ASSESSMENT LITERACY:
Identifying and Developing Valid and Reliable Assessments
Additional Support

Ohio Department of Education
Assessment Literacy

- Resources
  - Available on the Ohio Department of Education website: education.ohio.gov

- Online learning modules
  - Coming soon
Learning Targets

- Understand the connections between assessment literacy and other current educational work, including: Student Learning Objectives, Student Growth Measures, and Formative Instructional Practices.

- Define assessment literacy and understand why it is important to be able to select, develop, and review assessments.
Learning Targets

- Develop a trained eye to critically review existing assessments and design new assessments with regard to:
  - Foundations of assessment literacy (validity, reliability, and bias).
  - Quality assessment design (blueprints, alignment, rigor, stretch, and item design).
Consider This

“Research shows that teachers spend from a quarter to a third of their professional time on assessment-related activities. Almost all do so without the benefit of having learned the principles of sound assessment.”

**Ohio Standards for the Teaching Profession**

**Section Two: Ohio Standards for the Teaching Profession**

1. Teachers understand student learning and development and respect the diversity of the students they teach.
   - Teachers display knowledge of how students learn and of the developmental characteristics of age groups.
   - Teachers understand what students know and are able to do and use this knowledge to meet the needs of all students.
   - Teachers expect that all students will achieve their full potential.
   - Teachers respect students’ diverse cultures, language skills and experiences.
   - Teachers recognize characteristics of gifted students, students with disabilities and at-risk students in order to assist in appropriate identification, instruction and intervention.

2. Teachers know and understand the content area for which they have instructional responsibility.
   - Teachers know the content they teach and use their knowledge of content-area concepts, assumptions and skills to plan instruction.
   - Teachers understand and use content-specific instructional strategies to effectively teach the content concepts and skills of the discipline.
   - Teachers understand school and district curriculum priorities and the Ohio academic content standards.
   - Teachers understand the relationship of knowledge within the discipline to other content areas.
   - Teachers connect content to relevant life experiences and career opportunities.

3. Teachers understand and use varied assessments to inform instruction, evaluate and ensure student learning.
   - Teachers are knowledgeable about assessment types, their purposes and the data they generate.
   - Teachers select, develop and use a variety of diagnostic, formative and summative assessments.
   - Teachers analyze data to monitor student progress and learning and to plan, differentiate and modify instruction.
   - Teachers collaborate and communicate student progress with students, parents and colleagues.
   - Teachers involve learners in self-assessment and goal setting to address gaps between performance and potential.

4. Teachers plan and deliver effective instruction that advances the learning of each individual student.
   - Teachers align their instructional goals and activities with school and district priorities and Ohio’s academic content standards.
   - Teachers use information about students’ learning and performance to plan and deliver instruction that will close the achievement gap.
   - Teachers communicate clear learning goals and explicitly link learning activities to those defined goals.
   - Teachers apply knowledge of how students think and learn to instructional design and delivery.
   - Teachers differentiate instruction to support the learning needs of all students, including students identified as gifted, students with disabilities and at-risk students.
   - Teachers create and select activities that are designed to help students develop as independent learners and complex problem-solvers.
   - Teachers use resources effectively, including technology, to enhance student learning.

5. Teachers create learning environments that promote high levels of learning and achievement for all students.
   - Teachers create an environment that is physically and emotionally safe.
   - Teachers motivate students to work productively and assume responsibility for their own learning.
   - Teachers create learning situations in which students work independently, collaboratively and/or as a whole class.
   - Teachers maintain an environment that is conducive to learning for all students.

6. Teachers collaborate and communicate with students, parents, other educators, administrators and the community to support student learning.
   - Teachers communicate clearly and effectively.
   - Teachers share responsibility with parents and caregivers to support student learning, emotional and physical development and mental health.
   - Teachers collaborate effectively with other teachers, administrators and school and district staff.
   - Teachers collaborate effectively with the local community and community agencies, when and where appropriate, to promote a positive environment for student learning.

7. Teachers assume responsibility for professional growth, performance and involvement as an individual and as a member of a learning community.
   - Teachers understand, uphold and follow professional ethics, policies and legal codes of professional conduct.
   - Teachers take responsibility for engaging in continuous, purposeful professional development.
   - Teachers are agents of change who seek opportunities to positively impact teaching quality, school improvements and student achievement.
Ohio’s Standards for the Teaching Profession

Activity Instructions: Highlight standards and elements related to assessment.

1. Student Diversity
2. Content
3. Assessment
4. Instruction
5. Learning Environment
6. Collaboration and Communication
7. Professional Growth
Ohio’s Standards for the Teaching Profession

Standard 1

- Teachers understand what students know and are able to do and use this knowledge to meet the needs of all students.

Standard 4

- Teachers use information about students’ learning and performance to plan and deliver instruction that will close the achievement gap.
Standard 3: Teachers understand and use varied assessments to inform instruction, evaluate and ensure student learning.

Element 1: Teachers are knowledgeable about assessment types, their purposes and the data they generate.

Indicator (d): Teachers demonstrate an understanding of assessment related issues, such as validity, reliability, bias and scoring, by using assessments and the information from them.
Standard 3: Teachers understand and use varied assessments to inform instruction, evaluate and ensure student learning.

Element 2: Teachers select, develop, and use a variety of diagnostic, formative and summative assessments.

Indicators:

a) **Align** classroom assessment with curriculum and instruction.

b) **Purposely plan assessments** and differentiate assessment choices to match the full range of student needs, abilities and learning styles.

d) **Use assessments** to identify student strengths, promote student growth and maximize access to learning opportunities.
Ohio’s Teacher Evaluation System

Teacher Performance on Standards*
- Accomplished, Skilled or Developing Final Summative Ratings = Professional Growth Plan
- Ineffective Final Summative Rating = Improvement Plan

Teacher Performance on Standards: 50%

Student Growth Measures: 50%

Final Summative Rating
- Accomplished
- Skilled
- Developing
- Ineffective

Student Growth Measures
- A1 – Teacher instructs Value-Added subjects exclusively
  - Teacher level Value-Added: 50%
- A2 – Teacher instructs Value-Added courses, but not exclusively
  - Teacher level Value-Added proportionate to teacher’s schedule: 10 - 50%
  - LEA measures proportionate: 0 - 40%
- B – Approved vendor assessment teacher-level data available
  - Approved vendor assessment: 10 - 50%
  - LEA measures: 0 - 40%
- C – No teacher-level Value-Added or approved vendor assessment data available
  - LEA measures: 50%

*For evaluations in years prior to 2015-2016, teachers with average or higher growth were professional growth plan and those with less than average growth were improvement plans.

©2013. Battelle for Kids. All Rights Reserved.
Ohio’s Teacher Evaluation System

Ohio TEACHER EVALUATION SYSTEM (OTES)
Alternative Framework for 2015-2016

Teacher Performance on Standards

- Accomplished, Skilled or Developing Final Summative Ratings = Professional Growth Plan
- Ineffective Final Summative Rating = Improvement Plan

Formal observation and classroom walkthroughs/informal observations
- Pre-conference
- Observation
- Post-conference
- Complete performance rubric
- Mid-Year Review and Conference

Formal observation and classroom walkthroughs/informal observations
- Observation (both completed by May 1)
- Post-conference
- Complete performance rubric
- Written report (by May 10)
- Final Review and Conference

Student Growth Measures

- A1 – Teacher instructs Value-Added subjects exclusively
  - Teacher level Value-Added: 35%
  - 35%

- A2 – Teacher instructs Value-Added courses, but not exclusively
  - Teacher level Value-Added proportionate to teacher’s schedule: 10 – 35%
  - District measures proportionate: 0 – 25%
  - 10 – 35% 0 – 25%

- (or) B – Approved vendor assessment teacher-level data available
  - Approved vendor assessment: 10 – 35%
  - District measures: 0 – 25%
  - 10 – 35% 0 – 25%

- (or) C – No teacher-level Value-Added or approved vendor assessment data available
  - District measures: 35%
  - 35%

Alternative Components

- One or any combination of the following:
  - Student surveys
  - Teacher self-evaluations
  - Peer review evaluations
  - Student portfolios
  - District-determined component

Final Summative Rating

- ACCOMPLISHED
- SKILLED
- DEVELOPING
- INEFFECTIVE

*For evaluations in years prior to 2015-2016, teachers with average or higher growth wrote professional growth plans and those with less than average growth wrote improvement plans.
Student Growth Measures

Options to Measure Student Growth (OTES)

- Value-Added
- Approved Vendor Assessments
- Locally Determined Measures
  - Student Learning Objectives (SLOs)
  - Shared Attribution
  - Approved Vendor Assessments
    - Category A2 teachers only

Assessment literacy enables educators to effectively review, select, and develop valid, reliable measures.
Student Learning Objectives (SLOs)

- A measurable, long-term academic growth target that demonstrates a teacher’s impact on student learning.

- SLOs include:
  - Baseline and trend data
  - Identified student population or sample
  - Period of time covered by SLO
  - Content and Standards covered by SLO
  - The assessment that will be used to measure progress
  - Expected student growth including rationale for the growth target
Student Learning Objectives (SLOs)

- SLOs enable all educators to demonstrate their impact on student learning and receive recognition for their efforts. (ODE)
  - The strength of the SLO process is directly related to the quality of the assessments used to set and measure growth targets.
Connections to Formative Instructional Practices (FIP)

- Using clear learning targets
- Collecting and documenting evidence of student learning
- Providing effective feedback
- Student ownership of learning
Remember: The strength of the SLO process is directly related to the quality of the assessments used to set and measure growth targets.
Impact of Assessment Literacy

Assessment Literacy → Student Learning Objectives

→ Classroom Formative and Summative Assessments

→ Understanding Standardized Assessments

→ Student Learning
Balanced Assessment System

What do I know about my students?

Before Instruction:
Diagnostic Baseline Pre-Test

During Instruction:
Ongoing Formative

After Instruction:
Summative Post-Test End of Course

Formative or Summative Use
Pathway to Student Growth Measures

Remember: The strength of the SLO process is directly related to the quality of the assessments used to set and measure growth targets.
Identify Learning Priorities

- Longevity
- Leverage
- Levels

Learning Priorities: Longevity

- The knowledge and skills important today, tomorrow, and in the future.
- Example: Reading comprehension
Learning Priorities: Leverage

- The knowledge and skills used in more than one content area at a specific grade level (horizontal).
- Example: Creating and analyzing charts and graphs
Learning Priorities: Levels

- The knowledge and skills key to success in subsequent grade levels or courses (vertical).
- Example: Strong understanding of fractions to prepare for work with rational numbers.
Learning Priorities

- Introduced early
- Reinforced often
- Multiple opportunities for learning
Complex Standards

Unpack, Unwrap, Deconstruct

☐ Compare two three-digit numbers based on meanings of the hundreds, tens, and ones digits, using >, =, and < symbols to record the results of the comparison.

☐ Apply multiple criteria to evaluate the quality and effectiveness of music performance and composition including their own.
Apply multiple criteria to evaluate the quality and effectiveness of music performance and composition including their own.

<table>
<thead>
<tr>
<th>Skill</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apply</td>
<td>Multiple criteria of quality</td>
</tr>
<tr>
<td></td>
<td>Multiple criteria of effectiveness</td>
</tr>
<tr>
<td>Evaluate</td>
<td>Music performance by others</td>
</tr>
<tr>
<td></td>
<td>Music composition by others</td>
</tr>
<tr>
<td></td>
<td>Own music performance</td>
</tr>
<tr>
<td></td>
<td>Own music composition</td>
</tr>
</tbody>
</table>
Complex Standards?

- Compare two numbers between 1 and 10 presented as written numerals.
- Capitalize dates and names of people.
- People earn income by working.
- Use different body parts to strike a lightweight object (such as a balloon) and keep it in the air.
Partner Dialogue and Reflection

1. Share a **new awareness** or learning.
2. Share a **challenge** you may encounter in your implementation of growth measures.
3. Share an **opportunity** student growth measures may provide at your school or district.
Reflection: The Need for Assessment Literacy

- Connections of assessment literacy to:
  - Ohio’s Standards for the Teaching Profession
  - Student Growth Measures
  - Student Learning Objectives
  - Formative Instructional Practices (FIP)

- Balanced Assessment
- Learning Priorities
- Complex Standards
Remember: The strength of the SLO process is directly related to the quality of the assessments used to set and measure growth targets.
High-Quality Assessment
High-Quality Assessments

Validity: Does it measure what it intends to measure?

Reliability: Does it provide trustworthy results?

Bias: Does it offend or unfairly penalize?
Validity = Accuracy
Validity

Valid *inferences* of the test results

3 types of validity evidence:

- Content related
- Criterion related
- Construct related
Alignment

Since our ultimate goal is to improve learning and teaching through the use of the data generated from student growth measures, the assessments used must be closely aligned to articulated standards taught in the classroom.
Alignment

What do I want my students **to know** and **be able to do?**
Content and Skills

Identify Skills

who is telling the story at various points in a text.

Content
K.G2: Correctly name shapes regardless of their orientations or overall size.

- Ask student to point to a shape.
- Ask student to name the shape.
- Ask student to point to all of the shapes identified.
LS 5.2: Food webs can be used to identify the relationships among producers, consumers and decomposers in an ecosystem.

3. Which of the choices below describes the roles of the four organisms in the correct order from left to right?
   A. consumer, consumer, decomposer, producer
   B. decomposer, consumer, consumer, producer
   C. producer, consumer, consumer, decomposer
Aaron decided to compete in a decathlon.

What is a decathlon?
Create a mixed media landscape that conveys a story and depicts a well-known character from children’s fiction. Be prepared to discuss your artistic/stylistic choices as well as the cultural and/or social symbolism of the piece.

PE 6.4: Connect selected ideas, concepts and processes used in visual art with those used in other academic disciplines.
PR 6.2: Experiment with a variety of techniques and working methods when creating an original work of art.
RE 6.3: Explore and discuss how aspects of culture influence ritual and social artwork.
Is This Item Aligned?

Analyze the causes and consequences of major political, economic and social developments of the 1930’s with emphasis on the Dust Bowl.

Which of the following areas was most associated with “the Dust Bowls” of the 1930s?

- a. Area marked A
- b. Area marked B
- c. Area marked C
- d. Areas marked by D and E
Developing a Trained Eye

Analyze alignment claims

ALL NEW!
Now aligned to the Ohio Learning Standards Standards!
Is This Item Aligned?

3rd Grade Math

Let’s look at a classroom assessment for alignment. To what standard would these questions likely be aligned?

- Examine questions 1–3
- Examine questions 4–7
- Examine question 8
- Examine question 9
- Examine question 10
Are the expectations and cognitive demands consistent for the learning priorities, the instruction, and the assessment?
Content-Related Validity?

Content-Related Validity?

Content-Related Validity?


©2013. Battelle for Kids. All Rights Reserved.
Content-Related Validity

Excellent Content-Related Validity

Ways to Improve Validity

- Eliminate assessment items that contain content unrelated to what is intended to be measured.
- Ensure a representative distribution of assessment items.
- Ensure item alignment to both the content and the skill levels.
- Ensure that demonstration of mastery of content and skills being measured is not affected by content and skills not being measured.
Reliability = Consistency
Reliability and Validity

- A measure can be reliable, but not have validity
- A measure must be reliable to have validity
Reliability

- **Stability**: Does a student score in the same range when given the same assessment at different times?

- **Alternate Form**: Does a student score in the same range when given an alternate form of the assessment?
Ways to Improve Reliability

- Allow enough time to complete the assessment.
- Include enough items to accurately measure the content and skill indicated, including items of various complexity.
- Avoid ambiguous test questions.
- Provide clear directions.
- Develop a systemic administration procedure.
- Ensure consistent use of rubrics.
- Use multiple scorers (when possible) for items that are not selected response.
Bias

Assessment items and assessments that are biased lead to misinterpretations of the data and misinformed instructional decisions.
Bias and Validity

- Often, bias results in a measure of membership in a group more than a measure of a content objective (content validity).
- Bias causes confusion between what is being tested and who is being tested.
- Assessments written in a language in which the student is not fluent are biased.
- Bias can be positive or negative.
Is This Item Biased?

This is an item from an IQ test. What bias might this item have?

Rifle \(\text{Is To}\) Hunter

Saw \(\text{AS}\) ?
Is This Item Biased?

If you want to conserve resources, one really good way to do so is to:

A. Drive your car at higher speeds.
B. Use your washing machine for small loads rather than large ones.
C. Throw your newspapers away.
D. Take notes on your cell phone rather than on paper.
Is This Item Biased?

What did Garrett Morgan invent that helped improve transportation?

A. the steam engine
B. the traffic signal
C. the automobile
D. the airplane
Ways to Reduce Bias

- Be aware of possible sources of bias.
- Ask a colleague to review the assessment for possible sources of bias.
- Be aware that bias may exist if groups of students tend to do exceptionally poor or well on certain items.
Partner Dialogue and Reflection

1. Share a new awareness or learning.
2. Share a challenge you may encounter when developing valid, reliable assessments that are free of bias.
3. Share an opportunity that this work may provide in your school or district.
Reflection: Foundations of Assessment Literacy

- Validity (Content-Related)
- Reliability (Stability and Alternate Form)
- Bias
Assessment Inventory

- What local, state, and national assessments already exist in our district that may be used to fit our purpose?
- What other assessments must be developed or purchased to fulfill the purpose?

Assessment literacy skills related to sound item and assessment design are essential for those selecting or developing assessments.
## Webb’s Alignment Criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Alignment</td>
<td>Are the expectations consistent for both curriculum and assessments?</td>
</tr>
<tr>
<td>Depth-of Knowledge Consistency</td>
<td>Are the cognitive demands of curricular aims and assessments the same?</td>
</tr>
<tr>
<td>Range of Knowledge</td>
<td>Is the span of knowledge reflected in curriculum and assessment the same?</td>
</tr>
<tr>
<td>Balance of Representation</td>
<td>Does assessment reflect the priorities of the curriculum and the instruction?</td>
</tr>
</tbody>
</table>

Bloom’s Taxonomy vs. Webb’s Depth of Knowledge

- In Ohio’s Learning Standards, Bloom’s Taxonomy was helpful to identify the level of rigor.
- Teachers were encouraged to pay attention to verbs to identify the level of cognitive complexity.

**4th Grade:** Identify examples of cause and effect

**6th Grade:** Analyze examples of cause and effect
Many of Ohio’s learning standards require teachers to look beyond the verb to determine the level of cognitive complexity.

1st Grade: Describe characters, settings, and major events in a story using key details.

2nd Grade: Describe how characters in a story respond to major events and challenges.

3rd Grade: Describe characters in a story (e.g., their traits, motivations, or feelings), and explain how their actions contribute to the sequence of events.
Level of Complexity

- Bloom’s Taxonomy focuses on the type of thinking required to successfully answer the assessment item.
- Webb’s Depth of Knowledge focuses on how deeply the content must be understood in order to be successful.
- Both the thinking process (Bloom’s) and the depth of content knowledge (Webb’s) are important to the design of curriculum, instruction, and assessment.

Hess’ Cognitive Rigor Matrix

Bloom’s Taxonomy
- Create
- Evaluate
- Analyze
- Apply
- Understand
- Remember

Webb’s Depth of Knowledge
- DOK 1: Recall and Reproduction
- DOK 2: Basic Application of Skills and Concepts
- DOK 3: Strategic Thinking
- DOK 4: Extended Thinking

©2013. Battelle for Kids. All Rights Reserved.
Let’s go back to our 3rd grade math sample assessment.

Consider the Depth of Knowledge (DOK) Level for each of the standards on the back of the assessment.

Label each standard with the DOK level.
Assessment Blueprints
Start With a Plan

Assessment Blueprints

- Record the prioritized learning expectations that will be on the assessment.
- Identify the assessment methods to be used.
- Identify the level of complexity of the learning expectations and the assessment items.
- Determine how much “weight” each learning priority will receive.
- Balance the rigor of the assessment.
Assessment Methods
# Assessment Methods

<table>
<thead>
<tr>
<th>Selected Response</th>
<th>Constructed Response</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Multiple Choice</td>
<td>• Short Answer</td>
<td>• Product</td>
</tr>
<tr>
<td>• Matching</td>
<td>• Extended Response</td>
<td>• Visual</td>
</tr>
<tr>
<td>• True/False</td>
<td>• Essay</td>
<td>• Verbal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Physical</td>
</tr>
</tbody>
</table>
# Multiple Assessment Methods

<table>
<thead>
<tr>
<th>Students Demonstrate Learning through the Written Word</th>
<th>Students Demonstrate Learning through Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selected Response</td>
<td>Product Performance</td>
</tr>
<tr>
<td>Constructed Response</td>
<td>Physical/Verbal/Visual Performance</td>
</tr>
</tbody>
</table>

Learning is measured through correct or incorrect responses. Learning is measured through rubrics or scoring guides. These methods may involve a variety of responses and/or processes.
Determining Assessment Method

- Matching the learning expectation to the assessment method impacts validity.
- Some learning expectations may be able to be measured with multiple methods.
- Some methods are not appropriate for some learning expectations.
- Methods selected can also be influenced by:
  - Number of students
  - Turn-around time
Webb’s DOK and Assessment Method

- **Level I and Level II**
  - Usually one right answer
  - Usually assessed by selected response or constructed response

- **Level III and Level IV**
  - More than one correct answer or approach is possible and may involve real-world applications in new situations
  - Usually assessed by constructed response or performance
Which Assessment Method?

Identify the eight major planets in the solar system.

1. What is the DOK level of this learning expectation?

2. What methods would be good choices for assessment?
Which Assessment Method?

Make a line plot to display a data set of measurements in fractions of a unit (1/2, 1/4, 1/8).

1. What is the DOK level of this learning expectation?

2. What methods would be good choices for assessment?
Which Assessment Method?

Work effectively in a group to complete a task or solve a problem for which the group is held accountable.

1. What is the DOK level of this learning expectation?

2. What methods would be good choices for assessment?
Which Assessment Method?

Write arguments to support claims with clear reasons and relevant evidence.

1. What is the DOK level of this learning expectation?

2. What methods would be good choices for assessment?
Which Assessment Method?

Create a series of bar graphs that show kinetic energy, potential energy, and thermal energy for eight different positions on a roller coaster.

1. What is the DOK level of this learning expectation?

2. What methods would be good choices for assessment?
Which Assessment Method?

Compare and contrast visual forms of expression found throughout local regions and in different cultures of the world.

1. What is the DOK level of this learning expectation?

2. What methods would be good choices for assessment?
Weighting Assessment Items

How much evidence is enough?

- The broader the learning target is in scope, the larger the sample you will need to ensure it is covered thoroughly.
- The more important the learning target is, the larger your sample should be.
- The more important the decision to be made on the basis of the result, the larger the sample should be.
Stretch

- Does the assessment allow both high-achieving and low-achieving students to show growth?
  - Include questions of varying difficulty and complexity (foundational, basic, and advanced level)
  - Examine baseline results to inform decisions regarding the need for stretch on the assessment.
## Blueprint Example

### Levels of Cognitive Complexity

<table>
<thead>
<tr>
<th>Learning Expectations</th>
<th>Target DOK Level</th>
<th>Method</th>
<th>DOK 1 Recall or Reproduce</th>
<th>DOK 2 Basic Application of Skills/Concepts</th>
<th>DOK 3 Strategic Thinking</th>
<th>DOK 4 Extended Thinking</th>
<th>Total Points</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe how characters in a story respond to major events or challenges.</td>
<td>2</td>
<td>SR &amp; CR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5 points</td>
<td>20%</td>
</tr>
<tr>
<td>Ask and answer questions to demonstrate understanding of key details in a text.</td>
<td>1</td>
<td>SR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3 points</td>
<td>12%</td>
</tr>
<tr>
<td>Determine the lesson or moral of a story.</td>
<td>2</td>
<td>CR</td>
<td>1 point</td>
<td>2 points</td>
<td></td>
<td></td>
<td>3 points</td>
<td>12%</td>
</tr>
<tr>
<td>Identify difference in point of view of different characters.</td>
<td>2</td>
<td>CR</td>
<td>2 points</td>
<td>4 points</td>
<td></td>
<td></td>
<td>6 points</td>
<td>24%</td>
</tr>
<tr>
<td>Compare and contrast two or more versions of the same story.</td>
<td>3</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8 points</td>
<td>32%</td>
</tr>
<tr>
<td>Total Points</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6 points</td>
<td>24%</td>
</tr>
<tr>
<td>%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7 points</td>
<td>28%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12 points</td>
<td>48%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0 points</td>
<td>0%</td>
</tr>
</tbody>
</table>

**Rigor**

- 6 points: 24%
- 7 points: 28%
- 12 points: 48%
- 0 points: 0%
<table>
<thead>
<tr>
<th>Learning Expectations</th>
<th>Target DOK Level</th>
<th>Method</th>
<th>DOK 1 Recall or Reproduce</th>
<th>DOK 2 Basic Application of Skills/Concepts</th>
<th>DOK 3 Strategic Thinking</th>
<th>DOK 4 Extended Thinking</th>
<th>Total Points</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Describe how characters in a story respond to major events</td>
<td>2</td>
<td>SR &amp; CR</td>
<td>2 points</td>
<td>3 points</td>
<td>5 points</td>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ask and answer questions to demonstrate understanding of key details in a text</td>
<td>1</td>
<td>SR</td>
<td>3 points</td>
<td>3 points</td>
<td>12%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determine the lesson or moral of a story</td>
<td>2</td>
<td>CR</td>
<td>1 point</td>
<td>2 points</td>
<td>3 points</td>
<td>12%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify difference in point of view of different characters</td>
<td>3</td>
<td>P</td>
<td>2 points</td>
<td>4 points</td>
<td>6 points</td>
<td>24%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compare and contrast two or more versions of the same story</td>
<td>3</td>
<td>P</td>
<td>8 points</td>
<td>8 points</td>
<td>32%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total Points</strong></td>
<td><strong>2 points</strong></td>
<td><strong>3 points</strong></td>
<td><strong>5 points</strong></td>
<td><strong>6 points</strong></td>
<td><strong>0 points</strong></td>
<td><strong>7 points</strong></td>
<td><strong>12 points</strong></td>
<td><strong>0 points</strong></td>
</tr>
<tr>
<td><strong>%</strong></td>
<td><strong>24%</strong></td>
<td><strong>28%</strong></td>
<td><strong>48%</strong></td>
<td><strong>0%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3rd Grade Math

- Review the 3rd grade math assessment by blueprinting it.
  - Are the DOK Levels of the assessment items appropriate?
  - Does the assessment provide stretch?
  - What are the priorities of the assessment?
  - How rigorous is the assessment?
## Blueprint Example

<table>
<thead>
<tr>
<th>Learning Expectations</th>
<th>Target DOK Level</th>
<th>Method</th>
<th>DOK 1: Recall or Reproduce</th>
<th>DOK 2: Basic Application of Skills/Concepts</th>
<th>DOK 3: Strategic Thinking</th>
<th>DOK 4: Extended Thinking</th>
<th>Total Points</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Understand a fraction is part of a whole.</td>
<td>1</td>
<td>SR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>40%</td>
</tr>
<tr>
<td>2A &amp; 2B. Represent a fraction on a number line by defining the interval from 0 to 1</td>
<td>2</td>
<td>SR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td>30%</td>
</tr>
<tr>
<td>at lengths 1/b.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3A. Understand two fractions are equivalent if they are at the same point on a number</td>
<td>1</td>
<td>SR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>10%</td>
</tr>
<tr>
<td>line.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3C. Express whole numbers as fractions.</td>
<td>1</td>
<td>SR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>10%</td>
</tr>
<tr>
<td>3D. Compare two fractions and record the results of the comparison as &gt;, =, &lt;.</td>
<td>2</td>
<td>SR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>10%</td>
</tr>
<tr>
<td>Recognize comparisons are valid only when the two fractions compare to the same</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>whole.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Total Points**: 6 points 7 points 12 points 0 points  
**%**: 24% 28% 48% 0%
How to Use Assessment Blueprints

- **Define** the parameters of an assessment *before* you begin the creation process.
- **Review** a locally- or vendor-developed assessment for alignment to learning expectations, instruction, and cognitive complexity.
- **Confirm** in advance the information and knowledge you are planning to assess, which can guide instructional activities.
- **Share** blueprints with students to help make learning targets clear.
Create a blueprint for an assessment used to obtain a driver’s license.
Partner Dialogue and Reflection

1. Share a **new awareness** or learning
2. Share a **challenge** you may encounter in planning for assessments.
3. Share an **opportunity** that planning for assessment may provide in your school or district.
Reflection: Planning for Assessment

- Cognitive Rigor
  - Bloom’s Taxonomy
  - Webb’s Depth of Knowledge
- Assessment Blueprints
- Assessment Methods
- Weighting Assessment Items
## High-Quality Assessment Items

<table>
<thead>
<tr>
<th>Validity</th>
<th>Reliability</th>
<th>Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Aligned to the appropriate standards</td>
<td>□ Match the cognitive level presented in instruction</td>
<td>□ Clear, concise, and unambiguous</td>
</tr>
<tr>
<td>□ Clear, concise, and unambiguous</td>
<td></td>
<td>□ Sensitive to cultural differences</td>
</tr>
</tbody>
</table>

©2013. Battelle for Kids. All Rights Reserved.
Item Design
Parts of an Assessment Item

- **Stimulus**
  - Text or Graphic

- **Stem**
  - Presents the Problem or Task

- **Answer Options**
  - Used with Selected Response
  - Includes Distractors

- **Scoring Guide**
  - Correct Response
  - Rubric
The shapes shown are part of a design.

What do all these shapes appear to have in common?

A. All have four right angles.
B. All have at least one set of parallel sides.
C. All have four equal angles.
D. All have at least one set of perpendicular lines.
Parts of an Assessment Item

Stimulus: The Text “Annie and the Wild Animals”

Use details from the reading selection to complete the web.

- Animals that ate Annie’s corn cakes
  - snarling wildcat
  - giant moose
  - stag and family

Stem
## Scoring Guide

Include explanation and sample response

### Scoring Guidelines

<table>
<thead>
<tr>
<th>Points</th>
<th>Student Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2 point</strong></td>
<td>Response correctly completes both blanks.</td>
</tr>
<tr>
<td></td>
<td><strong>Sample Response:</strong></td>
</tr>
<tr>
<td></td>
<td>• Bear</td>
</tr>
<tr>
<td></td>
<td>• Wolf</td>
</tr>
<tr>
<td><strong>1 point</strong></td>
<td>Response correctly completes one blank.</td>
</tr>
<tr>
<td></td>
<td><strong>Sample Response:</strong></td>
</tr>
<tr>
<td></td>
<td>• Bear</td>
</tr>
<tr>
<td><strong>0 point</strong></td>
<td>Response is incorrect or irrelevant.</td>
</tr>
</tbody>
</table>
A Well-Constructed Stimulus Should:

- Not include extraneous reading material that is irrelevant.
- Only be used if it is essential to the associated question.
- Be correct and accurate.
A Well-Constructed Stem Should:

- Be in the form of a direct question (closed stem) or an incomplete statement (open stem) with direct questions being preferred.
- Avoid using negative statements such as *not* or *never*.
- Avoid irrelevant information.
- Be grammatically consistent within all options.
High-Quality Answer Options Should:

- Avoid repeating words or phrases.
- Be plausible to determine if students have mastered the knowledge or skill being tested.
- Include distractors based on common student errors or misconceptions.
- Be roughly the same length.
- Avoid choices of “all of the above” or “none of the above.”
Improving Selected Response Items

To increase reliability and minimize correct guessing

- Use 4-5 answer options.
- Avoid “cueing” correct answers from other questions.
- Avoid overusing one answer choice.
- Test higher-order thinking; avoid lifting material verbatim from instructional materials.
Improving Constructed-Response Items

- Clearly define the task.
- Avoid the use of question options.
  - When students have questions to choose from, they are actually taking different assessments.
  - Impacts reliability as it is impossible to discriminate between the student who could answer all options and the student who could not.
### Evaluating Assessment Items

#### 3rd Grade Math

<table>
<thead>
<tr>
<th>Alignment to Standards:</th>
<th>Is the Learning Objective clearly reflected in the assessment measure?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>All items in the assessment align to the standard(s) addressed in the SLO.</td>
</tr>
<tr>
<td>Somewhat</td>
<td>The assessment measure addresses the full range of topics and skills included in the SLO.</td>
</tr>
<tr>
<td>No</td>
<td>The focus of the assessment mirrors the focus of the curriculum and standards.</td>
</tr>
<tr>
<td></td>
<td>The items or task match the full range of cognitive thinking required during the course.</td>
</tr>
<tr>
<td></td>
<td>The assessment requires students to engage in higher order thinking where appropriate.</td>
</tr>
</tbody>
</table>

Comments:

<table>
<thead>
<tr>
<th>Stretch:</th>
<th>Will all students be able to demonstrate growth on this assessment?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>The test includes items that cover prerequisite knowledge and skills from prior years and appropriate, content-relevant items that will challenge the highest performing students.</td>
</tr>
<tr>
<td>Somewhat</td>
<td>Test items cover knowledge and skills that will be of value beyond the school year.</td>
</tr>
</tbody>
</table>

Comments:

<table>
<thead>
<tr>
<th>Validity and Reliability:</th>
<th>Is the assessment a valid and reliable tool for the intended purpose?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>The assessment does not include overly complex vocabulary.</td>
</tr>
<tr>
<td>Somewhat</td>
<td>Items or tasks are written clearly and concisely.</td>
</tr>
<tr>
<td>No</td>
<td>Clear scoring rubrics or guidance exists for open-ended questions or performance-based assessments.</td>
</tr>
<tr>
<td></td>
<td>The teacher has a plan for administering assessments consistently across classes.</td>
</tr>
</tbody>
</table>

Comments:
Performance Assessment

An assessment of a student’s skill, usually a product or performance, scored using a rubric.
Considerations for Performance Assessment Items

- Have multiple criteria that are being judged
- Reserved for high-level cognitive skills
- Allow for multiple approaches
- Multi-stepped
- Allow for reflection and revision
- Accompanied by a high-quality rubric which specifies quality standards
Rubrics
High-Quality Rubrics

- Help teachers score students’ work.
- Bring clarity to student expectations.
- Identify next steps in learning.
- Clarify curricular content.
- Assist students in self-assessing and increase the level of student ownership.
Types of Rubrics

- **Analytic**
  - Describes the work on each of the selected criterion separately
  - Provides more formative feedback
  - Links to instruction more closely

- **Holistic**
  - Describes the work by applying all the criteria at the same time
  - Produces an overall judgment
  - Used for scoring
Considerations for Performance Levels

- Start with fewer performance levels (3-4)
  - As the number of levels increase, the distinction between them becomes more granular and more difficult to distinguish.
- Determine if numbers or labels are more appropriate for your purpose and students.
- If labeling, you may want to start with the “goal level” (proficient, meets expectations, etc.) and fill in other descriptions as needed.
Considerations for Choosing Criteria for Rubrics

Criteria should:

- Represent learning priorities.
- Describe the observable evidence of learning.
- Be independent and distinct from one another.
- Work together as a whole to describe expected learning.
- Be able to be described by different performance levels.
- Not include elements related to work habits (neat, colorful, effort, etc.).

Considerations for Writing Descriptors

- Be precise in describing the performance level of each criteria across the continuum.
- Avoid vague language.
- Differentiate among levels clearly enough so that work can easily be categorized without ambiguity.
- Describe distinct levels of quality focusing on the same criteria across the continuum.
Is This a Quality Rubric?

- Examine the rubric for
  - Clear and distinct performance levels.
  - Criteria that are observable and independent yet work together.
  - Criteria that reflects appropriate learning targets and not work habits.
  - Descriptors that are clear and distinguish between levels.
Partner Dialogue and Reflection

1. Share a **new awareness** or learning.
2. Share a **challenge** you may encounter when developing high-quality assessment items.
3. Share an **opportunity** that this work may provide in your school or district.
Reflection: High-Quality Assessment Design

- Selected Response Items
- Constructed Response Items
- Performance Assessment
- Rubrics
## Challenges and Considerations

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Considerations</th>
</tr>
</thead>
</table>
| Assuring quality across all measures | - Establish processes for development, review, and administration.  
- Determine quality criteria.  
- Institute peer and/or content expert review. |
### Challenges and Opportunities

<table>
<thead>
<tr>
<th>CHALLENGE</th>
<th>OPPORTUNITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aligning initiatives to focus on assessment literacy and comprehensive measures of educator effectiveness</td>
<td>This work will lead to balanced assessment systems where sound assessment practices are used routinely in classrooms to measure, monitor, and respond to student growth.</td>
</tr>
</tbody>
</table>
Contact Information

REGIONAL FIELD SPECIALISTS

- Southeast and Central Regions
  - Tom Musgrave
  - 740-868-3556
  - tom.musgrave@escco.org

- Southwest Region
  - Cathryn Shaw
  - 740-703-5844
  - cathryn.shaw@fcesc.org

REGIONAL FIELD SPECIALISTS

- Northwest Region
  - Apryl Ealy
  - 419-674-3663
  - apryl.ealy@esc-cc.org

- Northeast Region
  - Tom Rounds
  - 216-544-6564
  - thomas.rounds@esc-cc.org