A STEADILY FLOWING
Stream of INFORMATION GIVES TEACHERS MUCH-NEEDED DATA

BY THOMAS W. MANY AND CHRISTINE JAKICIC

Teachers at Woodlawn Middle School in Long Grove, Ill., are no different than most. They were discouraged by the amount of testing taking place in their classrooms, frustrated with not having the kind of data they needed, and feeling overwhelmed by the data they did have.

The teachers routinely analyzed results of standardized achievement tests, disaggregating the data. Based on the data, they tried to change pacing, instructional strategies, and even to add new units of study to improve student learning, but they eventually realized that the standardized annual exams were not providing the critical information they needed. Annual exams did not offer timely information and weren’t linked to the local curriculum.

The teachers also used classroom assessments — quizzes, unit tests, and essays to measure and monitor individual students’ growth. Every teacher created his or her own assessments that reflected what had been taught, but all the assessments were different. The lack of consistency prevented teachers from analyzing how their students had done on an agreed-upon standard compared with other students in the school.

Woodlawn teachers found they had an either-or proposition: The available assessments were either highly formative classroom assessments that individual teachers had developed or highly summative standardized exams created by publishers or the state department of education. The school’s assessment system was out of balance. The teachers wanted to make better schoolwide instructional decisions, and they needed assessment data that would help them do that.

They needed a continuum of assessments.

Toward a balanced and coherent assessment system

“I needed to find more frequent ways to tell me if the students were learning. I realized that there were some things kids weren’t getting that I thought they were.”

— A 7th-grade Woodlawn teacher

In 2002-03, Woodlawn teachers took a significant first step toward a better assessment system. Teachers collaboratively developed three writing prompts to be used during the
year. They also agreed to begin using the Gates-MacGinitie Reading Assessment twice a year to measure vocabulary and reading comprehension. All students took the same test at the same time, and teachers compared individual students’ performances, as well as each class’s results. Instead of a once-a-year snapshot of student performance, these assessments gave teachers more periodic information about how their instruction affected student achievement. Teachers thought this step would answer the data question, but it turned out to be the first step of a much longer journey.

Initially, reading teachers used the Gates-MacGinitie results to support individual students. Reading teachers also looked for patterns within their own data, and the whole staff looked at the aggregate scores. While using assessments administered two or three times a year was more summative than using the traditional classroom assessments, the assessments also were far more formative than the high-stakes state exams given to students once a year.

At Woodlawn, teachers are organized in interdisciplinary teams that are responsible for specific groups of students. They also are members of a subject-specific team such as the math or science team. Interdisciplinary teams meet daily, and subject-specific teams meet once a month. Despite the fact teachers had been organized around teams, the only teacher teams really using the new data to improve instruction were the reading teacher groups.

However, the initiative was blossoming into a districtwide effort. The district offered a data retreat to train teachers on how to use data to make instructional decisions. Teachers learned to understand and analyze the data, to make hypotheses, and to set goals from the data. These teacher leaders in turn used their new knowledge to work with their colleagues at each school during bimonthly staff meetings. The

The teachers wanted to make better schoolwide instructional decisions, and they needed assessment data that would help them do that.
A balanced and coherent system describes assessment in one of four categories — classroom, common, district, and external — based on how formative or summative the assessment is. The types of assessment are not discrete or mutually exclusive, but help focus discussion on the potential uses of data generated by each category of assessment.

**Classroom assessments** are the most formative and include the informal checks of student learning that go on in classrooms every day. This type of assessment is designed by individual teachers and provides ongoing feedback about how students are learning. Classroom assessments generate the kind of descriptive information teachers use for regrouping, reteaching, and monitoring individual students’ mastery of skills and strategies as they progress through the curriculum in their individual classrooms.

**Common assessments** are more formative than summative and are closely linked to the local curriculum. Common assessments are typically developed by teams of teachers at the building level and given at about the same time to all students within the school who are in the same class, course, or grade level. These assessments provide feedback about how individual students are progressing through the curriculum in comparison with other students. Most schools use a quarterly, monthly, or even a unit schedule for common assessments, use them to help generate a grade, and see them as a way to trigger systematic interventions.

**District assessments** are more summative than formative, but many teachers use the results in more formative ways for diagnostic information about students. These assessments are given districtwide at the same time to all students in the same class, course, or grade level. Examples are the Gates-MacGinitie, NWEA-MAP, or DIBELS tests, which provide diagnostic information about individuals and/or groups of students and often are used as entrance and exit criteria for programs such as Title I reading and math labs. These tests use a pre/post, fall/spring format and often are scored by outside vendors.

**External assessments** are the most summative. These are commercially developed, standardized, nationally norm-referenced tests and high-stakes state assessments that students take periodically. These external assessments are not linked to local curricula. Data from these assessments are used to identify and track trends over time, evaluate the relative strengths and weaknesses of curricular programs within a school or district, and generate ranks or metrics for public accountability.

### Overview of assessment

<table>
<thead>
<tr>
<th>Classroom assessments</th>
<th>Common assessments</th>
<th>District-level assessments</th>
<th>External assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most formative</td>
<td>More formative</td>
<td>More summative</td>
<td>Most summative</td>
</tr>
<tr>
<td>Ongoing student and teacher assessment</td>
<td>Collaboratively developed and curriculum-embedded</td>
<td>Identify groups of at-risk students</td>
<td>Ranks and benchmarks</td>
</tr>
<tr>
<td>Quizzes, essays, and projects</td>
<td>Pyramid of interventions</td>
<td>Entrance and exit criteria</td>
<td>Standardized achievement tests</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DIBELS, NWEA-MAP, Gates-MacGinitie</td>
<td>State and provincial exams</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Programmatic support</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DAILY</th>
<th>WEEKLY</th>
<th>UNIT</th>
<th>MONTHLY</th>
<th>SEMESTER</th>
<th>ANNUAL</th>
</tr>
</thead>
</table>

Teachers began to talk about reading and writing across the content areas and gathering more formative data in the other subject areas as well. Teachers began to think about assessment in terms of quality, matching the type and purpose, rather than quantity, the amount and frequency of assessment.

As teachers received more systematic results, they became convinced they needed assessments even more connected to their classrooms to give them better information about student progress.

In 2003-04, teams of teachers began developing additional common assessments. Teachers from across the
Teachers were so invested that they asked for additional released time during the school day to continue their work.

The district met in their subject-specific teams to write assessments that they hoped would answer their questions about how to ensure each student was learning. As they met, each teacher completed one assessment that all could administer during the first quarter.

Teams decided whether to use selected response, performance assessments, or written response items. They decided how many questions and what level questions (knowledge, analysis, or evaluation) to use. Each team made decisions based on what information they thought would best help them know how to help their students.

Once this first assessment was completed, the issue of time became vitally important. Teachers wanted to continue writing additional assessments and knew they would need time to discuss the results. What they realized was that much more had occurred than just writing a test. The conversations about what was important to include in the test affected how they subsequently felt about what was important to teach. Many teams also explored their beliefs about the concept of mastery. These were topics that recurred in meetings during the rest of the year.

The monthly after-school subject meetings lasted an hour, but teachers quickly realized one hour a month wasn’t enough time to write new common assessments. The district superintendent and building principals made a commitment to teachers to help find more time.

The time came in a variety of ways. First, job-alike monthly meetings were devoted solely to this process; all other initiatives were put on hold. Teachers were so invested that they asked for additional released time during the school day to continue their work. The district provided substitutes for all teachers who taught a particular subject to meet to continue their work.

Teachers also offered creative ideas to find more time. Aides covered assemblies, plays, and other schoolwide events while interdisciplinary and subject-specific teams worked on writing assessments and analyzing the results. The physical education department took students on a field trip, which allowed academic teachers time to work together. Teachers also used faculty meetings to accomplish this task.

One problem that emerged was that teachers began the process with their colleagues at the other middle school. Although teachers believed the best situation was to do this work as a team, they also realized that sometimes they needed to work individually or in school groups to get the project accomplished. What teams discovered was that once they worked together through the first assessment, they were more comfortable allowing the process to proceed with individual or smaller groups bringing back draft assessments for the group to consider.

One teacher described a typical meeting. There was no formal leader for the four 7th-grade science teachers. They all worked together. During this meeting, teachers reviewed their unit on simple machines. They discussed how much time should be spent on the unit. Although earlier discussions around pacing were difficult, experience allowed them to make this decision quite easily. Once they agreed what items should be assessed, they began discussing how they taught certain concepts, and one teacher felt comfortable enough to admit to having a difficult time with this particular unit. The other three jumped in and shared activities and experiments. As one group member explained, “Once we got the idea that these assessments weren’t going to be used to compare teachers, we were much more open to saying, ‘I need help with this.’”

Momentum began to build. Building-level teams designed their own agendas, including time to discuss pacing and curriculum content, to review assessment results, and to determine how and by whom additional assessments would be written. For the first time in many years, teachers reported feeling that their meeting time was well-spent.

The formal job-alike meetings at the district level and informal subject-specific team meetings at the building level created “a conduit to talk about things that matter,” as one

**Woodlawn Middle School**
Long Grove, Ill.

**Grades:** 6-8  
**Enrollment:** 690  
**Staff:** 54 certified staff; 18 support staff.

**Racial/ethnic mix:**  
- **White:** 87%  
- **Black:** 1%  
- **Hispanic:** 2%  
- **Asian/Pacific Islander:** 10%  
- **Native American:** 0%  
- **Other:** 0%

**Limited English proficient:** 2%

**Languages spoken:** 7

**Free/reduced lunch:** 3%

**Special education:** 16%

**Contact:** Christie Jakicic, principal
Woodlawn Middle School  
6362 Gilmer Road  
Long Grove, IL  60047

**Phone:** (847) 353-8500  
**Fax:** (847) 939-8237  
**E-mail:** cjakicic@district96.k12.il.us

**CHRISTINE JAKICIC** is principal of Woodlawn Middle School.
**THOMAS W. MANY** is superintendent of Kildeer-Countryside School District 96. You can contact him at 1050 Ivy Hall Lane, Buffalo Grove, IL 60089, (847) 459-4260, (847) 459-2344, e-mail: tmany@district96.k12.il.us.

**HOPE MARVIN** is principal of Long Grove Elementary School. You can contact her at 6362 Gilmer Road, Long Grove, IL 60047, (847) 353-8500, (847) 939-8237, e-mail: cjakicic@district96.k12.il.us.
Initially, the teachers were concerned that they would be compared unfavorably to their colleagues, and this was probably the biggest barrier leaders faced. By the end of the second year, however, some were asking colleagues for help when their own results were not what they wanted. As teachers felt more empowered to share their ideas, they also became more willing to share their results.

Just as important, the school learned that collecting any and all available information didn’t always lead to improvement. Too often before, the faculty had started down the road of “data dump,” and teachers realized that more information is not always better. Data that aren’t useful get in the way and, as a result of discussions, staff eliminated some assessments to make time for others.

The district helped by discontinuing a standardized, nationally normed test, eliminating a complicated and underused running records assessment, and redesigning, combining, and consolidating a number of unit and chapter tests into district assessments given once a quarter. Teacher teams accepted responsibility for developing more frequent common assessments and eliminated those assessments that had been routinely given to students without careful consideration of what would be done with the data.

Their goal was to write more frequent common assessments to identify students’ progress over time. These common assessments were intended to be more formative. Teachers used the data to adjust their instruction rather than waiting for the results of a more summative exam too late to make a difference for current students.

Instead of feeling that assessments were taking time away from instruction, teachers began to appreciate that good assessments could provide answers to questions about student progress, highlight needed changes in the curriculum, offer information that would suggest modifications to existing instructional practice, and give pacing cues for optimal learning.

The common assessments were systematic enough to allow teachers to compare student performance against an agreed-upon standard and were more closely linked to the local curriculum.

With the added periodic district assessments, teachers now had meaningful information about their students.

For example, one teacher described getting the results of the MAP (Measure of Academic Progress, a Northwest Evaluation Association standardized exam purchased by the district) and identifying why one student was struggling with reading. Rather than an ambiguous comprehension score from the standardized exam, she could see from the MAP score that this student was deficient in vocabulary skills. She could now work with the child more specifically on those skills.

Common assessments helped create a clearer map of what to teach. Teachers had a better understanding of what material they should cover. Because these common assessments were their own work, teachers didn’t feel pressured from outside or that change was imposed on them. They recognized that the process of creating formative common assessments assured students would have consistent learning experiences classroom to classroom and school to school.

UNEXPECTED RESULTS

Through professional dialogue, Woodlawn teachers have learned that a system relying solely on the most summative and most formative assessments falls far short of what is needed for effective instruction. They need a balanced and coherent system that includes a variety of assessments not limited to the highly formative or highly summative assessments typically found in schools.

From the teachers’ work, an assessment model emerged that includes four types of assessment arranged along a continuum from most formative to most summative, using factors such as the frequency, use, and audience for assessment results as additional guides. Viewing assessment as a continuum, teachers are better able to use data to guide instruction and understand which assessments are most appropriate for what purpose. Teachers need the timeliness and flexibility of traditional classroom assessments and use the results of external assessments to calibrate and benchmark student achievement levels. They also need data from common assessments to compare student progress on a regular basis with other students against an agreed-upon standard. Finally, they benefit from the systematic data generated by periodic district-level assessments.

Woodlawn teachers moved from concern, frustration, and being overwhelmed by testing to discovering positive and unexpected outcomes of their work. Nearly 90% of the school’s students now are meeting state standards.