

# **Recommendations Related to the Operational Implementation of Performance Assessments Within Ohio's K-12 Assessment System**

A paper written for the Ohio Department of Education

Pearson

January 13, 2010



# Table of Contents

1. Introduction .....	1-1
2. Recommendations Related to Program Design, Local Buy-In, and Program Management.....	2-5
3. Recommendations Related to Development of Operational Performance Tasks ..	3-13
4. Recommendations Relating to Scoring, Reliability, Validity, and Comparability ....	4-17
5. Recommendations for a Web-based Performance Assessment Management System .....	5-25
6. Looking to the Future .....	6-33
7. References.....	7-35
Appendix A: Resources.....	A-37
Appendix B: EdExcel Overview.....	B-49

## Contributors

**Kelly Shasby Burling, PhD**, Research Scientist, Psychometric and Research Services  
**Robert P. Dolan, PhD**, Senior Research Scientist, Psychometric and Research Services  
**Jeri Frank, PMP**, Program Manager, National Services  
**David Full**, Proposal Manager, Proposal Services  
**Wesley E. LaMarche**, Director, Solutions Architecture  
**Paul Nichols, PhD**, Vice President, Psychometric and Research Services  
**Nivedita Shilpi Niyogi**, Vice President, National Services  
**Kurt Rogahn**, Senior Proposal Writer, Proposal Services  
**Daisy Vickers**, Director, Performance Scoring, New Business Development and Consulting  
**Walter “Denny” Way, PhD**, Senior Vice President, Psychometric and Research Services  
**Natasha J. Williams, PhD**, Senior Research Scientist, Psychometric and Research Services

# Recommendations Related to the Operational Implementation of Performance Assessments Within Ohio's K-12 Assessment System

## 1. Introduction

### Background Legislation

Ohio is taking proactive steps to reform statewide student assessment at the high school level. Enacted in July 2009, Ohio's House Bill 1 requires the State Superintendent and Chancellor of Higher Education to promulgate by rule a plan that replaces the Ohio Graduation Tests with a three-part assessment system consisting of the following components:

- A nationally standardized college entrance assessment
- A series of end-of-course (EOC) exams
- A senior project (ORC 3307.0712)

The nationally standardized college entrance examination, administered to all juniors, will cover the subject areas of English, mathematics and science. One administration will be covered at state expense and results will be shared with students and published on school, district, and state report cards. Ohio will use exam results for school- and district-reporting purposes and to inform students about college admissions and course placement decisions.

A suite of up to nine EOC exams will be given in English language arts, mathematics, science, and social studies. It is expected that the EOC exams will comprise a composite graduation requirement, that is, a composite score of all tests taken should reach or exceed a minimum composite score necessary for students to graduate. This compensatory model would replace Ohio's current conjunctive requirement to pass all tests. In addition, students who do not meet expected performance levels on any EOC exam will be given re-test opportunities and have the higher score applied to the composite score. Finally, it is encouraged that the tests should be designed, administered, scored and reported as part of a multi-state consortium which has the benefits of reducing costs and increasing efficiencies in a number of ways. The EOC exams will be aligned to the Ohio Core graduation requirements. Item formats will continue to consist of multiple-choice and constructed response items. In addition, the state is considering the inclusion performance tasks, embedded in curriculum materials

provided to supplement high school course syllabi. These performance tasks would provide the basis for locally developed or locally managed performance assessments.

The senior project will give students the opportunity to demonstrate what they know and can do, perhaps in the “21st century skill” areas, using measures other than standardized assessments. To meet the intent of House Bill 1, districts will have to certify whether students have successfully completed the senior project.

The motivation for this reform is to better prepare graduating Ohio students for success in college and the workplace. Ohio’s efforts align closely with the goals of the federal Race to the Top Assessment competition. They follow national trends aimed at increasing the rigor of high school instruction and assessment and increasing focus on 21<sup>st</sup> century skills such as critical thinking and problem-solving, interdisciplinary learning, creative thinking, technology, personal awareness and self-management.

Ohio’s assessment reform effort also recognizes recent attention given to assessment systems prominent in many high-achieving nations on international assessments, such as the Program for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS). Many of these systems combine centralized (state or national) assessments with local assessments given by teachers, which are factored into the final examination scores. The local assessments may include research papers, science experiments, presentations of various kinds, and collaborative projects. Though mapped to the centralized syllabus or standards for the subject, the local assessments are typically designed, administered, and scored locally but sometimes with state oversight or assistance.

## The Ohio Performance Assessment Pilot Project

Ohio has begun steps to build a new assessment system through the Ohio Performance Assessment Pilot Project (OPAPP), a research effort in collaboration with Stanford University’s School Redesign Network. The two-year pilot, funded by the Gates and Hewlett Foundations, began in 2008. The goals of the pilot are the following:

- Learn what it takes to construct and effectively implement performance assessments
- Build the capacity of Ohio educators and the K-12 system to implement a performance assessment system
- Evaluate the value and viability of incorporating performance assessments as one piece of Ohio’s assessment and accountability system

In the project’s first phase, a set of standards-based performance assessments and rubrics developed in English language arts, mathematics, and science will be field-tested across the state and scored between January and August of 2010.

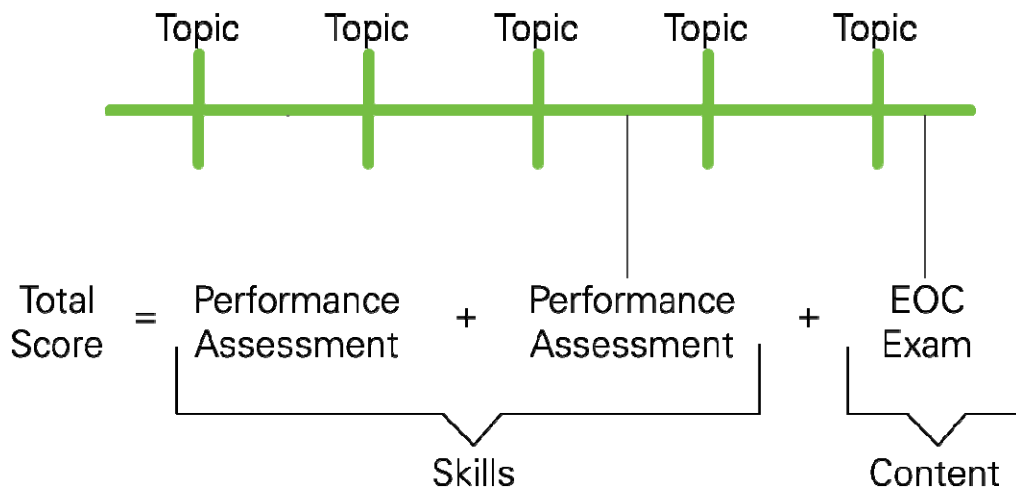
Phase II of the OPAPP will expand the performance assessment system to include social studies and career technical education. Based on Phase II plans, a task bank will be developed to support ongoing work, the network of participating Ohio districts, schools,

and educators will expand and their capacity for carrying on the work into the future will increase, and an initial technology platform to support the performance-based assessment system will be developed.

## Performance Assessments in Ohio’s High School Assessment Reform

One model envisioned by Ohio is to use performance assessments in conjunction with EOC exams. The figure that follows depicts how performance assessments could be combined with a summative EOC exam to yield a total score in a given subject.

### Course Instructional Progression



**Figure 1. Potential Combination of Performance Assessments and EOC Exam.**

The performance assessments and the summative EOC exam could combine to produce a total score that would contribute to the graduation composite. In addition, for selected EOC areas (for example, English II, Algebra I, Biology or Physical Science), the total score could be used for school accountability measures. From the standpoint of traditional psychometrics in the United States, the idea of using locally scored performance assessments in this manner is controversial. However, there are ways to achieve reliability and validity of performance assessment scores that contribute to a balanced EOC assessment system.

A second way performance assessments could directly relate to Ohio’s high school assessment reform is in the context of the “senior project” component. Although implementation of this component could take place in a variety of ways, many aspects of the OPAPP project likely will apply to these assessments, including training and professional development, scoring, and electronic storage and documentation of project artifacts.

## Purpose and Preview of the Paper

Ohio has invited Pearson to help conceptualize how the state might implement performance assessments operationally within its high school assessment system. Pearson brings a number of capabilities relevant to this task, including experience with developing tasks, training, and scoring for large-scale performance assessments, expertise in research and applied psychometrics, and a capacity for developing and applying assessment technology in large-scale systems. We see this as an opportunity to help focus and articulate our development and delivery strategies related to performance assessments in ways that will position us to collaborate with Ohio, its research partners, and other states interested in pursuing performance assessments as part of a balanced assessment system.

The purpose of this paper is to provide discussion and recommendations related to the operational implementation of performance assessments within Ohio's assessment system.

We have organized our discussion into the following four major sections:

- Program design, management, training, and professional development
- Developing operational performance tasks, rubrics, and scoring protocols
- Establishing and maintaining reliability, validity, and score comparability
- Developing and maintaining an e-portfolio / performance assessment management system

In each section, we concentrate on the operational challenges and threats to valid and reliable scores and interpretations that must be addressed as such a program is implemented. In two appendices listed at the end of the paper, we describe some Pearson capabilities that are relevant to portfolio assessment. In Appendix A, we describe Pearson tools and resources that might be applied within a web-based performance assessment management system. In Appendix B, we provide a brief overview of EdExcel, a U.K.-based Pearson company offering academic and vocational qualifications and instruments comprised of constructed response and performance-based assessments.

## 2. Recommendations Related to Program Design, Local Buy-In, and Program Management

Ohio has begun steps to implement a new assessment system—particularly at the critical high school level—through new legislation and new initiatives such as the Ohio Performance Assessment Pilot Project (OPAPP). Implementing an assessment system that emphasizes capturing student performance in the classroom as part of an accountability system will present the state with many challenges—and opportunities. We examine both in this section. We want to articulate a vision of what broad scale deployment of OPAPP would require, and additional challenges the system will face in the future as technological resources become more readily available in classrooms.

Based on the current model of performance tasks, administration, and scoring, one of the most significant challenges to statewide implementation will be logistics. Throughout this paper, we detail the processes and procedures that could be employed to create a technically sound and efficient performance assessment system that can be incorporated into the state’s accountability model. Each piece of the system is complex and requires thoughtful development, which in turn requires coordination of a multi-stage project schedule and implementation timeline.

Pulling all the pieces together, verifying the quality of their delivery, and monitoring the complex interconnected timelines of public awareness and engagement, item development, professional development, development of the technology platform, training on the technology platform, administration, scoring, and moderation requires vision, experience, organization, and the ability to deliver.

### Vision for Broad Scale Deployment

Statewide deployment of the OPAPP system will require a multi-stage project schedule and implementation timeline that includes the following:

- Statewide **public awareness and engagement** campaign for teachers, administrators, and other stakeholders on the purpose and intended benefits of the program
- Continued **assessment design and development** based on the work of the OPAPP project to expand the item bank and add subject areas
  - Work with teachers, School Redesign Network (SRN), and stakeholders to create additional tasks for each subject
  - Take tasks to review committees
  - Work with teachers, SRN, and stakeholders to create scoring rubrics
  - Work with teachers, SRN, and stakeholders to identify exemplars for scoring

- Develop, test, and deploy the **technology platform**
- Develop administration support
  - Provide online test prep/ task prep materials through data management system for local access prior to the administration windows
  - Provide teachers access to tasks and the data management system at the requested time (within the allotted window)
  - Provide technical support for the assessment and the platform
  - Monitor submissions of evidence for deadlines
  - Send confirmation receipts when artifacts are saved to the system
- Develop and implement model for **scoring** and **quality assurance**
  - Train-the-trainer model to create performance task assessment and scoring experts in each district and school, using the model developed during Phase I of OPAPP
  - Provide in-person training on the assessment tasks, embedding them in the curriculum as well as on the learning outcomes and the rubrics
  - Provide online training and qualification exercises for teachers
  - Provide scoring directors and supervisors for questions and guidance, available by phone, instant message, video-enabled chat, etc.
  - Develop and monitor a timeline for uploading, scoring, and submitting scores
  - Develop a quality assurance plan that involves teachers, provides automatic second scoring of failing and borderline scores, and includes sufficient external audit for technical quality (e.g., scoring consistency)
  - Implement scoring and moderation in a timely manner to provide rapid results to students and teachers, to influence instruction
- Build on the Phase I processes to create statewide **professional development**
  - Identify teachers from each school to participate in each of the aspects of assessment from task design through scoring and standard setting—create a “Professional Development University” model to support this
  - Following audit scoring, provide feedback to teachers on reliability and fidelity of their scoring
  - Develop a web-based platform to deliver professional development individually, support local trainer experts in local professional development meetings, and provide guidance for local scoring sessions, including access to scoring directors and supervisors who can view the same task evidence remotely
- Develop and implement an ongoing **validation research** agenda that includes projects described in the plan for Phase II of the OPAPP

## Requires a Plan and an Experienced Organization

It is critical to engage the people in schools and classrooms who will ultimately determine the success of the system. To achieve this requires launching a comprehensive plan to educate teachers and administrators about the program prior to their involvement (Maxwell, 2007). However, this puts an additional time burden on an already full schedule of activities necessary for administration and timely scoring of the performance tasks, as well as the development of the comprehensive technology platform that will support professional development, assessment development, deployment, scoring, and moderation.

An experienced assessment organization (vendor) can assist with this burden. OPAPP's evolution from pilot program to trusted element of Ohio's assessment system will require an organization with expertise in deploying complex, large-scale performance assessment programs and technology platforms. The organization should also be facile at managing relationships between multiple vendors, stakeholders, and agencies. The scope of this project and the expertise required to deploy it requires the involvement and cooperation of the SRN, vendors, ODE, LEAs, and stakeholders across Ohio. Experience with such diverse working relationships will facilitate the successful and streamlined deployment of a performance assessment system in Ohio.

This management responsibility will increase as the program expands and as the technological requirements for the system change.

Eventually, the performance assessment system could include extended projects and tasks that students complete online. To provide more information about students' content mastery, data can be collected about the processes and the steps they complete as they develop a final task product. Such a system could also provide support and scaffolding to students as they self-pace through their tasks, creating an assessment that immediately impacts learning.

### Administrator and Teacher Buy-In

Successful program implementation with beneficial learning outcomes for students may be significantly impacted by administrator and teacher confidence and commitment (Black, Harrison, Lee, Marshall, & Wiliam, 2002, Black and Wiliam, 2001).

Embedded performance-based assessments, in which teachers have a stake in the development, deployment, and scoring, might engage teachers in the assessment system and validates their expertise in content and as those who know their students best. External quality assurance efforts can also involve teachers and administrators; further sanctioning their professional expertise and providing opportunities for professional development while ensuring appropriate monitoring and technical quality (Maxwell, 2007).

**It does not work if imposed from the outside.** Infusing this level of local involvement in the accountability system into the classroom, schools, and districts, is not likely to

succeed when imposed on teachers from the outside (Maxwell, 2007). To achieve the full benefits of an embedded performance assessment system, teachers and administrators need to fully understand the program, recognize its impact in areas where it has already been used, and advocate for the changes such a program can provoke.

Ohio's performance assessments comprise one part of a comprehensive balanced assessment system. They provide a picture of student achievement that is both different and complementary to those created by classroom work, summative classroom assessments, and standardized selected response format tests.

**Performance tasks enhance insight for students and teachers.** In performance tasks like those developed by the OPAPP, students may have extended opportunities to engage with content, pose questions, and complete investigations that further their content knowledge and broaden their skills. Ideally, as students and teachers monitor student work on the performance task products, students will self-regulate their learning by identifying areas of understanding and confusion, and teachers will, similarly, gain insight into students' mastery of curricular and task objectives and thus adjust their teaching and guide students' inquiries accordingly (Black and Wiliam, 2001).

In a system where the objectives of learning, as defined in the performance outcomes, and the criteria against which learning will be measured (as defined in the common scoring rubrics) are clear to both teachers and students, aligned to state, national, and international curriculum standards, and designed to provide feedback in a timely manner, teaching and learning may heighten through deeper and more individualized understanding of content that students can transfer to other classroom or assessment situations.

**These new approaches have worked.** The list of "ands" in the previous statement was long. Aligning and implementing so many components in the context of a diverse state, an existing curriculum and assessment system, and through the bureaucracies of state and local education agencies, is a complex task. Yet, research suggests that in the instances when such programs have been implemented, from school level adoptions (Black and Wiliam, 1998, Black, Harrison, Lee, Marshall, and Wiliam, 2002) to national assessment programs (Maxwell, 2007, Stanley, MacCann, Gardner, Reynolds, and Wild, 2009), learning outcomes for students have been positive, with sufficient technical quality for the assessments to be part of a high-stakes system.

Perhaps the most significant factor in successful implementation of a performance assessment system able to provide information useful for both formative and summative interpretations, and one not mentioned in the previous list of "ands," is the engagement of the teachers and administrators in the process and its goals.

The elective engagement of teachers and administrators in a process for which they also have ownership of some of the development and scoring may provide validation of the teachers as professionals and, as shown in programs in Australia, can result in a more

successful system (ACACA, 1999). The initial public relations campaign also serves as initial training about the system.

The effort should include presentations at state and local conferences, dedicated training sessions, and testimonials from researchers and practitioners, all presented in-person across Ohio. At the same time, a publicly accessible website should contain written information, video clips of testimonials, video of training sessions and presentations, and contact information for leaders from the pilot sites. This concurrent effort is intended to show the State's engagement in the project and determination for its implementation, but also allow individuals to learn more and become engaged in the ideas on their own, becoming advocates for implementing the program in their schools as they learn about the potential benefits.

### Research Supports Training Approach

When statewide implementation is more imminent, Ohio and its assessment organization (vendor) should schedule systematic local training sessions, each led by a "local leader" who will also serve as a resource for local teachers and administrators throughout the program, such as the "coaches" used in Phase I. These initial in-person sessions should address the purpose of the program, performance objectives, performance tasks, and integrating the system into classroom work and schedules. In and of itself, such training on the objectives of curricular content has been shown to beneficially change instructional practices (Shavelson, Young, Ayala, Brandon, Furtak, Ruiz-Primo, Tomita and Yin, 2008).

Teachers often report feeling consumed by getting through curricular content without being able to focus on the intended learning outcomes. By having objectives clearly defined for themselves and students, and having rubrics that communicate performance expectations, teachers may concentrate on the process of learning and information that students retain (Shavelson, Young, Ayala, Brandon, Furtak, Ruiz-Primo, Tomita, and Yin, 2008), while research suggests students may self-regulate their learning and become more motivated to reach goals which have been clearly communicated (Yin, Shavelson, Ayala, Ruiz-Primo, Brandon, Furtak, Tomita and Young, 2008).

Brief training in the scoring procedures and the technology platform should be included so these topics are familiar. However, since they are not crucial at the outset of the program and can be supplemented by additional and web-based training, detail can come later.

Comprehensive training on the performance outcomes should be part of yearly professional development activities; ideally discussions from these sessions should inform the refinement of the objectives over time so Ohio teachers remain continually involved in creating and deepening their understanding of the content areas.

## Scoring as an Element of Professional Development

Additional professional development occurs by the teachers' participation in scoring. In training for scoring, participating teachers are provided examples of work ranging from very low to very high quality. In discussion with their peers, they come to consensus understanding on the components of the work products that represent content mastery and deep learning. This may deepen teachers' understanding of the content, the learning objectives, and possible methods for instruction and demonstration of learning (Darling-Hammond, McCloskey, 2008).

To verify the consistent application of standards across Ohio, all teachers would participate in scoring training and scorer qualification. Initial training should occur in a setting, either in-person or online, in which teachers have the opportunity to ask questions and discuss procedures and rubrics with their peers. Additional training can be deployed remotely, using web-based videos and interactive modules. Such methods have been successful in Pearson's distributed scoring model and in Pearson's scoring of alternate assessment performance tasks in several states.

Pearson provided award-winning online training for teachers involved in the Texas Alternate Assessment (discussed in Appendix A). We discuss in other sections the training mechanisms for scoring and the organization of scoring procedures. For now, we want to highlight the benefits of using technology to support scorer training and scoring, as well as the necessity for in-person scoring discussions to develop teachers' deep understanding of content standards, learning objectives, skills, and their application (Darling-Hammond, McCloskey, 2008).

Finally, we recommend professional development in the use of the scores generated during operational assessment and the information gained about student knowledge from the assessments. The greatest impact on the student learning will occur if scores are reported in the context of a system in which they can be interpreted and used to make timely instructional changes (Nichols, Meyers, and Burling, 2009). Scores must integrate into Ohio's data systems and the scores must have meaning when examined in conjunction with other data points.

## Training for the Technology Platform

Training requirements for the technology platform will depend, on some extent, on the complexity of the platform developed and the technology skills of people in the field.

No matter how intuitive and user-friendly the platform's design, systems, and interface, some training will be necessary. Technology skills may differ widely across the state. We recommend centralized in-person training of a local expert from each district or region—essentially a train-the-trainer approach. For regions with high technology needs, the in-person training can be requested for all teachers, or a sub-set of teachers from each school. All of this in-person training should be supplemented by comprehensive, user-accessible web-based tutorials and on-demand “help” features.

As the scoring and the technology platform are inherently connected, some scoring training will involve how to use the platform. However, the focus of the scoring training should be on content, learning objectives, rubrics, and scoring processes, so separate training to encourage teachers' facility with their system should be provided in advance.

## Summary

Developing and deploying a performance-based assessment system will best succeed if the system is not perceived as imposed upon from the outside. Rather, success will be significantly cultivated through teachers' engagement in the process and the implementation. Additionally, if teachers and administrators see the new system as an additional burden—instead of a means to complement their instruction and provide timely and meaningful information to benefit their instruction—it will be far more challenging to implement.

Success also will depend on high levels of teacher time commitment. This, too, should not go on top of teachers' existing responsibilities. It is likely to require state-level commitment to support scheduling time for teachers to participate in all aspects of the performance system. Implementing performance assessment is not an education reform panacea, but only one part of a system for improving student learning. Its strength lies in its engaging teachers and students in the process and the products, by providing them a share of ownership and responsibility, and validating the professionalism of the teaching force while supporting their depth of knowledge and instructional practices.



### 3. Recommendations Related to Development of Operational Performance Tasks

The involvement of teachers in the process of creating tasks is important for the professional development opportunities it may provide them. Involving teachers may improve assessment literacy and stakeholder buy-in, and help build important bridges between the work of the classroom and the work of the assessment system designed to measure instructional effectiveness. Research suggests that the involvement of teachers in developing and scoring performance tasks may help secure local understanding and an ownership stake in how the new assessment system and the performance tasks are developed, built, and validated, and scored (Black, Harrison, Lee, Marshall, & Wiliam, 2002, Black and Wiliam, 2001).

Teachers could be involved in following components, many of which have seen teacher participation during Phase I of the OPAPP:

- **Task development**—Training educators on the criteria for high quality tasks and involving educators in developing tasks and rubrics
- **Content review, bias and sensitivity review, data review, etc.**—Training educators on and facilitating their participation in review cycles, intended to develop in educators a full understanding of the development process, how tasks are aligned to standards, and the rigor associated with screening and placing tasks on assessments
- **Field test training and scoring**—Facilitating teacher involvement in field testing and evaluating the performance of tasks
- **Rangefinding**—Convening groups of teachers to set scoring standards and to score, discuss, and select papers to be used in scorer training
- **Scoring**—Teachers scoring the tasks locally, but with common training standards and materials to promote consistency. Teachers score completed tasks based on their training in the performance outcomes and rubrics. Additionally, they have an opportunity to gain an understanding of how close their students are to mastering the performance outcomes—and, equally important—where the student falls short, by evaluating their own students work with their expert content knowledge. Having this information about their own students in a timely manner can facilitate designing appropriate and targeted classroom instruction
- **Standard Setting**—Involving teachers in identifying levels of student mastery of the learning outcomes and associated performances of those outcomes in the performance tasks

This section describes recommendations for addressing each of those components in the broad scale implementation of the OPAPP.

Implementing the OPAPP statewide in Phase II will rely heavily on the pool of Ohio teachers and administrators with the hands-on experience of working with the SRN team in implementing Phase I of the Ohio Performance Tasks Assessment Pilot. To develop a sufficiently large bank of performance tasks to support an ongoing operational program, it will be critical to draw from this pool of experience.

## Training Teachers to Develop Performance Tasks

Teachers with experience gained in participating in the pilot can be trained and certified to serve as group leaders. These leaders will, in turn, train teams of teachers on the standards and task development, essentially a train-the-trainer approach. Trainers will learn the process for creating appropriate and meaningful tasks, and criteria for acceptance and usability on assessments. For this activity to serve as a comprehensive professional development exercise, it should take place state-wide, with representation from each region of Ohio, if not every district. Representatives will serve in a trainer capacity when they return to their region or district.

Centralized and district-level training sessions for teachers should be complemented with online training and certification to promote consistency and make sure that all trainees have the benefit of the same framework and are held to the same standards for quality. Online training offers Ohio considerable advantages, including identical training content for all trainees; low barriers to participation (“anywhere training”); and ease of access to training materials during the actual task development process for reference purposes. Educators will also be involved in building the online training, working collaboratively with task-development experts to incorporate best practices from the standard instructor-led model into the online components.

Once teachers have been trained and certified as content developers, they can work from their homes or offices, in a distributed model, to develop tasks. The tasks can then be sent electronically to a review committee for feedback and edits. This process continues until the committee deems the tasks acceptable, and will monitor the percentage of tasks accepted per content developer as well as monitoring the size of the overall pool.

## Associating Performance Rubrics to Tasks

As a part of task development training, the teachers will require training on the rubrics for a particular content. Rubrics within a content area are generic and can be applied to all tasks. As teachers develop tasks, they need to apply the rubric to the task to be sure it is applicable. They must possess an understanding of the rubric as they are developing the tasks. They need to think in terms of feedback to the teachers and students as they develop the tasks.

Pearson is participating in a research project on mathematics learning trajectories whose model involves associating comment codes with particular tasks. One set of comment codes is generic and pertains to a dimension, while another set of codes is specific to the task. The codes give insight into the students’ rationale and help to identify weaknesses or gaps. The codes also force the scorer to connect the work back to the outcomes

reflected in the rubrics. Comment codes in combination with scores give the classroom teacher information for future instructional planning.

The *Possible Evidence of Performance by Level on the Mathematics Rubrics* developed by the Stanford University SRN and ODE appear similar to the concept of providing comment codes by dimensions and, as such, would guide teacher rubric application training.

## Implementing Full-Scale Scoring

Moving from a pilot exercise to a full-scale scoring operation for Ohio's new high school assessment system will require heavy reliance on the expertise already developed among Ohio teachers—as well as reliance upon an experienced provider of scoring services.

Based on our extensive experience, scoring activities should take place along the following lines, with the scoring vendor leading the activities:

- A rangefinding committee, selected by the Ohio Department of Education (ODE), comprised of educators, and led by the vendor, will convene to select anchor (exemplar) student responses and training samples used to train teachers to score consistently across the state.
- The scoring leaders will incorporate the student responses into online training modules.
- Following a similar model as the one suggested for the tasks development, teachers involved in the task development, as well as scoring experts, will meet centrally to learn to consistently apply the rubrics to student responses from the pilot and any additional field testing of new tasks.

## Training the Teachers in Scoring the Assessment

Scorer training also will serve two purposes—professional development, and training teachers to score the assessments in a consistent and unbiased way.

The professional development component will include the links to curriculum and outcomes that help teachers better understand how their students perform in their classroom and how they and other students perform across their state. Professional development training can be modeled after the process used to train teachers to write tasks, with state-wide train-the trainer sessions.

Once the central training is completed, teachers will return to the districts and conduct training sessions for their fellow teachers, following the model at the central level. Online training can follow these local professional development events. Online training modules will provide teachers with a reference source or training refresher during scoring. Teachers can receive their scoring certification online by going through a qualification process. This certification should be a requirement for scoring student tasks, and will add to the consistency and validity of the scoring.

To provide Ohio with reliable results for its balanced assessment system, the scoring leaders will embed consensus-scored responses as a validity check for scorers. These validity responses should be selected prior to scoring.

### **Distributed Scoring Model and Consistency in Scoring**

To provide Ohio with results quickly, teachers will score the tasks locally from their homes or offices. Pearson's distributed scoring system allows teachers to access student work securely and remotely, submit scores, monitor their own performance, and receive coaching and support from scoring experts. Computer requirements are minimal, because the system was specifically designed to feature low barriers to participation. This model promotes wide teacher participation, rapid turnaround of results, and rapid feedback for their classrooms.

To facilitate more accurate and consistent scoring, the scoring leaders can embed consensus-scored responses as a validity check for scorers. These validity responses should be selected prior to scoring.

This sample scoring not only allows for monitoring the consistency in scoring across the state, but serves as a reliability check on classroom teachers. It will be monitored centrally with scoring leaders giving feedback to scorers, as well as calibrating them when necessary and checking the validity and reliability statistics. From this sample, a percentage of the student responses will be pulled and scored by professional scorers as an audit of the scoring process. Schools, districts, and the state will receive feedback from each of these scorings; helping teachers understand how their practices compare to other teachers' scoring.

Online training will be released to teachers as tasks are released. This will allow teachers the opportunity to practice scoring consistently with performance tasks throughout the year.

## **Summary**

Performance tasks may provide rich opportunities for students to demonstrate how much they know and can do. Scorers evaluating the completed tasks reward students for correct work and can gain an understanding of how close the student is to mastering the performance outcomes—a powerful tool for designing appropriate classroom instruction as well as for providing rich professional development for teachers. Through our scoring of alternate assessments, constructed response assessments in writing and math as well as our scoring of performance tasks for teachers seeking national certification from the National Board for Professional Teaching Standards (NBPTS), Pearson has extensive experience in scoring performance tasks. We can help Ohio with implementation of this critical element of its new balanced assessment system.

## 4. Recommendations Relating to Scoring, Reliability, Validity, and Comparability

Our recommendations related to scoring, reliability, validity, and comparability are intended to help Ohio with critical decisions regarding the incorporation of performance tasks into its new high school assessment system. In this section, we discuss the trade-offs that come with capturing work artifacts digitally, steps to enhance reliability and validity, and relevant considerations for a system of formative assessment and instruction.

Performance assessment puts the teacher at the heart of the assessment process as classroom teachers provide the judgments of student learning. To facilitate quality assurance efforts and provide evidence of student achievement, the assessment system must capture performance assessment responses and teachers' judgments quickly and accurately.

The expansion of locally administered and scored performance assessments raises a number of complex and sometime contradictory issues. These issues include the validity and reliability of performance assessment scores, the need to keep personal information private, huge increases in the amount of data to be organized and understood, and difficulty in preserving data for long-term use. Developing the policies, standards, and infrastructure needed to safeguard the integrity, accessibility, and stewardship of performance assessment data is a critically important task.

### Enhancing Reliability and Validity

Ohio is pursuing the creation of a balanced assessment system that may include objectively scored end-of-course (EOC) examinations as well as state-developed, locally administered and scored performance assessments. Student performance on the EOC examinations and the performance assessments may be combined to contribute to a total EOC score that would contribute to the graduation composite.

For some EOC subject areas, student performance on the EOC examinations and the performance assessments may be used for accountability purposes under No Child Left Behind. To be included in an accountability system, this balanced assessment system envisioned by Ohio must have evidence of validity and reliability. In this section, we will address the collection and evaluation of evidence for the reliability and validity of this balanced assessment system. Though conceptually and methodologically intertwined, we will address validity and reliability separately.

## Validity

The Ohio balanced assessment system is intended to accomplish goals beyond the goals typical of a conventional assessment. The Ohio balanced assessment system has the following three goals:

- Measure individual achievement in core knowledge as well as critical thinking skills
- Provide formative information useful to teaching
- Develop in teachers a deep understanding of learning standards and high-quality work

Because of these more ambitious goals, validity for the Ohio balanced assessment system is broader than conventional validity theory for educational tests. In this section, we will consider evidence and rationales supporting validity for each of the goals of the Ohio balanced assessment system.

### Measure Individual Achievement

Conventional validity theory is well placed to address how test design and scoring affects the validity of interpretations of students' core knowledge and critical thinking skills. The Ohio balanced assessment system includes the following two measures of student achievement:

- The EOC examinations
- The performance assessments

Each approach emphasizes different features of assessment, and this is reflected in the evidence and rationales used to support the two approaches.

The EOC examinations emphasize assessment of core knowledge and objective scoring. Though all types of validity evidence may be used to support test score inferences and challenging alternative interpretations for EOC scores, perhaps the most applicable are content relevance and representativeness and convergent and discriminant correlations.

Evidence of content relevance and representativeness might take the form of alignment between the content addressed by the EOC assessments and the knowledge, skills, and abilities taught in the class.

Evidence of convergent and discriminant correlations might take the form of correlations among different EOC assessments, performance assessments, classroom assessment scores, and even college entrance examination scores.

The performance assessments are intended to emphasize the assessment of critical thinking skills and scoring by teachers of their own students' work. The greatest threat to validity for performance assessment that includes the assessment of critical thinking skills

may be construct underrepresentation—for example, the situation in which the assessment is too narrow and fails to include important aspects of a construct.

One fruitful source of validity evidence applicable to this concern involves process models and process engagement that would indicate the degree to which students are analyzing, synthesizing, evaluating, applying knowledge, comprehending and evaluating data, formulating questions, and being innovative and creating new solutions. Evidence of process models and process engagement often takes the form of interviews with students immediately after attempting performance tasks, or recording think-alouds by students while attempting performance tasks (i.e., cognitive labs).

An additional threat to validity for performance assessment in which teachers score their own students' work comes in the form of construct-irrelevant variance, in which evaluations of student performance are contaminated by factors irrelevant to the focal constructs.

Perhaps the validity evidence most applicable to ruling out either positive or negative bias from teachers' scoring their own students would come from statistical analyses of performance assessment scores designed to monitor or correct local scoring biases. Assessment practices in other countries offer several models for moderation.

Under a statistical moderation model like that practiced in New South Wales, Australia, the EOC examination scores may be used as a criterion to adjust for performance assessment scores. In such a model, the idea would be to adjust performance assessment scores within a school using scores on the more "objective" EOC tests.

The mechanics of statistical moderation are relatively straightforward, and for example, can be expressed mathematically in sample notation as follows:

$$Y = \bar{Y} + \frac{S_Y}{S_X} (X - \bar{X}),$$

where

- $X$  is the EOC assessment scores,
- $Y$  is the moderated performance assessment scores,
- $\bar{X}$  is the EOC assessment mean score for the school,
- $\bar{Y}$  is the performance assessment mean score for the school,
- $S_X$  is the standard deviation of the EOC assessments for the school,
- $S_Y$  is the standard deviation of the criterion scores for the school.

From the standpoint of psychometric methodologies favored in the U.S., there are a number of concerns with such applications of statistical moderation. Obviously, the quality of the moderation is affected by the relationship between the criterion and moderating variable. If the relationship is very high, one would wonder why the more elaborate performance assessments are needed. If the relationship is not very high, there is little justification for using the EOC to moderate the performance assessment scores. Furthermore, the linear adjustments could advantage or disadvantage students with

extreme performances depending on the relative skewness of the moderating and criterion score distributions. Finally, the moderation approach can restrict individual student performance on the criterion variable (e.g., the performance assessments) depending on the overall school's performance on the moderating variable (e.g., the EOC test).

Thus, while statistical moderation is used to adjust reported scores in some assessments systems (e.g., some assessments in Australia and the United Kingdom), for the Ohio Performance Assessments it may be preferable to use the statistical methodology to identify schools or districts where performance assessment results seem unusual enough to trigger further scrutiny.

Social moderation is a second quality assurance strategy that can be used with performance assessments. Under a social moderation model like that practiced in Queensland, Australia, teachers from another school might re-evaluate a sample of student performances, and high levels of disagreement would suggest further investigation rather than automatic score adjustments

A Pearson company, EdExcel, offers a variety of qualifications resulting in both academic and vocational credentials in the United Kingdom. Appendix B of this paper provides a brief description of EdExcel's examination programs and their approaches to moderation and quality assurance.

### **Provide Formative Information**

Conventional validity theory is poorly suited to address claims that performance assessments provide formative information useful to teaching, as it privileges test score interpretation over test score use.

To be applicable to formative assessments, validity theory must take into account the claim implicit in formative use of assessment scores that the information derived from students' assessment performance can be used to improve student achievement.

How that information is used, and not what the assessment tells us about current achievement, will affect future achievement. For an assessment use to be appropriately labeled "formative," both empirical evidence and reasoned arguments must be offered to support the claim that improvements in student achievement can be linked to the use of assessment information by a teacher or the learners themselves.

Evidence and arguments that support the claim that assessment use provides formative information useful to teaching and learning requires causally linking information from performance on a particular assessment to the selection of instructional actions that, when implemented, lead to gains in student learning.

The emphasis on test score use to improve student achievement broadens the focus of validity investigation to include the coordinated assessment and instruction system within

which test score information is employed. For assessment information to serve a formative purpose, each component of the system must function effectively.

### **Develop Teachers' Understanding**

Conventional validity theory is completely unsuited to address the claim that the implementation of the assessment system develops in teachers a deep understanding of learning standards and high-quality work.

Not even consequential evidence of test score use, which is concerned with the interpretations of subject performance as the basis for actions, can accommodate such a far-reaching claim. But one can still muster evidence and arguments to support the claim that implementing the assessment system develops in teachers a deep understanding of learning standards and high-quality work.

Perhaps the validity evidence most applicable to this far-reaching claim would examine how the processes of scoring and quality assurance impact teachers' instructional practices and their understanding of students' learning.

One might argue that the processes of scoring and moderation are the mechanisms that help teachers in the following ways:

- They become more knowledgeable about the standards
- They expand their knowledge about how to teach the standards
- They calibrate their sense of the standards to common benchmarks

Investigating such claims should be part of a comprehensive validity research agenda.

### **Reliability**

Just as the more ambitious goals for the Ohio balanced assessment system require a broader conception of validity, these same more ambitious goals require a broader conception of reliability than conventional reliability theory.

In the following discussion, we define reliability as a measure of consistency in outcomes over replications of a procedure (Brennan, 2001). In this section, we will consider how to assess reliability for outcomes under each of the goals of the Ohio balanced assessment system.

### **Measure Individual Achievement**

The EOC examinations are intended to emphasize the assessment of core knowledge and objective scoring. Conventional reliability theory is well suited to estimating the reliability of scores from the EOC examinations.

Ideally, one can estimate the reliability of the EOC examination scores using a design in which forms are administered across different occasions to students. But such a study may not be practical due to the time and expense required.

Alternatively, one can estimate the reliability of the EOC examination using an internal consistency design, such as coefficient alpha, in which a single EOC examination form is administered only once.

A single form design using coefficient alpha will capture some, but not all, of the sources of error influencing student performance. For example, a single form design will capture the moment-to-moment variance associated with student scores, as with a loss of concentration, and the moment-to-moment variance associated with the environment, as with a sudden loud noise.

However, the performance assessments are intended to emphasize the assessment of critical thinking skills and scoring by teachers of their own students' work. This represents additional sources of error that should be considered in the assessments' reliability.

Generalizability theory is a second statistical framework used to analyze reliability. In contrast to classical test theory's conceptualization of error as a unitary entity, generalizability theory is concerned with identifying and estimating the contribution of multiple sources of error (or facets). Facets are sets of homogeneous conditions over which the measurement activity attempts to generalize.

For performance assessments that are scored by teachers, generalizability theory allows for the estimation of effects due to different teachers as well as the moment-to-moment variance captured by the use of coefficient alpha under conventional reliability theory.

### **Provide Formative Information**

In contrast to the focus on scores from a single assessment under conventional reliability theory, reliability theory for formative purposes must focus on the outcome of the system for learning—that is, student learning. Recall that the claim for the formative use of assessment is that to improve student achievement, one can use the information derived from students' assessment performance. Reliability for this system of coordinated assessment and instruction involves quantifying the consistencies and inconsistencies in producing positive student outcomes.

### **Develop Teachers' Understanding**

Similar to the reliability for formative purposes, reliability for an assessment system intended as a mechanism for teacher professional development must focus on the outcome of the system, i.e., teachers' understanding of learning standards and high-quality work. Again, reliability for this system involves quantifying the consistencies and inconsistencies in producing positive professional development.

## **Summary**

Our recommendations related to scoring, reliability, validity, and comparability are intended to help Ohio with critical decisions regarding its move to a new high school assessment system. Expansion of locally administered and scored performance

assessments raises a number of issues concerning validity and reliability of performance assessment scores, and in organizing and understanding the data yielded by the assessments. Developing policies, standards, and infrastructure needed to safeguard the integrity, accessibility, and stewardship of performance assessment data is critically important. In developing a balanced assessment system, it will be important for Ohio to work with its partners to focus on successfully addressing these issues.



## 5. Recommendations for a Web-based Performance Assessment Management System

### System Context

Supporting the performance assessments that Ohio envisions will require an easy to use, reliable online platform to allow for a broader inclusion of performance assessment tasks throughout the school year. Specifically, what is needed is a configurable Web-based Performance Management System.

The ideal solution will match the assigned performance tasks to the resources available to the student to complete the task.

To provide Ohio educators with the tools they need to provide a performance-based assessment, any web-based performance management system should support a sequence of events as follows:

- Develop performance tasks
- Assign them to students
- Capture and score the responses
- Assess student performance
- Evaluate effectiveness of the performance task

The system must be constructed to allow performance tasks to be assigned as a routine and regular part of classroom activity. The system must also allow the student to submit the work created and managed by the system for more rigorous scoring, potentially to be incorporated as part of a standardized end-of-course (EOC) examination.

The Web-based Performance Management System should provide interfaces for the following:

- **Teachers**, to enable them to accomplish the following:
  - Create performance assignment tasks and move them to a student in-box
  - Complete training on performance tasks using computer-based training
  - Assign scores to tasks they are eligible to score
  - Track number of assignments pending
- **Students**, to enable them to accomplish the following:
  - Access the individual in-box, for tracking assignments
  - Access links to websites containing performance tasks

- Upload assignments they have completed using common desktop software (Word, PowerPoint) and, if tasks require them, audio and video files
- “Tag” or associate a response with one (or more) performance assignments
- **External Systems**, to enable the following:
  - Associate performance tasks with a student’s multiple choice EOC exam responses
  - Link to student management systems for assignment of grades for completed tasks (assume most reporting will happen via existing student management systems)
  - Potential to interface with automated essay scoring technology

## Digitization of Responses

An online engine for performance scoring requires availability of digitized work products from students. A wide variety of methods exist to achieve this purpose.

In traditional high stakes testing, students use paper and pencil, and most test processing vendors capture images of the student work from answer documents that can then be routed for evaluation by scorers on local area networks or over the Internet.

To support the next generation of performance assessments, one must consider alternatives for digitizing work assignments, to enable evaluation by additional scorers other than the student’s classroom teacher.

The table that follows provides some alternatives for digitizing student responses.

	Advantage	Limitations
<b>Scanning</b>	<ul style="list-style-type: none"> <li>• Student can work with traditional classroom materials.</li> </ul>	<ul style="list-style-type: none"> <li>• May have limitations on color or paper size. <b>Mitigation:</b> Limit assignments to scannable paper sizes, such as 8.5x11.</li> <li>• Schools may not have sufficient scanning capacity for planned workload. <b>Mitigation:</b> Offer access to service bureau for scanning or piggyback on traditional assessment.</li> </ul>

	Advantage	Limitations
<b>Desktop Software</b> Microsoft Office OpenOffice.org	<ul style="list-style-type: none"> <li>Widely available</li> <li>Open Source and Commercial alternatives available</li> <li>Provides 21<sup>st</sup> century skills</li> </ul>	<ul style="list-style-type: none"> <li>Lab space may be limited. <b>Mitigation:</b> Match assignments to available lab space. Allow students to work from home or library. Add capacity to labs.</li> <li>Large number of alternatives makes faithful rendering at scorer workstations a challenge. <b>Mitigation:</b> Limit file formats supported. Require conversion to ISO Standard format such as PDF (ISO 32000). Convert at time of upload to standard format.</li> </ul>
<b>Video/Audio</b>	<ul style="list-style-type: none"> <li>Allows great flexibility how work products are created</li> <li>Large number of affordable video cameras without mechanical limitations of tape are available</li> <li>Provides 21<sup>st</sup> century skills</li> </ul>	<ul style="list-style-type: none"> <li>Works best for assignments that involve visuals, sounds, and actions that demonstrate learning by individuals or groups of students. <b>Mitigation:</b> Choose carefully the tasks presented in this format.</li> <li>Effective video creation requires quiet place, to avoid background noise that would materially affect ability to grade. <b>Mitigation:</b> Capture presentations one at a time. Allow students to work from home.</li> <li>Video scoring limits pace to real time for teacher scoring. Readers can read written work at their own pace (typically faster than if the document were read aloud). Scoring rates for audio and video tasks will be tied to the time it takes to view them. <b>Mitigation:</b> Specify maximum time for presentation as part of task instructions.</li> </ul>
<b>Web Application</b>	<ul style="list-style-type: none"> <li>Places least infrastructure demand upon schools and districts. Requires a web browser and Internet connection only at school.</li> <li>Avoids desktop licenses</li> <li>Simplifies administration chores</li> <li>Provides 21<sup>st</sup> century skills</li> </ul>	<ul style="list-style-type: none"> <li>Students may not be familiar with interfaces. <b>Mitigation:</b> Use tools for teaching students to use various applications</li> <li>Requires access to current web browsers and broadband Internet connections at schools.</li> </ul>

Table 1. Alternatives for Digitizing Work Assignments.

## Managing the Scoring Process

The principal advantage of scoring digitized materials is that the student's work can be made available rapidly to multiple scorers—simultaneously, if desired. To accomplish

this, the system provides a workflow configured to manage the scoring process for each student's response.

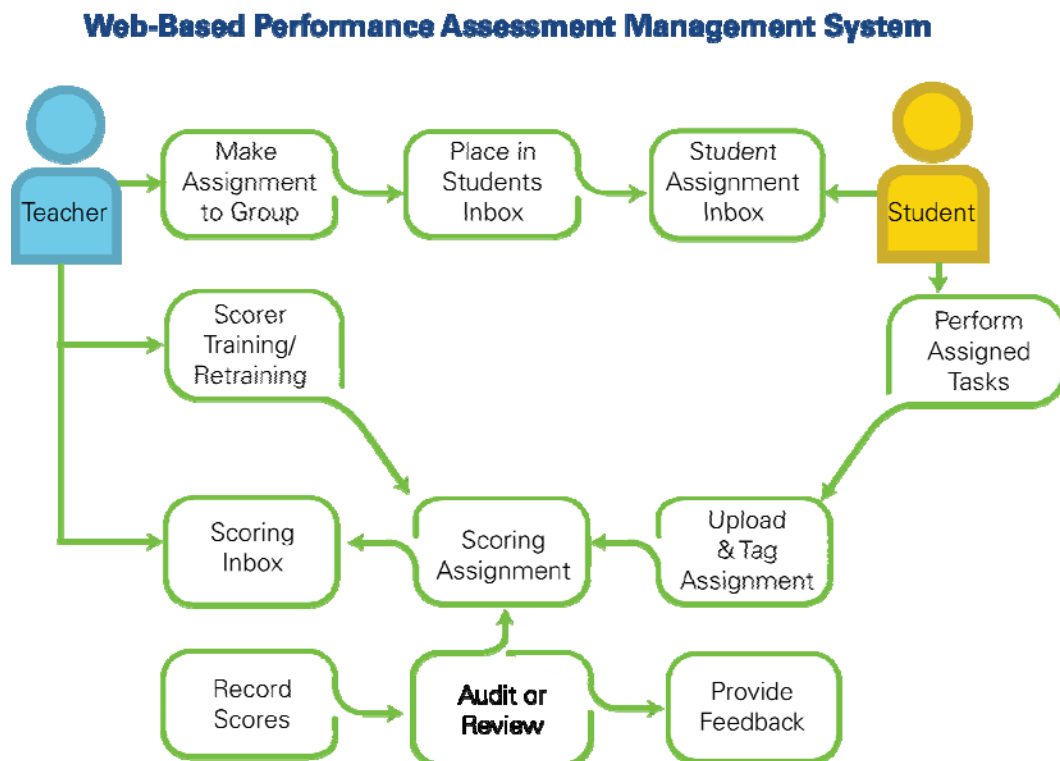
Management processes can be set up at the item level or workflows can be configured to require the classroom teacher to score his or her students' work. To check the consistent assignment of scores, the system can sample 20% of the responses at random from the classroom, and have a score assigned by another qualified scorer. Or, the workflow can be configured to obtain a second score for students on the borderline of a pass/fail threshold.

The designs for the workflow are limited primarily by the available scoring resources.

## Component Overview

Online assessment provides multiple advantages to Ohio educators. Therefore, all interfaces described in this section between people and systems are web based, unless otherwise noted. To provide security for Ohio's assessment system, all participants would be assigned individual user names and passwords.

Figure 2 below provides a visual overview of components for a Web-based Performance Assessment Management System.



**Figure 2. Overview of the Web-Based Performance Assessment Management System Components.**

## Teacher Interface

The system must provide a web-based interface for teachers to accomplish the following routine tasks associated with management of classroom assignments:

- **Develop, tailor, or extend a performance task matched to the student’s needs**—The system must provide an area for each teacher to maintain a set of “candidate” tasks that are under development. It may support an approval workflow for peers or administrators to comment or provide an approval for the content. The approval process may also include a process for confirming alignment of candidate tasks to state standards. Once completed, the teacher may elect to share completed tasks for use by other teachers.
- **Assign a previously developed task to students**—The system must allow teachers to make assignments to individual students. These assignments would join the list of assignments in the student’s inbox.
- **Track student completion of performance tasks**—As students complete work, their submitted work is associated with one or more of the inbox assignments. When the student has completed his or her work, the artifacts put into the system can be passed on for scoring. This interface could allow the student to upload draft or intermediate work products. The teacher would have the option to track, review, and provide comments on intermediate work products.
- **Provide a training interface**—The system should track that the teacher has completed the training necessary for scoring. Training provided would review the scoring rubric, provide exemplars for each score point, and provide the teacher an opportunity to mock score previously scored items to confirm understanding of the rubric.
- **Provide a scoring interface**—The system would maintain a queue of work for scoring, and would manage an interface that assigns work for review by the teacher. This interface would not divulge the identity of the student work at the time it was being evaluated. The work assigned to the teacher for scoring would be managed by the system in accordance with the established workflow for the item. The scoring interface collects scores, and optionally can permit the scorer to select comments from an approved list of comments consistent with the score point and the rubric.

## Student Interface

The system would provide a web-based interface for the student to:

- **Manage assignment inbox**—The system would provide the student with an inbox of assignments. Each assignment would contain all of the information necessary to complete the task. This would include a description of the task, and might include a structured form (for example, an outline in a Word document or PowerPoint template). The assignment could include links or references to materials (periodic table, etc.).

- **Manage work folders**—The system would provide the student with space in which to store tasks, including partially completed assignments. Students would have the ability to add, update, or remove items in their folders.
- **Submit assignments**—Once the student completed work on an assignment, he or she would associate one or more items from the work folder with an assignment. The student would then submit the work for scoring. A copy of the work at the time of assignment completion would be associated with the item. Once submitted, the defined workflow for the item would be invoked to manage the response through the scoring process.
- **Review scored assignments**—Once the scoring process was completed, the student would receive notification of the score assigned. The score might also include comments for the student to consider. The student and teacher would have access to the original work, scores, and any comments provided.

## Scoring Management Reporting

To verify that work is progressing as expected, any system of this magnitude will need to have a number of management reports, including the following:

- **Scorer training**—Provide reports to show teachers that have completed workshops and online training. These reports can be used to target populations requiring follow-up training.
- **Work in progress**—Reports to confirm all responses are receiving scores at the pace expected. These reports can be used to predict scoring completion timeframes.
- **Inter-Reader Agreement**—For all tasks that receive multiple scores, reports of inter-rater agreement can document the effectiveness of training and consistent application of rubrics across a pool of scorers.
- **Validity**—Another desirable system management option would be to have panels of content experts provide responses with known, agreed upon scores. These responses with known scores could be injected into the work assigned to scorers at random throughout the process. The scorer would not be able to distinguish these control responses from the other student responses when scoring. Reports would summarize comparisons between assigned scores and known scores for each scorer subjected to the validity responses.

## Student Results Reporting

The student results reporting function should include the following:

- **Student summary**—Reports summarizing work on all performance tasks can be assembled. These reports can contain a quantitative summary of scores assigned, as well as, narrative based upon rubric and assigned comments.
- **Teacher/Class Summaries**—Once scores are assigned; the data can be used to provide quantitative summaries of performance. Comparing teacher or classroom

performance to performance of peers within the school, district, and state can be used to identify relative strengths and weakness of the group.

## Interfaces to Other Systems

To provide Ohio with both utility and flexibility, the system should allow interfaces to other systems, as follows:

- Student Management Systems, including the following:
  - **Account management**—At the start of each semester/grading period, an interface with the student management system would allow the system to create or update student accounts with current class assignments. This would enable teachers to make assignments to students. This interface would also mark student accounts inactive for students no longer enrolled.
  - **Grade book**—Once scores are completed, they are periodically loaded to the grade book section of the student management system.

## Summary

Digitizing student responses provides Ohio with the ability to exploit a web-based performance assessment management system. Using this system, tasks can be tailored to individual students, students and teachers can manage and monitor their work, responses can be assigned to multiple scorers, reports can be generated, and interfaces with other systems can be developed. Such a system is achievable and will be an essential component of a balanced assessment system.



## 6. Summary and Looking Ahead

The purpose of this paper was to provide discussion and recommendations related to the operational implementation of performance assessments within Ohio's assessment system. In doing this, we focused on four major areas:

- Program design, management, training, and professional development
- Developing operational performance tasks, rubrics, and scoring protocols
- Establishing and maintaining reliability, validity, and score comparability
- Developing and maintaining an e-portfolio / performance assessment management system

From our considerations of these topics, we believe that a performance assessment system that contributes to Ohio's high school assessments is feasible.

The opportunities for performance assessments to become part of state assessment systems have become more salient because of the changes playing out through the Race-to-the-Top and Common Core assessment competitions. Although Ohio may not have anticipated this turn of events when the Ohio Performance Assessments project began, the progress and lessons learned from Phase 1 of the project dovetail with these assessment reform initiatives. Ohio and other states are organizing a consortium to apply for a competitive grant that can provide the resources needed to fully research and develop a performance assessment system such as has been described in this paper.

Despite these reform-based opportunities, there are obviously pitfalls as well. At a recent Board on Testing and Assessments workshop, Brian Gong provided some lessons learned from personal experiences with innovative performance-based assessments in Kentucky during the 1990s (Gong, 2009). His recollections included a laundry list of challenges, including the following:

- Construct definition and influence of scoring rubrics
- Creating tasks that validly elicit constructs
- Creating scoring that reliably and efficiently measure the construct
- Year-to-year equating
- Scorer accuracy and reliability
- Development and field testing of constructed response items
- Reporting and turn-around time
- Challenges related to stakeholder dissatisfaction
- Challenges related to politics and election battles

Clearly, in looking ahead for performance assessments, the lessons of the past must be considered. The Ohio Performance Assessments project stands the best chance of success as part of a balanced assessment system with components characterized by clearly defined constructs, purposes and uses. As a consortium organizes itself to move the project forward, future success will depend upon designing the assessment components and their intended uses in a coherent, efficient, and commonly understood fashion.

## 7. References

- Australasian Curriculum, Assessment and Certification Authorities (ACACA) 1999, *Principles for the integrity, quality and long-term credibility of certificates of achievement*, Queensland Board of Senior Secondary School Studies, Brisbane. <<http://www.qsa.qld.edu.au/publications/research/yrs11-12/assessment/lt-cred99.pdf>>
- Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5 (1), 7–73.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Testing, motivation and learning*. Cambridge: University of Cambridge, Faculty of Education: The Assessment Reform Group.
- Brennan, R.L. (2001) *Generalizability theory*, (2nd ed.) New York: Springer
- Darling-Hammond, L. and McCloskey, L., 2008. Assessment for learning around the world: What would it mean to be “internationally competitive?” *Phi Delta Kappan*, Vol. 90, No. 4 (December 2008): pp. 263-272.
- Furtak, E.M., Ruiz-Primo, M.A., Shemwell, J.T., Ayala, C.C., Brandon, P.R., Shavelson, R.J., and Yin, Y. (2008). On the fidelity of implementing embedded formative assessments and its relation to student learning. *Applied Measurement in Education*, 21(4), 360-389.
- Gong, B. (2009, December). *Innovative assessment in Kentucky’s KIRIS system: Political considerations*. Presentation at the workshop on “Best Practices in State Assessment”, sponsored by the National Academy of Science, Board on Testing and Assessment, Washington, D.C. Available at <http://www7.nationalacademies.org/bota/Brian%20Gong.pdf>.
- Maxwell, G. (2007). *Implications of proposed changes to senior secondary school syllabuses*. Queensland Studies Authority.
- Nichols, P., Meyers, J., and Burling, K. (2009). A framework for evaluating and planning assessments intended to improve student learning. *Educational Measurement Issues and Practice*, 28(3) 14-23.
- Shavelson, R.J., Young, D.B., Ayala, C.C., Brandon, P.R., Furtak, E.M., Ruiz-Primo, M.A., Tomita, M.K., and Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, 21(4), 295-314.
- Yin, Y., Shavelson, R.J., Ayala, C.C., Ruiz-Primo, M.A., Brandon, P.R., Furtak, E.M., Tomita, M.K., and Young, D.B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, 21(4), 335-359.
- Stanley, G., MacCann, R., Gardner, J., Reynolds, L., and Wild, I. (2009). *Review of teacher assessment: Evidence of what works best and issues for development*. Oxford University: Centre for Educational Assessment.



## Appendix A: Pearson Resources

Technology provides opportunity for Ohio to reach its bold goals in student assessment. Pearson has proven technology to power new modes of assessing student performance and help improve teaching and learning.

Together with our sister companies ([Edexcel](#), [Pearson Language Testing](#), and [Pearson VUE](#)), the Assessment & Information group of Pearson is the most innovative provider of assessment and education data management services in the world. Pearson annually administers more than 15 million online assessments in the United States and around the world.

Ohio seeks to replace its high school graduation tests with a three-part assessment system that includes a nationally standardized college entrance assessment in science, mathematics, and English language arts; end-of-course exams in science, mathematics, English language arts, and social studies; and a senior project. At the same time, Ohio wishes to explore new technologies, such as online administration and artificial intelligence scoring. Moreover, Stanford University has shown Ohio that other nations' assessment systems include intellectually ambitious performance tasks, in contrast to traditional multiple choice or short answer items from the United States.

Pearson has proven technology to power new modes of assessing student performance and help improve teaching and learning.

Pearson has technological resources to help Ohio respond to those challenges.

We can provide proven online test administration, through a single sign-on portal that offers convenient test management. Our portal includes a proven test application that students will find easy to use, and speedy scoring and reporting. (We say our system is “proven” because in the U.S. alone in 2009, we delivered 4.1 million tests online.) Through use of innovative items and computer adaptive testing, we can provide students with intellectually ambitious assessments that not only **measure** learning, but help **foster** learning. We offer both image-based and artificial intelligence scoring. And, we can configure a system to help educators manage the groundbreaking portfolio portion of Ohio’s new high school assessments.

We briefly describe, on the pages that follow, our PearsonAccess online portal, TestNav online testing application, innovative items, image-based scoring, artificial intelligence scoring, portfolio management system, and other Pearson resources.

## PearsonAccess

PearsonAccess provides an essential suite of tools for managing and coordinating test administration activities—from enrollment to test materials ordering, to giving students access to the test, to scoring and reporting. Fundamentally, PearsonAccess is our answer to the challenge of integrating all aspects of the assessment cycle instead of having disbursed and redundant systems. PearsonAccess provides a single interface that integrates paper and online testing and reporting, and makes programs more accessible and easier to use.

PearsonAccess.com

PEARSON

Visit our websites to learn more about our products:

**PEMSolutions™**  
The most comprehensive approach to assessment

**Perspective™**  
Enhancing reporting with web-based learning Resources

**PA Series®**  
Formative assessment for forecasting growth

**Everything Starts Here.**

CHOOSE YOUR STATE TO BEGIN:

Please Select ▼

Statewide testing requires the management of complex data sets and schedules with continual communication and seamless access—all key factors to success.

At Pearson, improving data management, communication, and access led to the creation of our new portal, [www.pearsonaccess.com](http://www.pearsonaccess.com). The portal is the single entry point to state testing products and services, including PEMSolutions™, Perspective™, PA Series®, and EDWARD™.

**TrueScores.com**  
Pointers and perspectives on psychometrics and test and measurement practices.

Copyright © 2008 Pearson Education, Inc. or its affiliate(s). All rights reserved. [Privacy Policy](#) [Terms of Use](#) [Rights and Permissions](#)

**Single Sign-On.** Ohio users will need only one login ID and password to access all data applications within PearsonAccess.

## Pearson’s Online Testing

Easy to use and easy to implement, our online testing system delivers a secure, reliable, and interactive testing experience on the same desktop computers already available in most schools. Our solution provides students a richer, more engaging testing experience, while test administrators benefit from an interface tailored to specific roles and responsibilities. Authorizations can be varyingly assigned to allow authorized administrators to update enrollment information, schedule tests, assign test forms, request additional materials and ancillary services, and view reports and data online.

Students access the TestNav™ delivery engine using a Web browser—no need to download special software. Our solution provides a secure testing environment with a robust suite of tools ranging from a notepad to a mathematical compass, and ancillary materials such as the Periodic Table of Elements or custom material of Ohio’s choosing. TestNav also supports audio, video, and interactive Flash-based items that parallel students’ classroom experience and the media-rich culture where they learn and play every day.

## Supporting Challenging Test Content

The computer presents tremendous opportunities to develop innovative assessment items that allow students to more fully demonstrate their abilities. These items enable students to show how they arrived at an answer, and allow scoring on how well they understand a particular process, rather than simply scoring their answers as only right or wrong.

Pearson has developed, delivered, and scored multiple intensive item simulations that allow students to interact with a range of experiments and interfaces. Examples include measuring viscosity of honey, charting the path of the sun and moon, and watching video on the life cycle patterns of a tortoise. All of these interactive simulations allow the user to run experiments, record the results, and then draw conclusions based on those results.

In addition to our own experience with innovative items, Pearson delivers partnered content. For example, in support of the National Assessment of Educational Progress (NAEP), Pearson has formed highly collaborative relationships with independent item development contractors, including Inspire Learning and TATA Interactive, to introduce the next generation of innovative items. Our success in producing, delivering, and scoring these assessments demonstrates the kind of productive working relationships Pearson supports and maintains throughout the testing industry.

Pearson's TestNav delivery platform supports the following innovative item types:

TestNav Innovative Item Capabilities	
Innovative Item Types	Supported by TestNav
<b>Multiple Selection</b>	Yes
<b>Gridded Response</b>	Yes
<b>Video Streaming</b> to enhance items	Yes
<b>Audio Streaming</b> to enhance items	Yes
<b>Click Response</b> for items to designate response location on screens <ul style="list-style-type: none"> <li>Click and drag response</li> <li>Multiple response clicks collected and stored</li> <li>Hot spots</li> </ul>	Yes
<b>Constructed Response</b> (essay, graphing, short phrase, single word, numerical response) <ul style="list-style-type: none"> <li>Delivered</li> <li>Captured</li> <li>Scored through artificial intelligence systems and/or human readers</li> </ul>	Yes
<b>Simulation Delivery</b> and response	Yes
<b>Simulation Response</b> varied based on user input	Yes
<b>Integration</b> of color, movement, zoom features into items, drawing capabilities	Yes

**An Interface to Meet Ohio Needs.** Pearson's TestNav online interface can a wide variety of item types, including multiple selection items, items with video and audio, click-and-drag items, and several variations of short- and extended-response items.

## Pearson's Automated Scoring Technology

If Ohio seeks to explore new technologies, such as artificial intelligence scoring, it pays to explore a proven solution. Pearson has experience in developing custom solutions based on our highly flexible technology and our Intelligent Essay Assessor (IEA).

Ohio can use IEA as a formative assessment tool or to enhance scoring for high-stakes assessments. IEA helps you quickly deliver accurate, valid assessments. With IEA, you select the prompts and provide the necessary number of human-scored samples to calibrate the system. We customize a web-based interface where students enter their responses to prompts and educators access reports.

IEA offers the following unique characteristics:

- Sophisticated algorithms that power the system
- Cross-subject testing for needed assessments
- Accurate results and scoring

- Data analysis to help evaluate program efficacy
- Plagiarism detection to protect test validity

Pearson’s automated essay scoring technology is based on Latent Semantic Analysis (LSA), a machine-learning method that acquires and represents knowledge about the meaning of words and documents by analyzing large bodies of natural text. LSA is all of the following:

- Theory of how people learn the meanings of words
- Mathematical system for computational modeling of thinking processes
- Text analysis tool

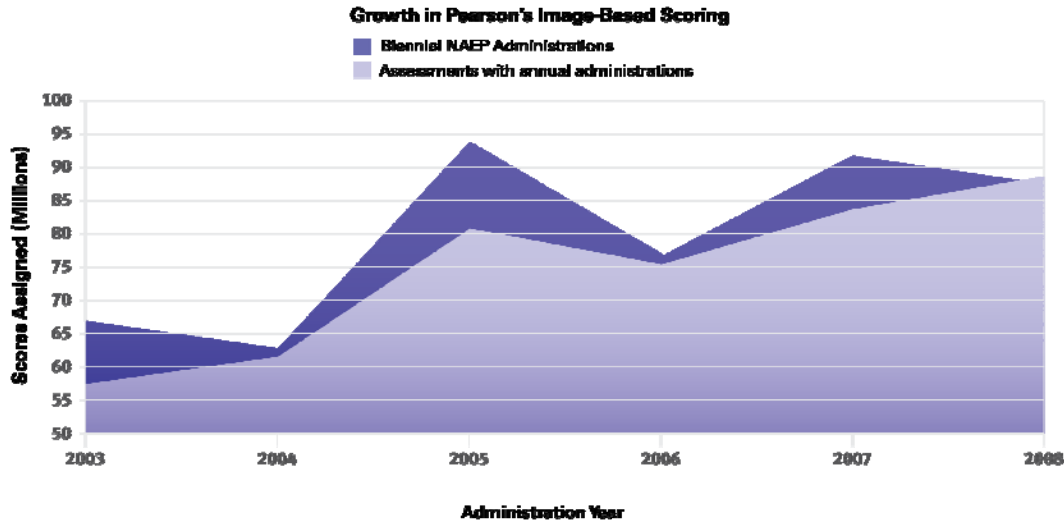
In addition to the underlying LSA technology, Pearson’s Knowledge Analysis Technologies™ (KAT) engine includes additional customized development and proprietary mathematical techniques to optimize LSA for specific automated scoring applications.

**KAT offers proven validity and reliability.** In tests over thousands of essays, Pearson’s automated scoring technology has proven as reliable as professional human scorers and more predictive of the average of two human scorers than the inter-rater reliability. This automated scoring technology also can be used to judge traits, such as writing applications and conventions with reliabilities equivalent to human graders.

Pearson’s automated scoring technology has proven as reliable as professional human scorers.

## Pearson Image-Based Scoring

To score its new high school assessments in time to provide teachers the data they need, Ohio will need a proven system with ample capacity. Pearson has the experience and capacity Ohio needs. Since its implementation in 2000, we have used our image-scoring to effectively manage scoring projects such major projects as the California Standardized Testing and Reporting (STAR) Program and the National Assessment of Educational Progress (NAEP).



**Growth in Image-Based Scoring Capacity.** Our scalable capacity allowed us to meet increased demand for performance scoring on state contracts as well as years with large National Assessment of Educational Progress (NAEP) administrations. In 2008, Pearson assigned 138.6 million scores for constructed-response items.

Pearson owns a series of U.S. and foreign patents for key components of image-based scoring. Our innovations in image-based scoring allow us to score state and national assessments of various types, content areas, administration cycles, and scale.

Ohio will benefit from our innovations in image-based scoring that allow us to score assessments of multiple types, content areas, administration cycles, and scale.

Performance scoring integrates multiple processes (routing work, scoring responses, monitoring quality, and tracking progress and workflow) into a single, efficient, user-friendly system. Electronic-format student responses are distributed to computer workstations, where we score them using qualified, trained scorers. We automatically route, to qualified personnel, responses requiring second scores or resolution reads. All scores assigned to student responses are automatically captured and available for review.

Performance scoring provides a number of advantages over other scoring systems and methods, as illustrated in the following figure:

Advantages of Pearson's Image-Based Scoring System	
Feature	Description
<b>Streamlined front-end processes</b>	Soon after student test booklets are scanned or students have entered their responses electronically, images of student responses are available at scorers' workstations. Electronic routing and tracking eliminates the costs associated with paper sorting, packetizing, and shipping and the possibility of losing any test documents.
<b>Increased security of test documents and student information</b>	Employees at scoring sites do not handle actual documents; scorers cannot print from their computers. We include no demographic information when routing responses to scoring personnel, preserving student privacy.
<b>Flexibility and focus through item-level scoring</b>	Since responses to different items from a single student are separated at the outset of the process, responses can be routed to the most appropriate group of scorers in a given location, allowing scorers to train and focus on one item at a time. Performance scoring also allows routing sets of items or linked items to groups of scorers. Scorers are not influenced by students' other responses.
<b>Dispersed evaluation</b>	By dispersing the various responses from any one student across a wider pool of scorers who work on separate items, performance scoring eliminates potential risks involved in having one scorer score all items from a given student, thereby removing the potential for a "halo effect" in scoring. A scorer's judgment of a student response on one item is not influenced by the quality of that student's other responses, as the scorer does not see a series of responses from the same student.
<b>Real-time quality monitoring</b>	Supervisory personnel at all levels and locations can monitor the progress of the project and the work of scorers through a series of online tools and real-time reports. Since reports are updated continuously and automatically, scoring staff can react quickly to any trends or issues that arise, rather than wait for monitors to be scanned or end-of-day reports to be printed and distributed.
<b>Automated backreading</b>	Our backreading tool allows our performance scoring staff to immediately search for responses scored by a particular scorer, responses receiving a particular score, or a combination of both. It also can easily retrieve responses where two scorers assigned scores with perfect agreement, adjacent agreement, or where the scores were non-adjacent. These multiple filters can be used to quickly pinpoint and correct individual scorer trends.
<b>Automated validity</b>	Pre-approved validity responses can be delivered online at a preset interval throughout scoring. This process is transparent to scorers, since validity responses are interspersed with "live" responses. The system automatically captures, calculates, and reports validity data at the item, reader, and room level.

**Image-based Scoring Advantages.** Image-based scoring provides advantages over paper-based scoring that result in more accurate scoring, reduced turnaround time, and increased cost efficiencies.

## Managing High School Projects

To provide Ohio with online tools for its performance-based assessment at the high school level, we can apply the experience we are gaining in observational assessments in alternate assessment and early childhood education.

Our work with portfolio-based alternative assessment enables Pearson to offer our integrated PearsonAccess online platform to help with managing the student project component of Ohio's new high school assessment.

Additionally, we are in the process of moving a number of early childhood observation-based assessments to the PearsonAccess online environment, including the Work Sampling System from Pearson Early Learning, a comprehensive framework for implementing observational assessment, and the Ounce Scale, used for measuring child progress.

Using our single sign-on system designed by usability specialists, we can develop a system in which Ohio high school teachers can access web-based applications that support paper-based and online testing.

Below we note several online score entry options Ohio may wish to consider for high school assessment administration—data entry and task management workflow. Both options, currently in use by existing Pearson customers, assume that teacher instruction for scoring the assessment occurs outside the online system.

**We can configure data entry.** Similar to the scannable, paper answer document, we can configure the online system for Ohio assessments to mirror paper answer documents and records. This option reduces the administrative burden of paper solutions. The data entry solution is configurable, so teachers see screens similar to paper documents.

**With the task management workflow,** teachers are free to conduct performance tasks as part of the high school assessment. At the teachers' convenience, they can enter scores into the secure, password-protected online system. With the workflow approach, teachers see an initial set of tasks, and based on the scores recorded, the system presents another set of tasks. We currently use this approach in Texas for its alternate assessment program.

**Use PearsonAccess to manage videos, images, and documents.** Our system offers capability to upload videos, images, and documents, should Ohio desire to digitize these artifacts. Ohio will know the status of score entry tasks because the online system includes status reports.

We can configure our system to enable teachers to accomplish the following:

- **Enter and upload scores** (for assessments/evaluations/observations completed locally), which could then be merged with other scores for a combined end-of-course or graduation assessment
- **Upload student work and artifacts** (these can be video files, images, etc.) for safekeeping and/or for scoring or second scores, or for auditing teachers through blind scoring
- **Create performance reports** that display the status of the score entry tasks (reports currently available include: students loaded into the system, completed students, status of the student in the workflow, and status of the collection of evidence)

Authorized users can also access district-, school-, and student-level reports via the online system. Online reporting provides a means for quickly accessing results at the users' convenience. Educators will have the ability to quickly view assessments results for students tested online or on paper through our online system.

Additional capabilities include data sorting and filtering and viewing data in PDF format. Reports and data files can be viewed via a published area within the portal, and saved onto the user's system, if desired.

## Video Training for Conducting Task-Based Alternate Assessment

**Showing teachers** and test administrators how to conduct task-based observational assessment **has more impact than just telling them.**

To coordinate, facilitate, and deliver yearly training for teachers and test administrators, we can assist with online, video-based training that will leave a lasting impact—at considerable economy to the ODE.

To provide material for training sessions, Pearson can go into select Ohio schools and districts to capture quality video of Ohio students performing OPAPP assessment tasks. Video provides authentic examples in a controlled environment, which allows us to illustrate specific nuances of the assessment. The ODE also can use video as part of practice activities that would allow administrators to rate actual Ohio students and get constructive feedback.

A variety of delivery mechanisms are possible, including video on DVD or online as a component of an interactive web-based training. The following graphic shows a possible interface for an interactive web-based training.



**Powerful Lessons from Video Training.** We can distribute video on DVD or, if the ODE desires, online as a component of an interactive web-based training. Here, we demonstrate a possible interface that could be used for interactive web-based training.

We are no strangers to such work and have collaborated with our state clients to deliver training that includes teacher-student pairs. Our online training modules developed in collaboration with the Texas Education Agency (TEA) earned first-place honors in the American Educational Research Association (AERA) Division H, School Evaluation and Program Development in Publications competition.

We have assumed costs for the development of high quality training videotapes of student teacher pairs for Ohio balanced assessment training.

### **A Powerful Training Tool Clarifying the Rating Process**

Capturing video of Ohio students actually conducting performance-based assessment objectives/tasks is a powerful way to help teachers gain experience with the assessment process that draws out the most common, unique, and challenging aspects of the rating process.

Capturing video of Ohio students actually conducting performance-based assessment objectives/tasks is a powerful way to help teachers gain experience with the assessment process.

Our video production team will work with the ODE to develop the video script and shots needed for the videotape training. We will work collaboratively to draw out the key concepts and messages critical to your audience. We have the equipment and the experience to videotape Ohio students as they perform tasks associated with the performance-based assessment

We can accommodate any level of involvement the ODE may desire during production of the training materials. Once we have completed the material, we can distribute it on DVD/CD, or provide the video for distribution on the state educational television system.

### **Pearson's In-House Video Capabilities**

Pearson can, if the ODE desires, construct an informational videotape (or DVD) for use by Ohio community and school staff. For our video needs, we generally use our in-house video production crews, which so far have recorded more than 30 educational projects.

Our in-house video production crews so far have recorded more than 30 educational projects.

For media duplication, we use a vendor in Omaha, Nebraska—Software Marketing Solutions (SMS). Impressed with this company's timely responses and quality products, Pearson has contracted with this preferred vendor for more than 10 years.

### **Summary of Video Capabilities and Experience**

- Our team of videographers has been shooting videos in school districts for over a decade.
- Our experience capturing video in special needs classrooms helps our team prepare before the shoot to capture the shots we need, while remaining flexible to adjust to conditions in the classroom to lessen the intrusion and fit the needs of students and teachers there.
- We bring the necessary professional equipment to capture high quality video for the training, but use strategies to help keep the classroom as comfortable for students as possible.
- We can create custom training interactions to meet the specific needs of Ohio content, but we also offer proven instructional interactions to enable us to bring Ohio training content online quickly.
- We can meet the challenges of delivering rich media in schools that have a wide range of network and computer system capabilities.
- If desired, Pearson can provide Ohio training content in a Section 508-compliant version, so your training content is available to the sight impaired if needed. Pearson can provide training with captions for hearing impaired users if needed. We can discuss the costs of this option.

## Summary

As Ohio contemplates changes to its assessment system to improve educational outcomes for students, the ODE needs to consider proven solutions from an experienced provider. Online assessment offers advantages and benefits to Ohio in terms of providing challenging assessment, generating results quickly to provide teachers with data in time to make a difference, and doing so economically. Pearson is prepared to work with the ODE to develop an online system that meets the needs of Ohio teachers and students.

## Appendix B: Overview of EdExcel

EdExcel, a Pearson company, is the UK's largest awarding body, offering academic and vocational qualifications and testing to schools, colleges, employers, and other places of learning, internationally as well as in the UK. The company serves as one of five main examination boards for England, Wales, and Northern Ireland. Internationally, EdExcel each year awards over 1.5 million certificates to students around the world.

Educational assessment in England includes National Curriculum assessments, a series of educational assessments used to assess the attainment of children attending maintained schools in England.

The National Curriculum exams are given in the following three key stages:

- Stage 1 is administered in the second year of schooling to students approximately seven years of age
- Stage 2 is administered to 11-year-olds in their sixth year of schooling
- Stage 3 is administered to 14-year-olds

The stage 1 tests are teacher-administered, the stage 2 and 3 tests are standardized National Curriculum Tests. Recently, the stage 3 tests have been abolished as a requirement, although teacher assessments are still required and the National Curriculum Tests are optionally available to schools. Currently, there is controversy over the National Curriculum Tests (also known as the SATs, or “Statutory Assessment Tests”), with critics claiming that their use is cruel to students, encourages teaching to the test, narrows the curriculum, and focuses disproportionate resources on borderline pupils.

Beyond age 14, the British system is credential-based, and regulated by The Office of the Qualifications and Examinations Regulator (Ofqual). The credentials are managed and awarded by the examining boards, such as EdExcel. Credentials include the General Certificate of Secondary Education (GCSE), the Advanced Level General Certificate of Education (A-Levels), and the BTECs, which are vocational credentials awarded exclusively by EdExcel. These credentials are described in the paragraphs below.

### GCSE

The General Certificate of Secondary Education (GCSE) is an academic qualification awarded in a specified subject, generally taken in a number of subjects by students aged 15–16 in secondary education. Some students may decide to take one or more GCSEs before or afterwards; people may apply for GCSEs at any point either internally through an institution or externally. The International version of the GCSE is the IGCSE, which can be taken anywhere in the world, and which includes additional options, for example, relating to coursework and the language used.

Education to GCSE level is often required of students who study for the International Baccalaureate or to GCE Advanced Level (A-level). GCSE exams were introduced as the compulsory school-leavers' examinations in the late 1980s by the then Conservative Party government, replacing the Certificate of Secondary Education (CSE) and GCE Ordinary Level (O-Level) examinations.

## A-Levels

The Advanced Level General Certificates of Education, universally referred to as the A-levels, are usually studied over a two-year period and are widely recognized around the world, as well as being the qualification that is the most common method used by British Universities to determine an applicant's suitability for academic subjects.

## BTECs

The Business and Technology Education Council (BTEC) was a subdegree-conferring body in the United Kingdom until 1996, when its functions were transferred to EdExcel. BTEC qualifications are undertaken in vocational subjects ranging from Business to Engineering and even Animal Care. They are offered across a range of levels (1 to 6) and have equivalence to other qualifications such as the GCSE (levels 1 to 2), A Level (level 3) and university degrees (levels 4 to 6). BTECs tend to be based on practical work or coursework rather than timed examinations.

BTECs are a growing qualification in schools and are said to have high acceptance by employers. In 2005/06, 260,000 students studied BTECs at college; 63,000 studied BTECs in schools; 23,000 studied BTECs at University; and 14,000 employees studied a BTEC while at work.

## EdExcel and Quality Assurance

EdExcel employs different procedures for scoring the GCE/GCSEs and the BTECs.

The scoring for GCEs is similar to the procedures used for scoring constructed response assessments in the United States. Examiners are trained and monitored in scoring centered in a structured fashion. Marks are recorded in an electronic system (ePEN, based on the technology developed by Pearson in the U.S.). Scoring quality assurance is monitored through processes known as "internal moderation" and "external moderation". Internal moderation occurs within the center and is achieved through standardized training, and monitoring the scoring of individual markers. External moderation involves sampling candidates' ePortfolios for each unit entered in each assessment series. The ePortfolios are burned to CD-ROM or DVD and mailed to a designated "moderator". The moderators' marks are compared to the center awarded marks, and this comparison may provide the basis for statistical adjustments that could be made to the marks of some or all of the candidates entered for that unit.

For the BTECs, an external verifier appointed by EdExcel performs quality assurance. The external verifier checks that the center-designed assignment briefs are consistent

with the purposes of the assessment and that the assessment is undertaken against the correct standard. The external verifier also audits the internal verification process, which is required of all centers. The external verifier provides verbal feedback to each center and submits a written report that will include an action plan for any needed improvements. External verifications of the BTECs take place annually according to published timelines.

