

Overview of Local Assessment Validation Evaluation

November 17, 2011

Carolyn Dumaresq Ed.D., PDE

Rich Maraschiello Ph.D., PDE

Marianne Perie Ph.D., Center for Assessment

Agenda for Today

1. Chapter 4 Local Assessment Validation Advisory Committee – Update
2. Discuss evaluation criteria
3. Review options for local assessments
4. Describe evaluation process
5. Review the *Local Assessment Validity Evaluation Handbook*
6. Determine needs for technical assistance

Chapter 4 Local Assessment Validation Advisory Committee Update

- Proposed the “Stand Alone” local option for the Keystone Exams that was approved by the State Board of Education.
- Developed the Local Assessment Validation Criteria and Rubric

Four Validity Criteria used in Evaluation

- Alignment: *Does the assessment do an effective job of measuring the knowledge and skills in the eligible content PDE developed for each Keystone subject?*
- Fairness: *Does the assessment provide each student with relatively equal opportunities to appropriately demonstrate what they know and can do?*
- Proficiency Levels: *Does the assessment include proficiency levels comparable to those used for the Keystone exam?*
- Consistency: *Does the assessment demonstrate consistency in scores across items, tasks, scorers, forms and years?*

Three Options for Developing Local Assessments

Option A	Option B	Option C
<p>One assessment replaces the Keystone exam</p> <p><i>Must demonstrate that local assessment measures the same content and requires students to show similar or greater levels of performance than the Keystone exam.</i></p>	<p>Several assessment components together replace the Keystone exam</p> <p><i>Must show how scores of components will be combined to make a summary judgment about student proficiency that is of equal or greater rigor than the Keystone exam.</i></p>	<p>An additional component is administered in conjunction with the Keystone exam</p> <p><i>Must show how the score from additional component will be combined with the Keystone score to make a summary judgment about student proficiency that is of equal or greater rigor than the Keystone exam.</i></p>

Evaluation Process

- LEAs complete a submission packet.
- Evaluators will review the evidence submitted for each validity criterion.
- Local assessments must meet the requirements for “satisfactory” on each dimension of the rubric in order to be approved.
- Each content area must be evaluated separately.

Evaluators

- PDE will release a Request for Information (RFI) to identify potential evaluators.
- Eventually an Intent to Qualify (ITQ) for evaluators will be issued by PDE.
- PDE will identify qualified evaluators and establish a list of approved evaluators.
- The LEAs will select an evaluator from the approved list.

Submission Packet

- Any LEA wishing to use a local assessment option must prepare and submit a packet of evidence to the selected evaluator.
- The Local Assessment Validity Evaluation Handbook provides a description of the submission template and sample evidence.
- All evidence will be evaluated using the rubric criteria on the following slides.

Evaluation Criteria	Superior	Satisfactory	Insufficient
Alignment	<p>In addition to the evidence characterizing the satisfactory level:</p> <ul style="list-style-type: none"> • Evidence of depth of knowledge alignment from results of “think-aloud” protocols or other similar analyses • Evidence from an external alignment study • No gaps in coverage of the standards, all items/tasks are aligned to specific standards, and depth of knowledge represented by the items/tasks matches the expectations for depth of knowledge in the standards 	<ul style="list-style-type: none"> • Documentation of adequate sampling of all content standards • Evidence from an internal alignment study that used a two-way alignment process • Few gaps in the coverage of the standards, all of the items/tasks are aligned to specific standards, and there is a range of depth of knowledge (including DOK 4) represented by the items/tasks • Plans for periodic review of alignment 	<ul style="list-style-type: none"> • Items represent content standards, but many standards are unaddressed • The content standards are represented well, but the depth of knowledge required to correctly answer items is not in alignment with the standards

Evaluation Criteria	Superior	Satisfactory	Insufficient
Fairness	<p>In addition to the evidence characterizing the satisfactory level:</p> <ul style="list-style-type: none"> • Universal design principles were adhered to in developing the assessment. • Assessment results are communicated in a manner that allows for equitable remediation opportunities • Analysis of distractor choices across student groups (for multiple-choice items) • Disaggregated results show no large discrepancies between total scores and item difficulties 	<ul style="list-style-type: none"> • Procedures are in place to ensure that the items allow individuals from all subgroups to demonstrate their knowledge • Documentation from bias and sensitivity reviews show the items are free of noticeable bias • Accommodations and alternate assessments are provided as needed/appropriate • Performance expectations are communicated clearly to all stakeholders • The district produces and examines results disaggregated by student groups to search for differences in opportunity to learn • Test administration and security protocols ensure that all students experienced an equitable test environment 	<ul style="list-style-type: none"> • Review procedures are in place, but lack the sophistication to dependably detect potential bias • Results are not disaggregated by important (e.g., ones identified by the state on state-level report cards) student groups

Evaluation Criteria	Superior	Satisfactory	Insufficient
Establishment of proficiency levels	<p>In addition to the evidence characterizing the satisfactory level:</p> <ul style="list-style-type: none"> • Evidence that items represent a wide enough range of difficulties so that the assessment may provide adequate information across the range of cut scores • A plan for evaluating the appropriateness of cut scores once data is available from the assessment (predictive association) • The process for establishing proficiency levels involved individuals from a diverse representation of roles within the school community • Sample items are included in the descriptive information regarding each proficiency level 	<ul style="list-style-type: none"> • The process for establishing proficiency levels followed a researched and validated methodology and documentation of the process is provided • A convincing rationale for the chosen method of used to recommend cut scores is provided • Panelists had knowledge of the content and were demographically representative of all potential panelists in the district • Performance level descriptors are written to a level equally or more rigorous than Keystone’s (adoption of Keystone descriptors is adequate) 	<ul style="list-style-type: none"> • The performance level descriptors are not as rigorous as to the Keystone descriptors • Percent correct or course grade measures define the cut scores • The cut scores are either too idealistic or too lenient (i.e., they do not conform to the performance level descriptors) • Reasonable cut scores have been advanced, but documentation of the process for establishing proficiency levels is lacking

Evaluation Criteria	Superior	Satisfactory	Insufficient
Consistency	<p>In addition to the evidence characterizing the satisfactory level:</p> <ul style="list-style-type: none"> • A plan for ongoing calibration of raters’ scores to ensure that raters don’t become more rigorous or more lenient from one year to the next • Test equating procedures ensure comparable test difficulty across forms and/or years • Inter-rater agreement and internal consistency (whichever is applicable) far exceeds minimum requirements 	<ul style="list-style-type: none"> • Evidence is presented for measuring inter-rater agreement on open-ended items and internal consistency (i.e., reliability) on closed-ended items • Numbers meet minimum requirements for inter-rater agreement and/or internal consistency • Evidence of training for consistency within and across years for scorers of open-ended items (if applicable) is presented • A plan for periodic review of the equivalence of test difficulty across forms and/or years exists 	<ul style="list-style-type: none"> • Inter-rater agreement and/or internal consistency is too low to support the uses of the assessment results • Inter-rater agreement was not calculated or numbers were not provided • Only one rater was used for every open-ended item (i.e., 0% read behind)

Local Assessment Validity Evaluation Handbook

- PDE and the Center for Assessment have developed a handbook intended to inform LEAs about the evaluation requirements for a local assessment.
- The Handbook was reviewed by the Local Assessment Validation Advisory Committee.

What's in the Handbook?

- Short descriptions of components of assessment development
- Suggested readings for more information on technical aspects
- Primer on validity evaluation
- Criteria for the validity evaluation
- Template for submission
- Sample completed templates and evidence
- Glossary

Part I: Test Development

- Two main sections: alignment and fairness
- Chapter includes:
 - Directions for determining alignment between items and eligible content and conducting an alignment study
 - Explanation of what evidence relates to validity claims and why
 - Suggested resources

Part II: Achievement Standards

- Includes sections on the development of Performance Level Descriptors and cut scores
- LEAs must use Keystone definition of Proficiency but can go beyond that for other levels
- Provides examples of commonly used standard setting methods and resources for further information

Part III: Technical Quality Requirements

- Includes information related to consistency:
 - Across items/tasks
 - Across forms
 - Across scorers
 - Across years
- Relevance of each section will depend on assessment design
- Provides suggestions for further resources

Part IV: Evaluation Process

- Validity Primer
- Submission Templates
 - Different for each possible option for creating a local assessment
 - For each, the chapter shows:
 - Blank template
 - Instructions for completing template and types of required and optional evidence
 - A sample completed template

Part V: Evidence

- Sample evidence for each of the four validity criteria is included here.
- Samples are not provided for every type of evidence.
 - Intended to get LEAs started and provide examples that are not readily available in textbooks
- This section will be updated as needed with additional samples.

Tentative Timeline

- November - June 2012 – LEAs review local assessment validation requirements
- July 30, 2012 - LEAs inform PDE if they will pursue the local assessment option
- August 2012 - June 2013 – LEAs develop/refine assessments and gather validity evidence
- TBD- Administer assessments and finalize validity evidence

Questions?

For more information on these topics
please visit
www.education.state.pa.us or email
RA-Local-Assessment@pa.gov

The mission of the Pennsylvania Department of Education is to lead and serve the educational community, to enable each individual to grow into an inspired, productive, fulfilled lifelong learner.