

Assessment Literacy Series



-Module 2- Test Specifications & Blueprints

Objectives

Participants will:

1. Develop test specifications that articulate:

- Number of items by type
- Item point value
- Depth of Knowledge (DoK) levels

2. Develop a blueprint that designates:

- Items per content standard across DoK levels

Helpful Tools

Participants may wish to reference the following:

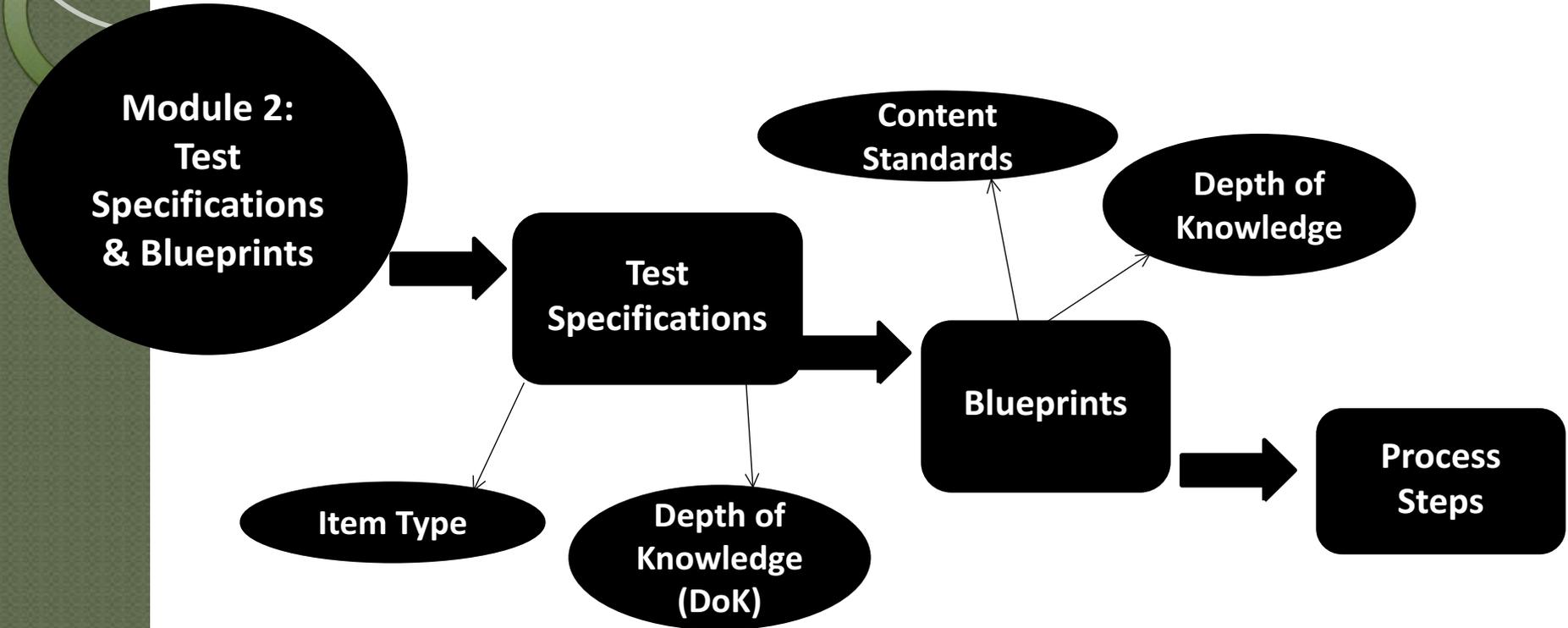
Guides

- Handout #3 – Test Specifications and Blueprint Example
- Handout #4 – Depth of Knowledge (DoK) Chart

Templates

- Template #3 – Test Specifications and Blueprint

Outline of Module 2





Specification Tables

Test Specifications

When developing test specifications consider:

- Sufficient sampling of targeted content standards
 - Aim for a 3:1 items per standard ratio
- Developmental readiness of test-takers
- Type of items
 - Multiple Choice (MC)
 - Short Constructed Response (SCR)
 - Extended Constructed Response (ECR)/Complex Performance tasks
- Time burden imposed on both educators and students

Test Specifications (cont.)

When developing test specifications consider:

- **Cognitive load**

- Aim for a balance of DoK levels

- **Objectivity of scoring**

- Each constructed response item/task will need a well-developed rubric

- **Weight of items (point values)**

- Measures should consist of 25-35 total points; 35-50 points for high school

- **Item cognitive demand level/DoK level**

- Measures should reflect a variety of DoK levels as represented in the targeted content standards

Test Specifications Example

[Handout #3]

Content Strand(s)	MC	SCR	ECR	Total
Expressions & Equations	4	0	0	4
Creating Equations	5	0	0	5
Structure in Expressions	3	0	0	3
Ratios & Proportions	3	2	0	5
Reasoning with Equations & Inequalities	4	1	0	5
Interpreting Functions	3	2	1	6
Real Number System	5	1	1	7
Grand Totals	27	6	2	35
				*Performance measure contains 35 items/tasks.
Content Strand(s)	MC (1 pt.)	SCR (2pts.)	ECR (4pts.)	Total
Expressions & Equations	4	0	0	4
Creating Equations	5	0	0	5
Structure in Expressions	3	0	0	3
Ratios & Proportions	3	4	0	7
Reasoning with Equations & Inequalities	4	2	0	6
Interpreting Functions	3	4	4	11
Real Number System	5	2	4	11
Grand Totals	27	12	8	47
				*Performance measure score based upon 47 points.

Test Specifications Example (cont.)

[Handout #3]

Content Strand(s)	DoK 1	DoK 2	DoK 3	Total
Expressions & Equations	1	2	1	4
Creating Equations	1	2	2	5
Structure in Expressions	2	0	1	3
Ratios & Proportions	0	5	0	5
Reasoning with Equations & Inequalities	3	2	0	5
Interpreting Functions	0	2	4	6
Real Number System	1	4	2	7
Grand Totals	8	17	10	35
				*Performance measure contains items/tasks with the following Level/DoK distribution: DoK 1 = 23% DoK 2 & 3 = 77%

Multiple Choice Items

- Stem (question) with four (4) answer choices
- Typically worth one (1) point towards overall score
- Generally require about one (1) minute to answer

Pros

- Easy to administer
- Objective scoring

Cons

- Students can guess the correct answer
- No information can be gathered on the process the student used to reach answer

Short Constructed Response Items

- Requires students to apply knowledge, skills, and critical thinking abilities to real-world performance tasks
- Entails students "constructing" or developing their own answers in the form of a few sentences, a graphic organizer, or a drawing/diagram with explanation
- Worth 1-3 points

Pros

- Allows for partial credit
- Provides more details about a student's cognitive process
- Reduces the likelihood of guessing

Cons

- Greater scoring subjectivity
- Requires more time to administer and score

Extended Constructed Response Items

- Requires students to apply knowledge, skills, and critical thinking abilities to real-world performance tasks by developing their own answers in the form of narrative text with supporting graphic organizers and/or illustrations
- Worth 4 or more points
- Entails more in-depth explanations than SCR items

Pros

- Allows for partial credit
- Provides more details about a student's cognitive process
- Reduces the likelihood of guessing

Cons

- Greater scoring subjectivity
- Requires more time to administer and score

Depth of Knowledge is...

- The complexity of mental processing that must occur in order to construct an answer
- A critical factor in determining item/task rigor

Level	Example of Verb	Example of Task
DoK Level 1	Recall	List three characteristics of metamorphic rocks.
DoK Level 2	Compare/Contrast	Describe the difference between metamorphic and igneous rocks.
DoK Level 3	Create	Develop a model to represent the rock cycle.
DoK Level 4	Construct	Using multiple sources, develop an essay on the rise of the Industrial Revolution.

Depth of Knowledge Chart

[Handout #4]

DoK Level	Definition	Verbs	Examples
1	Involves recall and the response is automatic. Activities require students to demonstrate a rote response, follow a set of procedures, or perform simple calculations.	define, duplicate, list, memorize, recall, repeat, reproduce, state, classify, describe, discuss, explain, identify, locate, recognize, report, select, paraphrase	<ul style="list-style-type: none"> Identify the main character. Subtract the numbers. Label the rivers on the map. Measure the length of your desk. List the steps in the water cycle.
2	<p>Activities are more complex and require students to engage in mental processing and reasoning beyond a habitual response.</p> <p>These activities make students decide how to approach a problem.</p>	choose, demonstrate, dramatize, employ, illustrate, interpret, operate, schedule, sketch, solve, use, write, appraise, compare, contrast, criticize, differentiate, discriminate, distinguish, examine, experiment, question, test	<ul style="list-style-type: none"> Summarize the events in the story. Describe the cause/effect of an event. Organize the data using a bar graph. Formulate a problem given data. Compare and contrast the main characters from the stories.
3	<p>Activities necessitate higher cognitive demands. Students are providing support and reasons for conclusions they draw.</p> <p>Typically, Level 3 activities have more than one correct response or approach to the problem.</p>	appraise, argue, defend, judge, select, support, value, evaluate, assemble, construct, create, design, develop, formulate, write	<ul style="list-style-type: none"> Support your ideas with details and examples. Design investigations for a scientific problem. Construct a model of the solar system. Using the graph, predict how many teeth would be lost by all the 2nd grade classes in the school and justify your answer.

Think-Pair-Share

- A pencil costs 5 cents at the school store. Matt wants to buy 3 pencils. How much will they cost?
- During a 12.5-hour period, the water level at a certain harbor started at mean sea level, rose to 9.3 feet above sea level, dropped to 9.3 feet below sea level, and then returned to mean sea level. Find a simple, harmonic motion equation that models the height h of the tide above or below mean sea level for this 12.5-hour period.

Think-Pair-Share

- After reading *Refrigeration by Evaporation*, write an informative essay explaining how Abba's invention responded to the needs of Nigeria's people and changed their lives. Use evidence from the text to support your ideas.



Process Steps

[Template #3]

1. Review content standards from completed *Targeted Content Standards Template* and insert content strand(s) into specification table.
2. Determine the number of items by item type (i.e., Multiple Choice, Short Constructed Response, Extended Constructed Response) for each content strand.
3. Ensure item type and cognitive level (I, II, III)/depths of knowledge (DoK) are assigned.
4. Assign item weights to each item type.
5. Assign number of passages (by type) when using literary works.

QA Checklist

- There is a sufficient sampling of targeted standards.
- The specifications reflect a balance between developmental readiness and time constraints.
- Time is considered for both educators and students.
- The cognitive demands reflect those articulated in the targeted standards.
- The measure allows for both objective and subjective scoring procedures.
- The measure consists of 35-50 points with the Level I/DoK I limited to one-third of the items/tasks.



Blueprints

Blueprints

- Content ID #
- Content Statement
- Item Depth of Knowledge (DoK)
 - Performance measures should reflect a variety of DoK levels.
- Sufficient sampling of content standards
 - Aim for a 3:1 item to standard ratio (3 items for every standard).
- Cognitive load
 - Aim for a balance of DoK levels among standards.
 - Design measures with at least 50% DoK 2 or higher.

Blueprint Example

[Handout #3]

Standard/ Content ID	Content Statement	Item Count	DoK 1	DoK 2	DoK 3
8.EE.1	Know and apply the properties of integer exponents to generate equivalent numerical expressions.	2	1	0	1
8.EE.2	Use square root and cube root symbols to represent solutions to equations of the form $x^2 = p$ and $x^3 = p$, where p is a positive rational number. Evaluate square roots of small perfect squares and cube roots of small perfect cubes.	2	0	2	0
A-CED.1	Create equations and inequalities in one variable and use them to solve problems.	5	1	2	2

Process Steps

1. List the standards by number and statement in the appropriate columns. Remember to aim for a 3:1 item to standard ratio.
2. Decide on the item count for each standard and fill in the appropriate column.
3. Determine the number of DoKs for each standard following the specified guidelines for “rigor”.
4. Repeat Steps 1-3 ensuring that item and DoK counts meet the specification requirements.

QA Checklist

- The blueprint lists the content standard ID number.
- The blueprint lists or references the targeted content standards.
- The blueprint designates item counts for each standard.
- The blueprint reflects a range of DoK levels.
- The blueprint item/task distribution reflects that in the specification tables.

Summary & Next Steps

Summary

Module 2: Test Specifications and Blueprints

- Developed test specifications and a blueprint to guide the item development process.

Next Steps

Module 3: Item Specifications

- Given the specifications and blueprint, develop items to measure aspects of the targeted content standards.