Using Real-World Tasks to Assess Student Performance

Aimee Mack came to a quick realization last spring as she and colleagues pored over Connecticut’s new teacher evaluation system: Their jobs would depend on being able to share evidence of their language learners’ growth with the entire school community.

“We knew that as a department we were going to be looking at data as part of our evaluation,” says Mack, a French teacher and world languages team leader at Brookfield High School. “That made me wonder: Does a standardized test for world languages exist?”

She soon heard about something that seemed to fit the bill: the ACTFL Assessment of Performance toward Proficiency in Languages. Called AAPPL Measure or simply AAPPL, the online, performance-based test assesses students in the Interpersonal, Interpretive and Presentational modes of communication, with tasks that require learners to listen, read, speak, and write about topics commonly explored in language classrooms.

Each task—whether typing out an e-mail message, video chatting with a recorded native speaker, or making selections based on an understanding of something heard, read, or viewed—occurs in the context of a Standards-based classroom. Students can take one or all portions of the test; they then receive reports that classify their demonstrated level of performance and offer detailed suggestions for moving up. (Sample tests in Arabic, Chinese, French, German, Spanish, and Russian can be found at aappl.actfl.org/demo. More languages are being added, including English.)

AAPPL’s availability comes as states grapple with two interrelated, high-stakes challenges: (1) finding ways to evaluate teachers’ effectiveness in the classroom and (2) identifying assessments that reliably gauge what students know and are able to do. This quest is especially critical given that many states weigh—or soon will weigh—students’ standardized test scores heavily in teacher evaluations, and also because pressure has mounted to include traditionally non-tested subjects such as world languages and the arts in the assessment mix.

“We’re really seeing a confluence of things, from Race to the Top to the shift in teacher evaluations,” says Craig Waterman, Assessment Coordinator for District-Determined Measures at the Department of Elementary and Secondary Education in Massachusetts, where at least three districts have begun using AAPPL. “But student learning can’t just be (measured as) growth on traditional standardized tests. That’s not the full range of what we do in schools. The majority of teachers are teaching in content areas that aren’t currently assessed using standardized tests. Those teachers want feedback about what they’re doing. The goal is to expand the range of what we’re assessing, to cover the full range of learning that takes place in school.”

With such aims in mind, Mack piloted AAPPL last spring with 80 of her French students, who took a version of the test that assesses learners in the Novice-to-Intermediate range as described by the ACTFL Proficiency Guidelines 2012 and the ACTFL Performance Descriptors for Language Learners. (A second version targets Intermediate-to-Advanced learners.) Based on her positive experience, Mack and her colleagues had all their French and Spanish students complete the Interpersonal Speaking and Listening portion of the test this past fall. They plan to re-assess in the spring.

Mack and instructors at other districts that have piloted AAPPL say the content and format of the test have fueled discussions about what needs to happen in their classrooms in order for students to achieve proficiency gains. AAPPL’s proving grounds include Utah, where the state’s dual language immersion students now take the test.

“We know that instruction is always driven by what you test,” says Gregg Roberts, the state’s World Languages and Dual Immersion Specialist. “Students need proficiency targets so their language will be stretched. And we like having an external assessment because it’s standardized and we can see how well our programs are doing.”

Language departments have experimented with a wide variety of assessments in recent years, from district-generated tests to commercial models. In 2001, ACTFL and other organizations worked with the Center for Applied Linguistics to create the framework for the Foreign Language National Assessment of Educational Progress (FL NAEP) which targeted 2003 as the launch year for a Spanish-language test for students in Grade 12. The design of the FL NAEP pushed the boundaries of large-scale testing of languages by including a simulated conversation and designing authentic tasks to assess Interpretive and Presentational modes of communication. Due to funding cuts, the FL NAEP was never officially administered.

Out of that initiative came much of the research and thinking that ultimately shaped AAPPL, which was further enhanced by a FLAP grant and pilot testing in Glastonbury Public Schools and funding from several STARTALK projects. Language Testing International (ACTFL’s official testing office) programmed a production version and built the Internet delivery, rating, and client reporting platforms.
How AAPPL Is Scored

AAPPL test takers receive a score report that rates their level of performance in each of the three communication modes from N-1 up to I-5, along with concrete recommendations for reaching the next level.

A score of N-1 reflects the abilities described as Novice Low in the ACTFL Proficiency Guidelines 2012. Scores of N-2 and N-3 reflect Novice-Mid abilities, with N-3 being a stronger performance within the Novice-Mid range; N-4 reflects Novice-High abilities. This means that a learner who receives the score of N-4, in addition to performing all Novice level functions fully, also shows performance at the Intermediate level, but not consistently.

Within the Intermediate level, a score of I-1 reflects the abilities described as Intermediate Low in the ACTFL Proficiency Guidelines 2012. Scores of I-2, I-3, and I-4 are all in the Intermediate-Mid range. Given that Intermediate Mid represents a broad range of abilities, the delineation allows learners and teachers to determine where the performance falls within that broad range and to track progress within that range. I-5 reflects Intermediate-High abilities. This means that a learner who receives the score of I-5, in addition to performing all Intermediate-level functions fully, also shows performance at the Advanced level, but not consistently.

<table>
<thead>
<tr>
<th>ACTFL Proficiency Guidelines</th>
<th>ACTFL Performance Scale</th>
<th>AAPPL Measure Performance Score</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intermediate High</td>
<td>Intermediate</td>
<td>I-5</td>
<td>B</td>
</tr>
<tr>
<td>Intermediate Mid</td>
<td></td>
<td>I-4</td>
<td></td>
</tr>
<tr>
<td>Intermediate Mid</td>
<td></td>
<td>I-3</td>
<td></td>
</tr>
<tr>
<td>Intermediate Mid</td>
<td></td>
<td>I-2</td>
<td></td>
</tr>
<tr>
<td>Intermediate Low</td>
<td></td>
<td>I-1</td>
<td></td>
</tr>
<tr>
<td>Novice High</td>
<td>Novice</td>
<td>N-4</td>
<td>A</td>
</tr>
<tr>
<td>Novice Mid</td>
<td></td>
<td>N-3</td>
<td></td>
</tr>
<tr>
<td>Novice Mid</td>
<td></td>
<td>N-2</td>
<td></td>
</tr>
<tr>
<td>Novice Low</td>
<td></td>
<td>N-1</td>
<td></td>
</tr>
</tbody>
</table>

and district leaders wondered how students in different schools would fare, says Matt Rega, Director of Assessment. They were pleased, he says, when they read through the score reports a few days after tests were administered. “We found it was pretty consistent across buildings, with all (scores) falling within one sub-level on the proficiency scale,” notes Rega.

Minnetonka began using AAPPL this past fall as an alternative to homegrown assessments and a previous commercial test. From here on, the district plans to assess immersion students once a year in the spring as part of a districtwide drive to gauge progress toward grade-level proficiency targets.

AAPPL offers a comprehensive, one-stop assessment that students say they enjoy and which allows teachers to use a common vocabulary of proficiency when discussing student work and progress,” Rega says. “Teachers have even created end-of-year targets where students can see if they’re on track for achievement on [Advanced Placement] language exams. Students—and their parents—will know that if they continue to meet those targets that they’re in good shape to score high on the AP exam by the end of ninth grade.”

Minnetonka’s adoption of AAPPL reflects a growing wave of interest in language immersion programs in Minnesota and other parts of the country. Rega credits his own district’s program with helping reverse a negative enrollment trend by attracting students from other districts.

Seven years ago, he says, enrollment had dropped to about 6,600 students in Grades K–7. Now, six years after immersion came to Minnetonka, enrollment stands at 9,800. Rega says, with more than 2,200 students enrolled in the immersion program.

In Utah—a state that has become a national model of success for dual language immersion students—AAFPL is used as a national model of success for dual language immersion students.
immersion—the use of AAPPL is an important component of students’ proficiency-based learning, according to Roberts. Immersion students receive proficiency ratings on their report cards rather than grades, which parents have learned how to interpret at open house meetings. The state even brings in university professors trained in conducting modified oral proficiency interviews (OPIs) to check that students are on track to hit proficiency targets.

Many of Mack’s students in Connecticut viewed AAPPL as “just something else we were doing in class,” she says. Still, it was a different way of being evaluated. “Students had had a bit of one-on-one assessments with me, but the whole computer-based element was new to them. I like that the creators of the assessment go to lengths to make it look like (the native speaker) is listening. The kids are sometimes fooled: Is this person really seeing me?”

At a time of shrunken school budgets around the country, cost was a factor in Brookfield’s decision to adopt a commercial test. AAPPL costs $20 per test taker for the complete, four-part assessment. The interpersonal portion alone, which is rated by certified evaluators, costs $10 per student and the presentational writing portion $5. The interpretive reading and listening portions are offered together for $5.

Mack and her colleagues were asked to present—and justify—their wishes. For Mack, the sole French teacher at her high school, a key argument was the difficulty in finding an independent rater to evaluate large numbers of speaking samples. After hearing this and other teachers’ input, school board members approved the Interpersonal Speaking and Listening portion of the test for all French and Spanish students.

Brookfield took that action as part of an overhaul of its foreign language program, undertaken with guidance from Glastonbury Public Schools’ world languages department.

Glastonbury, which offers six languages taught by more than 50 teachers, played an important role in piloting AAPPL through work funded by a five-year FLAP grant. The district piloted AAPPL with high school students, recently added it as a benchmark measure for eighth graders, and may try it out at the fifth-grade level, says Rita Oleksak, Director of the district’s foreign language department.

“We’re looking at what the results tell us,” says Oleksak, a former ACTFL President who was recently named NADSFL Supervisor of the Year. “It presents a good opportunity for teachers to think through, analyze, and interpret the data. For instance, what if you have a student who’s great in class but only comes out at Novice High? And then you have a quiet student who comes out at Intermediate Mid. A teacher might say of the first student, ‘But he talked and talked!’ Then we have to ask, ‘But what did he really say?’”

Already, feedback from students who have taken AAPPL has led some teachers to change their strategies and emphases in class. Some of Mack’s students, for example, said they struggled to recall some of what they had learned in earlier years of study. “Many upper-level students said they wished they had remembered some older vocabulary, for example, they hadn’t really had school vocabulary since French I.” That, she says, made her “think about spiraling and always going back to reinforce the most basic skills. It also has forced me to focus more on speaking and reading. Part of my evaluation is: How can I incorporate the things that are emphasized in the test in my everyday instruction?”

Mack also has taken time to explain to students the value of the score report and the strategies for improvement that each includes. “We’ve talked about [the Novice ratings of N-1 and N-2] and what each means. We had never had that conversation before. Before, it was, ‘You’re getting an A or a B.’ This isn’t about a letter grade. It’s about improving. It’s the idea that everyone is at a different level and all they’re trying to do is go up a level. If my students achieve that goal when we do the spring assessment, that’s information I’ll want to share with people.”

Performance assessments such as AAPPL, with the emphasis on measuring what students can do with language, go to the heart of effective communication, says Camargo of WestEd, a devout linguaphile who is currently teaching himself Korean.

“Language is about communication and exchange. It’s not information alone, but using information in a particular context to move things or other people. What AAPPL does is measure the efficacy of that exchange, and that’s very novel,” says Camargo.

What’s more, he adds, “it focuses on what it takes to move along that continuum from Novice to Advanced. It answers this very basic question: ‘When it comes to communication, can this person get the job done?’”

It answers this very basic question: “When it comes to communication, can this person get the job done?” —Carlos Camargo

Third-grade dual immersion students in French, Spanish, and Mandarin took the Interpersonal Speaking and Listening portion of AAPPL on a required basis for the first time last spring.

“Our elementary immersion program is driving our secondary program, as parents demand that the high school program incorporate proficiency targets as well,” Roberts says. “I always tell colleagues (in other states), if your state standards are proficiency-set, it’s easier to get immersion schools in place and attract younger students. That gets the high school and secondary teachers to take notice, and they start to really create and design their curriculum around proficiency.”

For students and teachers, AAPPL hasn’t produced the type of anxiety usually associated with standardized tests, Roberts says.

“Teachers see the benefits of having the (speaking) prompts as part of their instructions. And the kids love the test. They say, ‘Wow can we do that again?’” Roberts says. “Parents (love it), too. They say, ‘Someone’s finally telling me what my child can do with the language they’re learning.’”

Many of Mack’s students have taken AAPPL with the emphasis on measuring what students can do with language, go to the heart of effective communication, says Camargo of WestEd, a devout linguaphile who is currently teaching himself Korean.

“Language is about communication and exchange. It’s not information alone, but using information in a particular context to move things or other people. What AAPPL does is measure the efficacy of that exchange, and that’s very novel,” says Camargo.

What’s more, he adds, “it focuses on what it takes to move along that continuum from Novice to Advanced. It answers this very basic question: ‘When it comes to communication, can this person get the job done?’”