The Ohio Performance Assessment Pilot Project: A Test-bed for Ohio's Task Dyad Learning System Lauren Monowar-Jones Terrence Moore Ohio Department of Education

The Ohio Performance Assessment Pilot Project (OPAPP) is funded by of Ohio's Race to the Top federal grant. OPAPP is piloting performance assessment in high schools and elementary schools in Ohio. The tasks being piloted are designed to reflect the Task Dyad Learning System model (Moore, Monowar-Jones, & Xu, 2012) and are delivered to students in an online environment. This pilot is a miniature large scale assessment that includes reviews by external committees and a scoring plan that is scalable.

Background.

The Ohio Performance Assessment Pilot Project (OPAPP) began in 2008 with Gates Foundation money. At that time, the Ohio Department of Education (ODE) was interested in exploring whether performance assessments could be implemented in schools in Ohio. ODE partnered with Stanford's School Redesign Network and several curriculum-embedded performance assessments were designed and piloted in 33 high schools in 11th and 12th grade classrooms around Ohio. Teachers were involved in the development process, but did not write the tasks or create the rubrics. Once the tasks were implemented in the classrooms, teachers were brought together to score the student work.

Based on observations during the second year of this pilot, ODE altered the trajectory of the pilot. We noticed that there were many positive outcomes from the initial use of curriculum-embedded performance assessments in Ohio schools.

- Teachers reported that students were more engaged in the learning associated with the tasks.
- For some of the tasks, students seemed to be able to demonstrate greater depth of understanding than was anticipated.
- Students produced large bodies of work that demonstrated mastery or growth in many skills that were desired.
- During a second administration, teachers adjusted their classroom styles and were able to get their students to achieve even better results the second time around.

And we also noticed some serious drawbacks to the use of curriculum-embedded performance assessments.

- The large size of the rubrics (six to seven pages) was too cumbersome and made it difficult to train teachers to use it reliably.
- The complexity of the student products (usually 10 to 20 pages) made it difficult to train teachers to score them reliably.
- The length size of the student products also made scoring a long process (usually 20 to 30 minutes per student sample) to score.

Moore, et.al. (2012) note that all of the positive outcomes are related to teaching and learning and all of the drawbacks are related to measurement. To preserve the benefits and remove the drawbacks, what is now known as Ohio's Task Dyad Learning System was developed. The system is comprised of two components: a Learning Task and an Assessment Task. The Learning Task is very much like a curriculum-embedded performance assessment, but it is not used for measurement. The Assessment Task is the component used for measurement. The Learning Task is for formative purposes while the Assessment Task is better for summative purposes.

OPAPP was designed to pilot this model and to pilot an assessment system that is scalable. Further, the pilot uses an online delivery system for the tasks and the pilot has an extensive professional development component for teacher participants. At its inception, the pilot was designed to mirror the plans of both the PARCC and SBAC plans for summative performance assessments, because Ohio had not yet chosen a consortium, but was actively participating in both. The hope was that the pilot would inform the development of the Next Generation Assessments for Ohio (both the Common Core assessments that will be provided by the assessment consortia and the assessments that Ohio will administer for Ohio's Revised Standards in Science and Social Studies). While there are still no specific plans to include performance assessment in the Ohio assessments of Science and Social Studies, the inclusion of these subjects in the pilot allows for the possibility of their inclusion in Ohio's Next Generation Assessments.

Pilot Details.

OPAPP is designed to pilot several different features of a potential future assessment system at once. The intense scrutiny and feedback from teachers has allowed the successful teasing out of factors relevant to the successful implementation of an operational performance assessment system. Additionally, the pilot involves the interaction of many individuals who are participating in different roles and on different levels. There are, from schools, classroom teachers, administrators, specialists (special education, gifted, ELL teachers), technology liaisons, and media specialists. Additionally, coaches were hired by ODE to support the teachers, and higher education experts were hired by ODE to (1) support the teachers, and the development of tasks, and (2) to engage higher education in a dialogue about teacher preparation. Ohio also contrated with Measured Progress for professional development (PD) and task development, TaskStream for the online delivery system, Tvsvpro for videography, and AIR for evaluation of the project as a whole. And finally, there are the many ODE staff involved in the project, albeit mostly behind the scenes on things like task development, support and delivery of some PD, data analysis, survey management and delivery, and the management of seemingly trivial but essential things like attendance, contact hours, purchases of supporting materials, mailings, etc.

As stated earlier, one of the main things this pilot is doing is piloting Ohio's Task Dyad Learning System. The dyad, itself, is a concept that must be grasped by the participating teachers. This has been accomplished somewhat successfully through professional development. Like students, teachers learn by doing and the dyad system is internalized through the process of implementation. Currently, over 100 high school teachers in nine school districts in Ohio have just completed their first dyad implementation. The Learning Tasks were implemented in February and March and the Assessment Tasks just closed on April 13, 2012. Next week, the student responses will go before a range finding committee of teachers who will score a subset of the student responses and the Measured Progress scoring staff will create training materials. Starting on April 30, 2012, participating teachers will meet to score the Assessment Tasks they implemented over the first two weeks of April.

The scoring sessions in May will bring about a crucial moment for these teachers. Last year, we piloted a quasi-dyad system. We asked Stanford's SRN to create Assessment Tasks to go with the curriculum-embedded performance assessments left over from work in 2008. We piloted those tasks and scored them with teachers. Following the scoring of the Assessment Tasks, so many teachers had moments of revelation that it was hard to miss the importance of *doing* the implementation and scoring. (Ironic: Doing is important for understanding performance tasks!)

Teachers in OPAPP will be implementing Learning and Assessment Tasks, scoring Assessment Tasks and participating in *writing* Learning Tasks. The idea of engaging teachers in the writing process was to build capacity in the state to support these formative curriculum tools beyond the pilot project. Starting in the Fall of 2012, our first expansion cohort will begin training on task writing. The tasks they create will be the ones they pilot in the second semester of that academic year starting in January 2013. Measured Progress will still design the associated Assessment Tasks, because of their nature as secure items, but the creation of Learning Tasks by teachers and for teachers is a sustainability goal of OPAPP.

Another key aspect of the Next Generation Assessments that this project is piloting is the online delivery system. Originally, the reason for going online was to stop the insanity of mailing hard copies of student work while trying to protect student identity through coding student aliases for labeling work samples. The inclusion of this dimension has proven to be very fruitful for learning how to deal with

many aspects of online testing. These aspects include, of course, the issue of availability of technology, but also teacher and student technological literacy, and the development of tasks for a non-native online system. We did not anticipate that this last bullet would be a hurdle at all. It seemed that developing tasks and putting them in an online environment would be easy. However, there is more groundwork that needs to be laid to make this process seamless to the users. In retrospect, this makes sense, but still we did not anticipate the struggles we've encountered along the way.

Finally, the project was designed to pilot a scalable assessment system. This involved engaging committees of stakeholders to review the tasks for fairness, and sensitivity, and content. Also, we have employed range-finding committees to help with the development of materials for training teachers to score the task responses. These committees mimic Ohio's current assessment system and use extant structures to support this interaction with stakeholders.

Another issue of scalability is the pilot's emphasis on teacher professional development. In OPAPP, teachers are participating in eight days of professional development in Stage 1 of the pilot (the first semester of participation). When this system is scaled up, this level of intense professional development will be too costly to maintain. For this reason, we have been working with the videographer (Tvsvpro) and Measured Progress to develop professional development modules that will be accessable online. Measured Progress will work with ODE and Tvsvpro to create nine modules:

- One module on Formative Instructional Techniques for both high school and elementary school teachers.
- Two modules for high school teachers only one on Feedback and one on Reengagement.
- Two modules for elementary school teachers only one on Feedback and one on Reengagement.
- Two modules on scoring.
- Two modules on Learning Task development.

Lessons Learned.

Over the course of the nearly nine months that this project has been running, we have learned a lot. At this point, we would call the pilot very successful in that it has already taught its developers many lessons for how to scale up. To begin with, the engagement of so many people in so many different roles has really been a benefit to the project. In schools where the implementation has involved a cohesive relationship among all the participants (standard teachers, specialists, technology liaisons, media specialists, and administrators), we have seen the greatest success stories. For this reason, in the application process for Cohort 4, we are asking for districts to describe their existing relationships among these groups as a way of both preparing them for the pilot needs and to determine a selection criterion to allow the pilot to avoid roadblocks that may exist further down the road in implementation.

From the quasi-pilot we have results that indicate that the dyad model is successful in promoting student learning of the target skills. With the upcoming data from the pilot this semester, we will have more information about the successes and challenges of the dyad model.

We have asked the original (cohort 1) teacher participants to make an attempt to create Learning Tasks. So far, those efforts have resulted in some changes in our plans for training the current teacher participants. Once we implement those plans, and examine the Learning Tasks that teachers produced, we will have more insight into that process as well.

The use of the online delivery system has brought to light some expected and some unexpected challenges. How the task appears in the online environment affects the task design process in ways we did not anticipate. We did anticipate difficulty with technology access, although we did not expect that to happen in the first pilot. To mitigate the effects on students, we asked teachers to pilot in only ONE section of ONE class. We were aware that many schools had limited access to technology (a computer lab or two per school is typical), but we did not think that getting students in and out of the labs would present such a huge logistical challenge to school teachers and administrators. We are currently seeking

out constructive long-term solutions to this issue. Further we underestimated the challenges with teacher familiarity or comfort with technology. This has proven to be a roadblock for some teachers who don't use technology for learning and for teachers who use other technology for teaching (i.e. math teachers use graphing calculators, not computers or Excel). Students seemed more willing to adapt to this change than teachers, but not always. In particular, math proved to be a very challenging subject. We are looking into creating or using existing applications that can be embedded into the online delivery environment to solve these issues.

So far, scalability has only proven to be challenging because of the fast cycle of development. We are developing in one semester and piloting the next. It has been challenging to find individuals who will sit on committees to do the work. But given the short timescale on which we are operating, we have been very successful with this aspect so far.

The plans for testing the scalability of the professional development are underway. We have a storyboard for the first module and are working to identify the video clips for this module by the end of April 2012. We hope to have the first three modules ready for use in the Fall when Cohort 4 (the first group of teachers to use the modules) will come on board.

References

Moore, T., Monowar-Jones, L., & Xu, M. (2012). *The Task Dyad Learning System: Concept and Experiments in Technology Assisted Formative Assessment*. Paper accepted for presentation at 2012 National Conference on Student Assessment (CCSSO), Minneapolis.