2*. Meeting Summary. 2 Feb. 2013. Seattle Public Schools. n.d. Web. 30 Mar. 2013.
- Darling-Hammond, Linda, and Frank Adamson. *Beyond Basic Skills: The Role of Performance Assessment in Achieving 21st Century Standards of Learning*. Rep. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education, n.d. Stanford University. Web. 15 Mar. 2013.
- Neill, Monty. "Multiple Measures." *Pencils Down: Rethinking High-Stakes Testing and Accountability in Public Schools*. Ed. Wayne Au and Melissa Bollow Tempel. Milwaukee: Rethinking Schools, 2012. 231-234. Print.
- Roschewski, Pat, Jody Isernhagen, and Leon Dappen. "Nebraska STARS: Achieving Results." *Phi Delta Kappan* 87.6 (2006): 433-37. Web. 16 Mar. 2013.
- Sahlberg, Pasi. *Finnish Lessons: What Can the World Learn from Educational Change in Finland?* New York: Teachers College Press, 2011.
- Steptoe, Sonja. "How Nebraska Leaves No Child Behind." *Time* 30 May 2007: n. pag. *Time.com*. Web. 19 Mar. 2013.
- Tashlik, Phyllis. "Changing the National Conversation on Assessment." *Phi Delta Kappan* 91.6 (2010): 55-58. *ProQuest*. Web. 19 Mar. 2013.
- Williams, Kathy, and Barbara Minter. "Failing Our Kids: Why the Testing Craze Won't Fix Our Schools." *Pencils Down: Rethinking High-Stakes Testing and*

Accountability in Public Schools. Ed. Wayne Au and Melissa Bollow Tempel.

Milwaukee: Rethinking Schools, 2012. 9-11. Print.

"Wyoming Steers Clear of Exit Exams." *The National Center for Fair & Open Testing*.

N.p., n.d. Web. 03 May 2013.