

Testing the Dyad Learning System in Ohio Schools: The Ohio Performance Assessment Pilot Project (OPAPP)

Lauren Monowar-Jones
Terrence Moore
Ohio Department of Education

Abstract. *With Ohio’s Race to the Top (RttT) federal grant, Ohio is piloting performance tasks in high schools and elementary schools. Ohio is piloting a particular model of performance tasks: Ohio’s Task Dyad Learning System. The Ohio Task Dyad Learning System consists of a Learning Task and an Assessment Task. The Learning Task is used formatively for instruction and the Assessment Task is used for summative assessment. The purpose of this poster is to describe Ohio’s experiment to test the use of the Task Dyad Learning System and a scalable on-line assessment system. This paper describes the experimental design in which the proposed dyad learning system is being tested.*

Background

The Ohio Performance Assessment Pilot Project (OPAPP) is complex and has many objectives including the objective to test the use of the Task Dyad Learning System. Additionally, however, the pilot is testing an online task delivery, storage and retrieval system, online teacher scoring of tasks and the use of intense professional development designed to promote the use of curriculum tools and give students feedback that results in increased student learning.

Ohio’s Task Dyad Learning System

OPAPP is piloting a performance assessment system that involves both a formative component and a summative component – Ohio’s Task Dyad Learning System. Ohio’s Task Dyad Learning System consists of a Learning Task and an Assessment Task. Figure 1, below, shows the Task Dyad Learning System in a Closed Loop Model.

Closed loop control system model

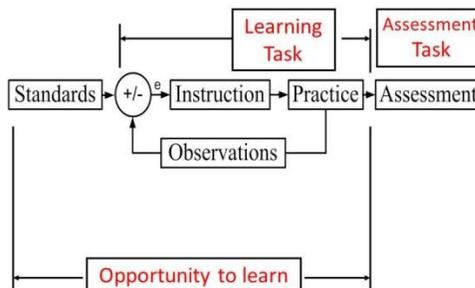


Figure 1.

The Learning Task is a formative curriculum-embedded performance task. For each Learning Task there are several Assessment Tasks so that the learning task might be used in multiple years while the assessment task could be used but a single time. Learning Tasks may contain many learning goals aligned to a wide range of standards. Assessment Tasks are limited to alignment to one Topic or Cluster in the standards. Each Assessment Task will assess a skill that was practiced in the Learning Task. It is important for this to be true to guarantee Opportunity to Learn. Students who complete the Learning Task should have had an opportunity to learn and practice the skill that is assessed and students should have received feedback from their instructor on their performance of this skill so that the student can improve.

Objectives or purposes

The purpose of this paper is to describe Ohio’s experiment to test the use of a dyad learning system and a scalable assessment system. The rationale behind the development of Ohio’s Task Dyad Learning System is detailed in Moore, Monowar-Jones, & Xu (2012). This paper describes the experimental design in which the proposed Task Dyad Learning System will be tested.

Additionally the pilot will test an online task delivery, storage and retrieval system, online teacher scoring of tasks, and the use of intense professional development designed to promote the use of curriculum tools. Table 1 below shows the structure of the project and the schedule of events for the participants.

Table 1.

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
Cohort 1	Pilot Online System	Pilot Task Writing				
Cohort 2 (High School)	PD	Pilot Tasks Score	Pilot Tasks Score Develop Tasks	Pilot Tasks Score Develop Tasks	Pilot Tasks Score	
Cohort 3 (Elem School)		PD	Pilot Tasks Score	Pilot Tasks Score Develop Tasks	Pilot Tasks Score Develop Tasks	Pilot Tasks Score
Cohort 4 (High School)			PD (online)	Field Test Tasks Score	Field Test Tasks Score	
Cohort 5 (Elem School)				PD (online)	Field Test Tasks Score	Field Test Tasks Score

Perspective(s) or theoretical framework

The Ohio Performance Assessment Pilot Program is piloting a performance assessment system that involves both a formative component and a summative component, called Ohio’s Task Dyad Learning System. Teachers will use the Learning Tasks and the Assessment Tasks in their classrooms.

Cohort 1 consists of about 50 high school teachers who have been participating in OPAPP since 2008. Cohort 2 consists of about 120 newly recruited high school teachers. Cohort 3 consists of about 90 elementary school teachers.

Cohort 2 and 3 teachers receive 64 hours of professional development on the use of formative assessment techniques that support implementation of performance tasks before they pilot any tasks. In the second semester of the program for both of these cohorts, the teachers will pilot a Learning Task and an Assessment Task with their students. The tasks will be delivered via an online system that delivers the tasks, stores the student responses and allows teachers to score the student responses.

Through the proper use of these tools, classrooms should be transformed into places where students are almost always actively engaged in learning. The expectation is that there will be an increase in student learning and achievement as well as a significant shift in classroom culture. The aim of the evaluation program for this pilot will be to measure the changes in the classroom environment as well as the changes in students learning and achievement.

The first question about the dyad learning system that must be asked is this: Do students who complete the Learning Tasks do better on the Assessment Tasks? From the experiments we ran in 2011, the answer to this question seems to be “yes”. This is important because teachers will be more motivated to use the learning tasks and the formative feedback techniques the learning tasks support when the easiest way to get points on the Assessment Task is to use the Learning Task and to become skilled in formative feedback.

Other key research questions related to the dyad system that this pilot will attempt to answer are:

- What steps do students take during the Learning Task that result in success on the Assessment Tasks?
- What valid inferences can be made about student achievement from the scores on Assessment Tasks?
- How many Assessment Tasks are necessary to make valid inferences about student achievement?
- How reliably can teachers score student work on Assessment Tasks?
- How efficient can a scoring system be that includes teachers scoring student responses delivered via an online system?

Methods, techniques, or modes of inquiry

The pilot that is underway will arm teachers with the skills to deliver the Learning Tasks, and then pilot the Learning Tasks and the Assessment Tasks in real classrooms. Students and teachers will be video-recorded while the tasks are being implemented to document the experiment and to study how students respond to the stimuli.

Additionally, students, teachers, administrators, coaches and parents will be surveyed often to see where their understanding, perception and beliefs about the system are situated and whether they change. Also, an alternative evaluation tool will be used to validate the inferences made from student performance on the piloted Assessment Tasks.

Teachers, students, coaches, administrators and parents will be interviewed to probe how they perceive the use of the dyad system as a part of the classroom culture and the assessment system structure.

Resources, evidence, objects or materials

The data sources for this experiment will be the students, the teachers, the coaches, the administrators and the parents. The evidence we will be seeking is evidence of learning, evidence of improvements in teacher technique or ability to implement the tasks successfully. These data can be gleaned from the surveys, video interviews and video recordings of the classroom environment as well as the test scores.

Results and/or substantiated conclusions or warrants for arguments/point of view

The soundness of the design of the experiment is perhaps a more appropriate designation for this section as the paper itself is about the (quasi-)experiment(al) design and not the experimental results. The pilot is designed to allow for extensive data collection to be used to evaluate many different aspects of the pilot project. In addition to surveys, pre-post assessments that run parallel to the performance tasks and interviews, the pilot includes video recordings of operating classrooms where tasks are being implemented.

These video recordings can be used to substantiate and support the claims made based on the data (quantitative measures) and interviews (qualitative/rich measures). The experiment, as designed, leaves much room for many research questions to be investigated.

In particular, we are interested in understanding the validity of the inferences that we will draw from the performance Assessment Tasks. For this, we envision the use of multiple-choice short-cycle assessments used pre- and post- task (both learning and Assessment Tasks) to demonstrate that the tasks are assessing the domain of knowledge and skills identified by the task writer. Also we are interested in doing a comparison of student performance on tasks across different content areas to demonstrate that the tasks are measuring content-specific skills and knowledge.

In the area of inter-rater reliability we will devise an operational scoring plan that allows for live assessment of inter-rater reliability and has built-in corrective action. Corrective action for a teacher who did not have high enough agreement on scores of the “known” papers would be to attend an online re-training professional development session designed for teachers with low agreement ratings. The student work would be re-scored by a regional moderation panel using moderators who train teachers to score student work.

The purpose of the piloted concept is to transform instruction, so the evaluation of the pilot should attempt to measure this transformation. We are interested in gathering data on this transformation from at least three sources: teachers, students, school administrators, and possibly parents.

From students, achievement as measured by state assessments and as measured by teachers (via a report card) will be collected to compare to student achievement as measured by the Assessment Tasks. A measure of the value that students place on the instructional approach we are piloting and student perceptions of the instructional approach we are piloting are two more data points we would

like to have. We hypothesize that these measures may change over the course of the pilot, so these data will be taken at several points throughout the pilot program.

From teachers, we are interested in learning at least about teacher beliefs about teaching, about student learning, as well as about student value of the instructional strategy we are piloting. We are also interested in teachers' perception of their own teaching style and how that changes over the course of the pilot program.

From administrators, we envision data collection on at least student achievement by school as measured by state assessments, the administrator's perception of teacher collaboration over time, the administrator's perception of student value of the instructional strategy being piloted, the administrator's perception of the teaching techniques used and a running tab of all transformational processes implemented in schools to accommodate the pilot program (i.e., extra prep time, PD, support).

Scientific or scholarly significance of the study or work

The primary significance of this work is to grow the body of knowledge and inform the assessment consortia about the use of the dyad learning system. Currently, neither consortium is considering providing curricular components that align to their assessments (as might be expected in a country dominated by state educational policies). Should this practice be deemed more reliable, more equitable, and more valid, this support the need for a dyad learning system.

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.