# Interim Assessment Identification and Evaluation Process

## Phase 2: Identifying and Prioritizing Assessment Characteristics & Evidence of Assessment Quality

Selecting, designing, or developing assessments that can be used to support a vision of teaching and learning requires careful planning around that vision. Phase 2 of the Interim Assessment Identification and Evaluation process serves to delineate the characteristics and features a formal interim assessment must demonstrate in order to provide for information that meets your targeted needs. To do so, this document poses questions focused on clarifying:

- the questions you want to answer with the interim assessment results;
- key interim assessment design features (e.g., content representation, length, duration, item format)
- the types of score comparisons that need to be supported;
- how and when the test should be administered and scored; and
- the information that should be reported on individual and aggregate score reports.

Engaging in this purposeful analysis will help those charged with selecting/designing and implementing interim assessment solutions to:

- more effectively and efficiently engage in the review and evaluation of assessment options;
- facilitate discussion of assessment needs with stakeholders and vendors; and
- identify resources, supports and guidance that will be necessary to support implementation (vendor, state, or locally provided).

This document is structured in three parts which are described below.

**Part 1 - Clarification of Use –** In Part 1 the user will document the highest priority uses and any information gaps that the interim assessment is intended to support (per Phase 1 of the Interim Assessment Specifications Process) and provide examples of the types of questions the results of the assessment must answer in order to use the results as intended.

**Part 2 – Defining Assessment Characteristics –** In Part 2 the user will answer a variety of questions focused on identifying the design, administration and reporting features that need to be in place in order to use the interim assessment results as intended and address the questions posed in Part 1.

**Part 3 – Identifying and Evaluating Evidence of Technical Quality:** Part 3 is a resource designed to help users understand and identify the types of evidence necessary to support decisions about the appropriateness of an interim assessment given its intended interpretation and use.

## Part 1: Clarification of Use

Consider the responses you provided to Activity 4 in Phase 1 of the Interim Assessment Identification and Evaluation process.

- What is the highest priority information need or assessment gap that was identified?
- Who needs this information and how will they use it to meet their specific goals and vision for student learning?
- Is a formal interim assessment the best way to collect this information given the manner in which it is intended to be used, or is it best collected through the review of student work or demonstrations resulting from informal classroom assessments?

To expand upon the last bullet above, it is important to note that not all uses will be best served by a formal district assessment. A formal assessment is one that is typically administered, scored and reported in a standardized manner to allow for the comparison of student results. Teacher-directed uses that require a nuanced understanding of student's current capabilities in a specific skill area (e.g., reading fluency, ability to analyze or communicate) may be better addressed through the collection and review of student work or informal assessments (e.g., presentations, projects, read-aloud activities). Prior to moving to Part 2 of this process it is important to determine whether a formal instrument will provide the type of information needed to support your intended use, as this tool was developed with the goal of helping articulate desired assessment characteristics and features that support the development, selection and review of formal assessments.

**Clarifying the Use of Assessment Results**

While you may wish to use the results from an interim assessment in a variety of ways, different uses necessitate different decisions about assessment design, administration and reporting. Therefore, to ensure an assessment meets your highest priority need, it is important to describe what you want to do with the results in as much detail as possible.

Table 1 provides *examples* of the ways in which districts may want to use the results of an interim assessment. Each row indicates the intended use and user of the assessment and provides examples of the types of question(s) a district may want to answer with the assessment results. Stating the questions to be addressed highlights both the information prioritized by the district and the claim the assessment results must support about students (e.g., college and career readiness, proficiency, on-grade level, mastery, above average), programs (e.g., effective, aligned to the curriculum), educators or schools to use the results as intended.

As shown in Table 1, how the results are intended to be used (Column 2) should align to an assessment information need identified by a district through the review of its assessment system (i.e. Phase 1). If it does not, the design of the assessment and the information it provides may not contribute to the existing assessment system in a meaningful way. Ultimately, it is the assessment need and vision (from Part 1) in combination with the intended use and interpretation of the assessment results that defines the required assessment characteristics and features. Knowing just one piece of the puzzle is not enough.

**Table 1. Examples of Uses and Questions to be Addressed**

| Information Need | How do you want to use the results | Primary User(s) | Examples of specific questions to be answered |
|---|---|---|---|
| Information about how likely schools are to meet interim achievement targets by the end of the year. | Predict performance on the end-of-year assessment for purposes of accountability. | District | What percentage of the schools within my district is predicted to meet the state's interim targets for proficiency by the end of the year? |
| Information about the extent to which students are prepared for the summative assessment. | Predict performance on the end-of-year assessment to identify where instructional support should be targeted.<br><br>Determine where/if students need additional practice using the summative assessment interface or responding to certain types of tasks | Teacher | Which of my students are least likely to meet expectations on the end of year state summative assessment?<br><br>Which content areas (e.g., reportable categories) appear to be the most problematic for my students?<br><br>Do students know how to use the online tools? What item or task types caused confusion? |
| Information about how far above or below grade level students are in a specified content domain (e.g., reportable category). | Assign students to work groups for instructional purposes<br><br>Identify appropriately leveled remediation tasks and activities. | Teacher | What is Erika's current reading comprehension level? How far above/below grade level is she performing?<br><br>To what extent are my students meeting grade level expectations related to the use of ratios? |

| | | | |
|---|---|---|---|
| Information about relative areas of strength and weakness within a specified curricular unit or content domain. | Help teachers' understand how to identify and focus re-teaching and remediation efforts.<br><br>Help students understand where they need to focus their attention to meet expectations. | Teacher<br><br>Student | What are the students in my class able to do when it comes to using functions?<br><br>What general writing elements am I struggling with most? |
| Information about how student performance is changing in response to instruction. | Evaluate progress within a specific content domain throughout a period of instruction to identify students who are falling behind. | Teacher | Has Erika demonstrated growth in her understanding of functions since the beginning of the school year?<br><br>Which students are not progressing, or progressing at rate that is not sufficient to be *on-grade level* by the end of the school year? |
| Information about the standards or skills that students struggle with the most within a given content area. | Identify professional development needs at the school and or district level.<br><br>Determine where the curriculum needs to be enriched or revised. | School<br><br>District | Which standards appear to be the most problematic for elementary students within our school?<br><br>Which of the science and engineering practices are students in our district struggling with the most? |
| Information about student performance against district-defined performance expectations (e.g., in social studies, art, etc.) | Identify schools within the district that are most/least effective at supporting students in meeting local expectations for performance. | District | Which schools in our district are most effective?<br><br>What schools are providing equitable access to opportunities for all students? |
| Information about student and aggregate performance against the standards at the end of each marking period/unit. | Evaluate the quality of a new set of curriculum materials.<br><br>Evaluate the alignment of instruction to curriculum. | District | Is our new math curriculum effective at improving student performance? How are students performing relative to last year? |

**Activity for Part 1:** *Please refer to your responses to Question #6 from Activity 4 in Phase 1 of the toolkit (see page 15). What assessment information need do you believe to be the highest priority? How do you intend for teacher, students, districts or schools to use that information given your vision for teaching and learning? What types of questions should the results allow you to answer?*

*In the table below identify your two **highest priority information needs** and the way in which you intend to use the results from an interim assessment. For each use, identify the primary user (i.e. district, school, teacher, students/parents) and provide examples of the types of questions you want to answer with the results. For your reference an example has been provided in the first row of the table. (Note: In order to clearly reflect the claims the data are intended to support, it may be useful to pick one grade and content area to use an example when articulating your questions.)*

| High Priority Assessment Information Needs | Use of assessment results | Primary User(s) | Example of question(s) to be Answered with Results |
|---|---|---|---|
| *Example: Information about student and aggregate performance against the standards at the end of each marking period/unit.* | *Identify broad concepts/skills that require re-teaching before moving to a subsequent unit.*<br><br>*Identify and students that are performing far below expectations for targeted remediation* | *Teachers* | *How well do students understand the range of skills covered within the first marking period of Grade 3 mathematics?*<br><br>*What skills within this unit were most problematic for the students within my class?*<br><br>*Which students require additional practice with the concepts and skills addressed in a given unit?* |
| **Highest Priority** | | | |
| **Second Highest Priority** | | | |

## Part 2: Defining Assessment Characteristics

Now that you have identified the highest priority need and intended uses of the assessment results, the next step is to think about the assessment design, administration and reporting characteristics necessary to use the results as intended. It is important to note that assessment characteristics are not independent and some uses dictate that certain features hold. For example, an assessment designed to predict performance on a state summative assessment requires broad content representation, substantial administration time to allow for enough items to be administered, and reporting that provides a prediction. However, diagnosing specific areas of strength and need requires a narrow focus on targeted content, administration that minimally impacts instructional time and reporting that provides immediate and detailed feedback.

There are many assessment characteristics that conflict with one another and are difficult to meet simultaneously. Some examples include:

- Broad content representation AND short administration times;
- Short administration time AND the provision of detailed, diagnostic information about a student's areas of strength and need;
- Broad score comparability (i.e., across districts/schools/classrooms) AND high flexibility with respect to timing and frequency of test administrations; and
- High stakes test use (e.g., promotion/retention) AND minimal test security.

A careful consideration of these constraints and trade-offs is necessary to identify a reasonable and coherent set of assessment characteristics and features.

**Activity for Part 2: For this activity please consider only your highest priority assessment need.** *Answer the following questions in consideration of your vision for teaching and learning and the primary questions you hope to address with the assessment results. The questions are organized in terms of test design, administration and reporting. For each option, scenarios, interdependencies and other considerations are provided to inform your decision making. After you have gone through the activity for your first need, you can repeat it again for your second high priority need as a means of evaluating the degree to which a similar set of assessment features are appropriate.*

## Test Design

*Directions: Consider what the design of the assessment should look like in order for the results to be used as intended. Answer questions 1-8.*

1. **Should the assessment(s) measure only on-grade-level content, or is it appropriate/necessary to measure content addressed above or below a student's grade level? (Select one)**

| Options | | Examples for when it might apply |
|---|---|---|
| a. | Test items must align to grade-level content standards. | The primary goal is to make inferences about a student's performance relative to the content expectations defined for that grade level**.** |
| b. | Test items may align to on and off grade-level content standards. | The primary goal is to:<br>• identify where a student falls along a learning progression that spans across multiple grade levels, or<br>• identify the knowledge and skills a student has/has not mastered within a content area/domain regardless of the grade-level with which he/she is associated. |

***If you answered 'a' to Question #1, then go to question #2.***

***If you answered 'b' to Question #2, then go to question #3.***

**2. What is the target sampling domain for each test? (Select one)**

| Options | | Examples for when it might apply |
|---|---|---|
| a. | The complete set of grade-level content standards (e.g., mini summative design) | The primary goal is to understand the degree to which students' are meeting a representative sample of the grade level content standards at one or more points throughout the school year to<br>• predict student performance on the end-of-year summative assessment,<br>• use for a pre- , mid-way, or post-assessment against end-of-year grade-level expectations, or<br>• inform local accountability determinations. |
| b. | A reportable category or sub-set of related grade-level content standards (e.g., modular design) | The primary goal is to understand the degree to which students are meeting the expectations defined by a sub-set of related grade level content standards as defined by a reportable category, unit or lesson within a single grade (e.g. G6 The Number System, G6 Proportional Reasoning, G6 Text Types and Purposes). The intended uses may include to<br>• evaluate or monitor student or aggregate performance<br>• inform/differentiate instruction, or<br>• Identify general areas of strength and need. |
| c. | A specific grade level skill, standard or learning objective | The primary goal is to understand the degree to which students are meeting the expectations defined by a specific skill or content standard within a grade (e.g., G6 Mathematics: Apply the properties of operations to generate equivalent expressions; Grade 6 ELA: Explain how an author develops the point of view of the narrator or speaker in a text). For example:<br>• to evaluate student or aggregate performance<br>• to diagnose specific areas of strength or need<br>• to identify areas of additional professional development in a content area |

**3. What should be the target sampling domain for each test? (Select one)**

| Options | | Examples for when it might apply |
|---|---|---|
| a. | The expectations associated with a given content area or construct of interest | The primary goal is to determine a student's current performance level or general proficiency in a broad content area (e.g., what grade level is Lisa currently performing at in Mathematics) unrestricted by grade. The intended uses may include to:<br>• inform placement decisions,<br>• identify students, classrooms or schools having students performing below grade level. |
| b. | A reportable category or sub-set of related content standards that go across grades | The primary goal is to make inferences about students' performance in a particular sub-domain or reporting category that goes across grades. **(**e.g., Number Sense, Proportional Reasoning, Reading Comprehension). The intended uses may include to:<br>• evaluate or monitor student growth within and across years,<br>• inform/differentiate instruction,<br>• identify general areas of strength and need. |
| c. | A specific skill, standard or learning objective that goes across grade levels | The primary goal is to make inferences about a student's areas of strength and need relative to a specific standard or general skill that develops across multiple grade levels. |

4. **What mode(s) of administration would best support the intended use of the results?** While often considered an administration characteristic, the mode of administration is directly related to the item formats that can be supported (e.g., technology-enhanced items). In addition the mode of administration should be familiar to the student and consistent with the manner was instructed and practiced in the classroom.

| Options | | Considerations |
|---|---|---|
| a. | Computer only | **Accessibility**: Requires the range of accommodations/supports required by the target test taking population to be available on computer.<br><br>**Item Types:** May not be appropriate for all types of performance-based tasks. Depends on the type of response or demonstration necessitated by the standards.<br><br>**Immediacy of Feedback/Scoring:** Allows for immediate scoring and reporting of results for most types of items. Some constructed response tasks will still need to be human scored.<br><br>**Logistics**: Requires student to have access to computers or other compatible devices for instruction, practice and assessment. |
| b. | Both paper and computer form | **Accessibility**: Paper-based forms may be necessary if the required range of accommodations to support the target test taking population is not available on computer, or if there are requirements that restrict the use/availability of sufficient technology in some classrooms and schools.<br><br>**Item Types:** If score comparability is desired (across modes) some innovative item types may be restricted from use.<br><br>**Immediacy of Feedback**: Paper-based administrations may not provide timely feedback due to the need for hand scoring and/or scanned scoring of responses. |
| c. | Paper only | **Accessibility**: May be necessary if technology access is limited or if there are requirements that restrict the use of technology in classrooms and schools.<br><br>**Item Types:** Limits the types of items that can be administered to students.<br><br>**Immediacy of Feedback**: May not provide timely feedback due to the need for hand scoring or scanned scoring of responses. |

**5. If the test is to be administered on computer would a fixed form or a computer adaptive assessment best serve the intended use?**

| Options | | Considerations |
|---|---|---|
| a. | Fixed Form – all students receive the same set of items on a given test form | • Useful for identifying specific items or tasks that are problematic within a classroom because all students are administered the same content.<br>• Best if you are assessing a narrowly defined content domain or a grade-specific skill /standard that does not lend itself to an adaptive design (e.g., writing to a prompt).<br>• A large number of items may be necessary to effectively differentiate the performance of high or low ability students. |
| b. | Adaptive Testing (AT) – students are administered items, sets of items, or forms based on how they responded to previously administered items | • Can provide for more accurate estimates of student performance by administering items (or sets of items) aligned to a student's ability level<br>• May allow for greater differentiation of student performance at the high and low end of the ability scale.<br>• Requires a large item bank to ensure that the benefits of adaptive testing are realized.<br>• Typically requires a computer-based administration. |

**6. What item formats should be represented on the assessment? (Select all that apply)**

| | Options | Considerations |
|---|---|---|
| a. | Selected Response | **Efficiency**: Allows for the greatest content coverage in a defined amount of time; can be scored quickly and reliably<br><br>**Mode**: Computer or paper<br><br>**Features**<br>• tend to address knowledge and skills at lower levels of cognitive complexity<br>• are not appropriate for addressing standards that require students to develop a product or provide an explanation for their response<br>• depending on when items are administered in the school year, the easiest to aggregate across administrations for comparisons of performance (e.g., classes, schools, and districts) |
| b. | Open-ended (OE) or constructed response (CR) | **Efficiency**: Less efficient than multiple choice and TEI items because students are generating a written response<br><br>**Mode**: Computer or paper<br><br>**Features**:<br>• high quality OE/CR items allow for the assessment of higher-order thinking skills<br>• some of these responses can be scored immediately using artificial intelligence, but many will need to be scored by educators using provided scoring rubrics<br>• scores may be less reliable, depending on scoring requirements and training provided<br>• depending on who scores the item, difficult to compare across administrations (i.e., independent scorers are often not comparable without high quality training) |

| | | |
|---|---|---|
| c. | Technology enhanced Items (TEIs) | **Efficiency**: Potentially more engaging and authentic way of assessing students if they are familiar with the online tools necessary to respond to different types of TEIs.<br><br>**Mode**: Computer only<br><br>**Features**:<br>• may not be appropriate for addressing standards that require students to develop a product or provide an explanation for their response<br>• students may need practice responding to these types of items prior to testing to promote familiarity<br>• equivalent representations of most TEI items are not possible on paper-based forms<br>• depending on the timing of administration and student opportunity to practice these item types, responses can be easily aggregated for comparisons |
| d. | Performance -based task (e.g. writing to a prompt; reading aloud; conducting an experiment) | **Efficiency**: Least efficient. Most will take at least one class period to administer.<br><br>**Mode**: Online or paper (depending on the task)<br><br>**Features**:<br>• can provide authentic, real world demonstrations of student learning<br>• some responses may be able to be scored using artificial intelligence scoring (e.g., writing responses), but many will need to be scored by educators using provided scoring rubrics<br>• scores may be less reliable, depending on scoring requirements and training provided<br>• depending on who scores the item, difficult to compare across administrations (i.e., independent scorers are often not comparable without high quality training) |

**7. What accommodations should the assessment support/embed given the intended use and test taking population? (Select all that apply and add additional, as needed.)**

| | **Options** |
|---|---|
| a. | 4-function calculator, scientific calculator, graphing calculator |
| b. | Braille/Refreshable Braille |
| c. | Video Sign Language |
| d. | Text-to-Speech |
| e. | Native Language-English Translations |
| f. | Customized Administration Time |
| g. | Large Print |
| h. | Captioning |
| i | Color contrast capability |
| J | Illustration glossaries |
| k. | Read aloud |
| l. | Extended time |
| m. | |
| n. | |

## Test Administration

There are several factors related to the design of the assessment and the manner in which results are intended to be used that can influence test administration decisions. The factors include such things as: the granularity of the content assessed, desired level of score comparability, and the types of inferences the data is intended to support (e.g., growth vs. status). For example, an interim assessment designed to help educators identify where additional support is needed after a brief period of instruction (i.e. a lesson or unit) will likely be short and administered once. On the other hand, tests designed to predict performance on the state summative or measure progress on the grade level content standards will be longer and administered at multiple times throughout the year.

In most cases, some level of score comparability will be necessary to compare performance across individual students or groups of students. To make accurate inferences about the relative performance of students (at the individual or aggregate level) the conditions for administration must be established to support those comparisons. The broader the comparability desired the more constrained administration decisions related to timing, security and ownership will need to be.

*Directions: Think about the way in which you intend to use the assessment results. What administration and scoring conditions need to be in place to support the use of results as intended?*

8. **Will the assessment be administered on an individual student basis or to groups of students at the same time (e.g., classroom, school, and district)? (Select all that apply)**

| Options | | Considerations |
|---|---|---|
| a | Individual | Appropriate if the goal is to<br>• evaluate student performance at a point in time defined by the educator<br>• allow students to self-monitor their performance toward the attainment of a specified standard or learning goal. |
| b | Group | Appropriate if the goal is to:<br>• evaluate the performance of one or more groups of students at a common point in time. |

**9. What level/type of score comparability is needed to use the results as intended? (Select all that apply)**

| | Options | This level of comparability is necessary when you need to… | Uses |
|---|---|---|---|
| a | Within a classroom | • compare and aggregate the performance of students **within a class**. | Evaluate needs for differentiated instruction within a classroom on the assessed content. |
| b | Within a school | • compare the performance of students **within a school.**<br>• aggregate and compare the performance of classrooms **within a school**. | Inform programmatic decisions at a school level (e.g., needs for professional development, required curricular supports, etc.)<br><br>Inform decisions related to the need for differentiated instruction for students across classrooms within a school. |
| c | Within a district | • compare the performance of students **within the district.**<br>• aggregate and compare the performance schools **within a district**. | Inform programmatic decisions at a district level and support district monitoring of performance at the school level. |
| d | Within the state | • compare the performance of students **within the state**.<br>• aggregate and compare the performance of schools and districts **within the state**. | Inform programmatic decisions at a state level and support state monitoring and evaluation of performance at the district and school level. |

**10. Who (i.e., what entity) should determine when the test is administered?**

| Options | | Considerations |
|---|---|---|
| a | State | If the primary user of the results is the state department of education, and/or a primary goal is to support the comparison of schools and districts, the state should establish the guidelines for administration, including the window in which assessments should be administered. (Note that the window may be defined by identifying days on the calendar, or specifying when the assessment should be administered after the completion of a period of instruction. |
| b | District | If the primary user of the results is district leaders and/or a primary goal is to support the comparison of schools and classrooms within the district for monitoring, program evaluation, educator evaluation, etc. then the district should establish guidelines related to when the assessment should be administered. |
| c | School | If the primary user of the results is the school (i.e., principal, teacher teams and/or curriculum coordinator) and/or a primary goal is to support the comparison of classrooms and students within the school for, program/curriculum evaluation, informing student placement decisions, etc. then the school should establish guidelines related to when the assessment should be administered. |
| d | Teacher | If the primary user of the results is the teacher, and the primary goal is to inform instructional decisions at the classroom level the teacher should determine when the assessment should be administered |
| e | Student | If a student is in charge of determining when he/she is ready to take a test along a learning continuum they should have some control as to when the test is administered. |

**11. Given the design of the assessment and the intended use of the results, how often should the assessment is administered? (Select one)**

| Options | | Appropriate when the primary goal is to |
|---|---|---|
| a | Once, prior to instruction. | • evaluate students' current level of understanding within the target content domain in order to plan for or differentiate instruction |
| b | Once, at the end of instruction. | • evaluate how well students have grasped the expectations associated with the target content domain in order to determine where additional support and remediation may be needed at the individual or classroom level |
| c | At the beginning and end of a period of instruction | • measure progress or growth in the target content domain<br>• evaluate the impact of an instructional strategy focused on the sample content domain |
| d | At fixed points throughout a period of instruction (e.g., weekly, monthly, every 9 weeks). | • measure progress with respect to a broad content domain or construct that is the focus of instruction for multiple weeks or months<br>• evaluate the need for student interventions, program eligibility, or performance against exit requirements |
| e | At multiple points throughout the school year. | • measure progress with respect to a broad content domain (i.e., grade level content standards) or a construct that students will be instructed on throughout the year<br>• predict performance on the state summative assessment at different points throughout the year |
| f | As often as needed, as defined by the teacher or student. | • allow for students to evaluate their understanding and address misconceptions as they progress through a unit or course<br>• allow for educators to evaluate individual or aggregate progress throughout instruction, as needed, to direct remediation and/or needs for differentiated instruction |

**12. Given the manner in which results are intended to be used, what level of test security is necessary? (Select one)**

| Options | | When may it apply? |
|---|---|---|
| a | Low | • Results are not intended to inform accountability decisions about students or teachers.<br>• There needs to be flexibility around when individual students take a test.<br>• Scoring must be conducted locally.<br>• Test content is intended to be accessible to teachers and students after administration so it can be used as a tool for identifying and addressing misconceptions. |
| b | Moderate | • Results may be used to inform grading or student level instructional decisions.<br>• The same test forms and/or items may be administered on multiple occasions or used to make inferences about student growth or improvement. |
| c | High | • Assessment results will be used to make accountability decisions about educators or students.<br>• Results will be used in a way that requires strict levels of score comparability between students, classrooms, schools and districts.<br>*Note: High test security does not preclude content from being distributed on an item release schedule.* |

## Reporting

Most interim assessments provide more than just an estimate of student achievement in the content domain; they provide supplemental scores and information that are intended to inform decision making. It is important that the information reported or provided by an assessment aligns with the intended use of the results and the specific inferences about students the data are intended to support. For example, a Grade 3 ELA assessment may report how a student performed in relation to the students in his/her class and the degree to which a vendor-established "on-grade" benchmark was achieved. While this information is important, if the goal is to understand and evaluate student progress over time this information would not be sufficient.

The different types of information provided by interim assessments can be broadly classified into the following five categories: Achievement Status, Predictive, Diagnostic, Growth/Progress and Pedagogical. For illustration, Table 2 provides examples of how the different categories of information are often represented on score reports and examples of the types of questions they are intended to address. It is important to note that in many cases a variety of different types of information can appear on the same score report. In addition, pedagogical information is unique because, unlike the other categories, it cannot be provided in isolation. It is presented as a response to (or as a direct result of) a student's performance on the assessment and may be associated with any of the categories of information presented in Table 2.

**Table 2. Categories of Information Provided by Interim Assessments**

| Type of Information | Purpose(s) of Information | Types of Scores | Examples of Potential Use | Examples of Questions Addressed by this Information | Primary Evidence Needed |
|---|---|---|---|---|---|
| **Achievement Status** | Describe student (or aggregate student) performance at a single point in time | Scaled Score; Raw Score<br><br>Normative: Rank; national/local percentile rank; grade equivalent; stanine<br><br>Criterion Referenced: Performance Level; On-grade level, mastery or college-ready designation | Inform student evaluation, grading or promotion<br><br>Inform local accountability (educator evaluation)<br><br>Evaluate whether students are meeting expectations within a given content domain. | What is the average scaled score in my class?<br><br>How did Lisa's math score compare to that of other 8th grade students in the nation?<br><br>Which students in my class are performing on grade level in reading comprehension? | Evidence that scores are reliable and the test was designed to provide for accurate and intended interpretations about a student's performance in the content domain. |

| Type of Information | Purpose(s) of Information | Types of Scores | Examples of Potential Use | Examples of Questions Addressed by this Information | Primary Evidence Needed |
|---|---|---|---|---|---|
| **Predicted Performance** | Forecast a student's predicted performance on the state summative assessment, the publisher's benchmarks, locally established benchmarks or a nationally recognized criterion measure. | An expected score or performance level on the state summative exam[1];<br><br>A score or performance level indicating the probability that a student will be "proficient" on the state summative exam | Track individual or school performance toward expectations.<br><br>To identify groups of students in need of remediation. | What percentage of students in my school is predicted to meet or exceed Proficiency on the state summative assessment?<br><br>What is the probability that Lisa will meet the College and Career Ready benchmark on the college entrance exam? | Evidence that the procedures and data used to predict student performance support accurate and intended claims about a student's future performance on the criterion measure. |
| **Growth/ Progress** | Describes growth in achievement over time within a content domain. | Gain score, student growth percentile, performance level indicating the degree of growth observed (can be either criterion- or norm-referenced); growth trajectory | Identify students who are falling behind.<br><br>Identify schools or classrooms that are exceeding expectations | Which students are not growing as rapidly as their peers, or as rapidly as needed to meet academic goals (e.g., college and career readiness)?<br><br>What is the impact of a new educational program or intervention on student learning? | Evidence that the data and procedures used to calculate and report growth support accurate claims about student progress within the domain. |
| **Diagnostic Information** | Identify specific areas of strength and need within an assessed content domain (student or aggregate) | Sub-score performance levels or percentile ranks; summaries of specific skills/concepts requiring additional support or remediation | Inform the identification of students for intervention, remediation, or eligibility<br><br>Identify specific strengths and needs to inform instruction. | What are the specific skills that my students are struggling with?<br><br>Which students should I recommend for participation in reading remediation? | Evidence that the assessment was designed to support intended claims about a student's relative areas of strength and need. |

[1] Or some other criterion measure

| Type of Information | Purpose(s) of Information | Types of Scores | Examples of Potential Use | Examples of Questions Addressed by this Information | Primary Evidence Needed |
|---|---|---|---|---|---|
| **Pedagogical Information** | Link students/teachers to appropriate instructional resources or activities. | Student location on a learning trajectory; Lexile or Quantile Score;<br>Link to practice items or tasks for students.<br>Links to instructional guides for educators. | Inform the identification and selection of appropriate instructional supports and/or resources aligned to student needs | What remediation activities should I provide to Ricardo given his performance on the assessment?<br><br>What instructional strategies are available to support improved instruction of Text Dependent Analysis? | Evidence that the provided information/ resources are of high quality and that their use leads to improved student outcomes. |

*Directions: For the remaining questions in Activity #2, think about the specific questions you want to answer with the assessment results in order to use the results as intended, as defined in Activity 1. Refer to Table 2 as needed when providing your response(s).*

13. **What types of information must the assessment must report in order to respond to these questions and use the results as intended? (Select all that apply)**

| Options | | Notes |
|---|---|---|
| a | Achievement Status | Please answer question 14 |
| b | Predicted Performance | Please answer question 15 |
| c | Growth/Progress | Please answer question 16 |
| d | Diagnostic Information | Please answer question 17 |
| e | Pedagogical Information | Please answer question 18 |

**14. In order to answer your questions of interest and use the results as intended what, if any, information must be reported about a student's current achievement? (Select all that apply)**

| Options | |
|---|---|
| a | How it compares to *defined standard, cut score or benchmark (e.g., proficiency, mastery, on-grade level, on-track, etc.)*<br><br>*If so, what is the standard of interest?* |
| b | How it compares to that of *his/her peers (e.g., in the class, school, district, state, nation)?*<br><br>*If so, what is the comparison group(s) of interest?* |
| c | Other<br><br>*Please explain:* |

**15. In order to answer your questions of interest and use the results as intended what, if any, information must be reported about a student's predicted performance on the state summative assessment or another criterion measure? (Select all that apply)**

| Options | |
|---|---|
| a | The predicted scaled score (or scaled score range) |
| b | The predicted performance level or level of attainment (e.g., Below Basic, Basic, Proficient and Advanced; College and Career Ready; ) |
| c | The probability of meeting a defined performance level (e.g., 35% probability of meeting/exceeding proficiency) |
| d | *Other: Please explain* |

16. **In order to answer your questions of interest and use the results as intended what, if any, information must be reported about a student's growth or progress? (Select all that apply)**

| Options | |
|---|---|
| a | How much growth was observed over a defined period of time (e.g., trimester, course, since last administration)<br><br>If so, what is the time period of interest? |
| b | Whether it is sufficient to meet a specified standard or expectation for performance at some point in the future *(e.g., proficiency, mastery, on-grade level, on-track, etc.)*?<br><br>*If so, what is the standard of interest?* |
| c | How it compares to that of his/her peers (e.g., "students like her/him," all students in the class, school, district, state, nation)<br><br>*If so, what is the comparison group of interest*? |
| d | *Other: Please explain* |

17. **What kind of diagnostic information should the assessment provide to support the intended use (e.g., sub-scores, detailed descriptions of relative areas of strength and weakness, item-level analyses and distractor rationales)? (Please describe)**

**18. What kind of pedagogical information are you looking for the assessment to provide? (Please describe)**

## Part 3: Identifying and Evaluating Evidence of Technical Quality

The evidence necessary to evaluate the quality of an assessment is explicitly tied to the manner in which the results are to be used and the claims they are intended to support. While there are certain types of evidence that must always be considered (e.g., alignment, item and form quality, reliability) what that evidence looks like and the manner in which it is prioritized and evaluated will vary across contexts. For example, the evidence necessary to support a claim of "college readiness" that informs school-based decisions about a student's eligibility for graduation, will be different than the evidence collected to support a claim of "mastery" that informs a teacher's decisions about whether a student is ready for the next unit of instruction. The former requires evidence reflecting a) the precision of test scores for making accurate decisions about "readiness[2]" b) the appropriateness of the readiness performance standard or cut-score and c) the relationship between test results and other measures reflecting college readiness. While similar *types* of evidence may be gathered to evaluate claims related to mastery, how that evidence is prioritized and the standards of quality to which it is held may differ. For example, the cut score used to define mastery may be based solely on a teacher's judgement regarding the percentage of questions a student must answer correctly in a particular content area (e.g., proportional reasoning). While this would not be defensible for an assessment used to determine eligibility for graduation, it may be completely acceptable in a classroom-based context.

It is due to these context-specific factors that it is not feasible to have a common set of guidelines dictating how evidence of technical quality should be weighed and evaluated for all tests. The collection and evaluation of evidence is always in service to a given score interpretation and specified test use.

The goal of Part 3 of this tool is to help test users identify and prioritize evidence of assessment quality in light of the assessment characteristics and features identified (in Part 2) as necessary to interpret and use the results as intended. While some forms of evidence are necessary to support all types of assessments, most are tied to the characteristics and features reflected in Part 2 of this tool. If your goal is to select or evaluate the appropriateness of an existing assessment for meeting your highest priority assessment information need, as articulated and operationalized in Parts 1-2, Part 3 can be used to inform your discussions with vendors and support decisions about the degree to provided/available evidence suggests the assessment will meet your information needs. If the goal is to develop a new assessment, Part 3 can be used to ensure the test is designed, administered and reported in a manner that supports the intended use(s).

Table 3 lists 25 assessment quality claims. These claims reflect the statements that need to be supported, by evidence, to evaluate the technical quality of the assessment and use of results as intended. If using an off-the-shelf assessment, evidence should be provided by the test developer that supports the uses and interpretations the assessment was *designed* to support as articulated in technical manuals, user's guides and other documentation. Intended uses and interpretations of results are also reflected in the types of information provided on score reports, as previously discussed. The claims

[2] However this is ultimately defined for the program.

are presented in the three categories defined in Part 2: Test Design, Test Administration and Reporting. For each claim the questions from Part 2 that inform decisions regarding the appropriateness of the evidence given your assessment information needs is also provided.

**Table 3. Assessment Quality Claims**

| **Test Design** | **Related Questions from Part 2** |
|---|---|
| **1.** The intended purpose and uses of the assessment are clearly stated. | 1-18 |
| **2.** Test items and passages align to the intended content standards or academic expectations. | 1 |
| **3.** Test blueprints reflect an appropriate distribution of content, item types and cognitive demand within forms given the *target sampling domain*. | 2 or 3 |
| **4.** Item development and review procedures and materials in place to ensure all newly developed items are fair and meet technical quality standards. | 4, 6 |
| **5.** Test development and review procedures in place to ensure forms meet the content and statistical quality requirements reflected in test blueprints and specifications. | 4-7 |
| **6.** Test forms are designed to ensure scores can be compared across forms, occasions and students in the manner intended. | 7,9,11 |
| **7.** There are appropriate accommodations in place to support the measurement of all students in the intended test taking population. | 7 |
| **Test Administration** | |
| **8.** The assessment provides for the conditions necessary to administer the test to individuals or groups of students as needed to support the intended use. | 8-11 |
| **9.** The assessment provides for the level of test security and content access deemed necessary to support the intended use of results. | 12 |
| **Reporting - Achievement Status** | |
| **10.** Achievement results can be interpreted and used as intended. | 14 |
| **11.** Reported achievement results are reliable. | 14 |
| **12.** Score reports and other resources (e.g., user's manual/interpretive guides) provide guidance to ensure reported achievement results are interpreted and used appropriately. | 14 |
| **Reporting– Prediction** | |
| **13.** The assessment was designed to support predictions of performance on the state summative assessment or other criterion measures. | 15 |
| **14.** Predicted results can be interpreted and used as intended. | 15 |
| **15.** Predicted results are reliable. | 15 |

| | |
|---|---|
| **16.** Score reports and other resources (e.g., user's manual/interpretive guides) provide guidance to ensure predicted results are interpreted and used appropriately. | 15 |
| **Reporting -Diagnostic** | |
| **17.** The assessment was designed to provide diagnostic information about student's strengths and weaknesses in the content domain. | 16 |
| **18.** Diagnostic results can be interpreted and used as intended. | 16 |
| **19.** Diagnostic results are reliable. | 16 |
| **20.** Score reports and other resources (e.g., user's manual/interpretive guides) provide guidance to ensure reported diagnostic information is interpreted and used appropriately. | 16 |
| **Reporting– Growth** | |
| **21.** The assessment was designed to provide information about student progress/growth in the content domain. | 17 |
| **22.** Student growth or progress information can be interpreted and used as intended. | 17 |
| **23.** Reported information about growth/progress is reliable. | 17 |
| **24.** Score reports and other resources (e.g., user's manual/interpretive guides) provide guidance to ensure overall achievement scores are interpreted and used appropriately. | 17 |
| **Reporting– Pedagogical** | |
| **25.** Provided links to pedagogical information are supported by evidence. | 18 |

The final section of this document outlines the key questions and sources of evidence that users should consider in relation to each claim. Questions highlighted in red reflect those focused on the appropriateness of the evidence in relation to the user's needs. (Note: This section is still under development. For illustrative purposes the first three claims have been provided in Table 4.)

**Table 4. Questions Supporting the Evaluation of Evidence (*Working*)**

| Claim | Key Questions to Pose | Potential Sources of Evidence |
|---|---|---|
| 1. The intended purpose and uses of the assessment are clearly stated. | Are the purpose, uses and inferences the assessment was designed to support clearly and consistently stated?<br>To what extent are the purpose, uses and inferences consistent with my needs? | Technical Report, User's Guides, Marketing Materials |
| 2. Test items and passages address the target content standards or academic expectations. | Are the content standards/frameworks/ expectations the assessment(s) were developed to measure clearly articulated (e.g., Common Core State Standards)? | Technical Manual, User's Guides, Marketing Materials<br><br>Test and item development and scoring procedures/ materials (e.g., item |

| | | |
|---|---|---|
| | What evidence is provided to demonstrate that items and passages align to the intended standards/frameworks/ expectations? | specifications; scoring rubrics; item writing training materials)<br><br>Passage selection (or development), review and evaluation procedures (e.g., text complexity)<br><br>Independent alignment studies. |
| | To what extent are these content standards/frameworks consistent with the state/local expectations that are my focus for assessment? | |
| 3. Test blueprints reflect an appropriate distribution of content, item types and cognitive demand within forms given the *target sampling domain*. | Is the target sampling domain (e.g., grade 3 standards; Number Sense) associated with the test blueprint clear? | Documentation summarizing the required characteristics of each operational test form with respect to content representation, cognitive demand (e.g., Depth of Knowledge), and item formats. |
| | Is the distribution of test content represented by the blueprint appropriate for making inferences about student performance in the target sampling domain? | |
| | Does each test form require an appropriate range/distribution of cognitive complexity? | |
| | Are the item types represented appropriate for making inferences about student performance in the target content domain? | |
| | Is the target domain, as defined by the test blueprint, appropriate to support my assessment information needs? | |