**Quality Control Checklist**

|  |  |  |  |
| --- | --- | --- | --- |
| Content Area | Course/Subject Area | Reviewer | Date |
|  |  |  |  |

**Part I: MATERIAL SCREENING**

|  |  |  |  |
| --- | --- | --- | --- |
| Task ID | Task | Status | Comment |
| 1.1 | Purpose statement | □ |  |
| 1.2 | Content standards (selected) | □ |  |
| 1.3 | Specifications table | □ |  |
| 1.4 | Assessment blueprint | □ |  |
| 1.5 | Operational form | □ |  |
| 1.6 | Score key and/or Scoring rubric(s) | □ |  |
| 1.7 | Administrative & scoring guidelines | □ |  |

**Part II: FORM/ITEM RIGOR**

|  |  |  |  |
| --- | --- | --- | --- |
| Task ID | Task | Status | Comment |
| 2.1 | Operational form is developmentally appropriate (100% on grade-level) | □ |  |
| 2.2 | Operational form is rigorous (60% DoK 2 or higher) | □ |  |
| 2.3 | Operational form matches the targeted standards (100% accuracy) | □ |  |
| 2.4 | Operational form has sufficient item/task density (5 items/points) | □ |  |
| 2.5 | Operational form reflects the content pattern (95% coverage) | □ |  |
| 2.6 | Items/tasks are assigned correctly to the targeted content standards | □ |  |
| Task ID | **Task** | **Status** | **Comment** |
| 2.7 | Items/tasks are assigned the correct cognitive level | □ |  |
| 2.8 | Items/tasks are developmentally appropriate (readability, content focus) | □ |  |
| 2.9 | Items/tasks have been screened for sensitive subject matter | □ |  |
| 2.10 | Items/tasks have been screened for potential bias (e.g., contextual references, cultural assumptions, etc.) | □ |  |
| 2.11 | Items/tasks have been screened for fairness, including linguistic demand and readability | □ |  |
| 2.12 | Items/tasks have been screened for structure and editorial soundness | □ |  |

**Part III: STANDARDIZED PROTOCOLS**

|  |  |  |  |
| --- | --- | --- | --- |
| Task ID | Task | Status | Comment |
| 3.1 | Specifications and/or blueprints reflect the operational form | □ |  |
| 3.2 | Administrative guidelines for teachers are clear and standardized | □ |  |
| 3.3 | Item/task directions for test-takers articulate expectations, response method, and point values | □ |  |
| 3.4 | Accommodation guidelines for SWD, 504, ELL, and others are referenced | □ |  |
| 3.5 | SCR/ECR scoring guidelines and rubrics are standardized | □ |  |

**Note 1:** The term “item” is used generically and encompasses any stimuli presented to the test-taker, including reading passages, tasks, etc.

**Note 2:** The Quality Control Checklist is typically used prior to test administration.

**QUALITY CONTROL CHECKLIST**

**Procedural Steps**

Step 1. Identify two subject matter experts (teachers) with experience in teaching the content upon which the assessment is based.

Step 2. Complete Part I by organizing and reviewing the operational form, answer key, and/or scoring rubrics, blueprint, administrative guide, etc.

Step 3. Complete Part II by screening each item/task and highlight any “potential” issues in terms of content accuracy, potential bias, sensitive materials, fairness, and developmental appropriateness and then flagging any item/task needing a more in-depth review.

Step 4. Complete Part III by examining the assessment protocols and identifying any shortcomings.

Step 5. Review the Quality Control Checklist; Prepare for any needed item/task revisions.

**ASSESSMENT QUALITY RUBRIC**

**DIMENSION I: DESIGN**

| **Task** | **Descriptor** | **Technical Evidence** |
| --- | --- | --- |
| **I.A** | The assessment’s design is appropriate for the intended audience and reflects challenging material needed to develop higher-order thinking skills. The purpose of the performance measure is explicitly stated. |  |
| **I.B** | The assessment’s design has targeted content standards representing a range of knowledge and skills students are expected to know and demonstrate. |  |
| **I.C** | Specification tables and blueprints articulate the number of items/tasks, item/task types, passage readability, and other information about the assessment. |  |
| **I.D** | Items/tasks are rigorous (designed to measure a range of cognitive demands/higher-order thinking skills at developmentally appropriate levels) and of sufficient quantities to measure the depth and breadth of the targeted content standards. |  |

**DIMENSION II: BUILD**

| **Task** | **Descriptor** | **Technical Evidence** |
| --- | --- | --- |
| **II.A** | Items/tasks and score keys were developed using standardized procedures, including scoring rubrics for human-scored, open-ended questions. The total time to administer the assessment is developmentally appropriate for the test-takers. |  |
| **II.B** | Items/tasks were created in terms of: (a) match to the targeted content standards, (b) content accuracy, (c) developmental appropriateness, (d) cognitive demand, (e) bias, (f) sensitivity, and (g) fairness. |  |

**DIMENSION II: BUILD (cont.)**

| **Task** | **Descriptor** | **Technical Evidence** |
| --- | --- | --- |
| **II.C** | Administrative guidelines contain step-by-step procedures used to administer the assessment in a consistent manner, including scripts to orally communicate directions to students, day and time constraints, and allowable accommodations or adaptations. |  |
| **II.D** | Scoring guidelines were developed for human-scored items/tasks to promote score consistency across items/tasks and among different scorers. These guidelines articulate point values for each item/task used to combine results into an overall raw score. |  |
| **II.E** | Summary scores were reported in terms of raw and standard scores. Performance levels reflect the range of scores possible on the assessment and use statements or symbols to denote each level. |  |

**DIMENSION III: REVIEW**

| **Task** | **Descriptor** | **Technical Evidence** |
| --- | --- | --- |
| **III.A** | The assessment was reviewed in terms of: (a) item/task distribution based upon the design properties found within the specification and blueprint documents, and (b) item/task and form performance (e.g., levels of difficulty, complexity, distracter quality, bias, and other characteristics) using pre-established criteria. |  |
| **III.B** | The assessment was reviewed in terms: (a) editorial soundness, (b) document consistency, and (c) linguistic demand. |  |

**DIMENSION III: REVIEW (cont.)**

|  |  |  |
| --- | --- | --- |
| **Task** | **Descriptor** | **Technical Evidence** |
| **III.C** | The assessment was reviewed in terms of the following alignment characteristics:   * Content Match (CM) * Cognitive Demand/Depth of Knowledge (DoK) * Content Pattern (CP) * Item/Task Sufficiency (ITS) |  |
| **III.D** | Post administration analyses were conducted on the assessment (as part of the refinement process) to examine items/tasks performance, scale functioning, overall score distribution, rater drift, etc. |  |
| **III.E** | The assessment has score validity evidence that demonstrate item responses were consistent with content specifications. Data suggest that the scores represent the intended construct by using an adequate sample of items/tasks within the targeted content standards. Other sources of validity evidence such as the interrelationship of items/tasks and alignment characteristics of the assessment are collected. |  |
| **III.F** | The assessment’s reliability coefficients are reported for the assessment, which includes estimating internal consistency. Standard errors are reported for summary scores. When applicable, other reliability statistics such as classification accuracy, rater reliability, etc. are calculated and reviewed. |  |

**ASSESSMENT QUALITY RUBRIC**

**Procedural Steps**

Step 1. Identify two subject matter experts (teachers) with experience in teaching the content upon which the assessment is based.

Step 2. Complete Dimension I by reviewing the operational form, answer key, and/or scoring rubrics, blueprint, administrative guide, etc. that were created during the Design Phase of the assessment.

Step 3. Complete Dimension II by conducting alignment and other types of reviews. Also, review documentation on the procedures used in item/task development.

Step 4. Complete Dimension III after administration of the assessment. Evaluate the psychometric evidence of the assessment and how those data were used in refining the assessment for future administrations.