SELECTING HIGH-QUALITY DISTRICT ASSESSMENTS:

THE NEW YORK STATE DISTRICT ASSESSMENT PROCUREMENT GUIDANCE

APRIL 2025

NYS ED.gov

**New York State
EDUCATION DEPARTMENT**

Knowledge › Skill › Opportunity

Assessments are essential for supporting students by offering information that guides districts, schools, and educators in improving teaching and learning. To be effective, a district assessment system must be carefully designed so that the selected tools and practices provide a complete and coherent picture of students' performance and progress. New York's Educational Assessment Strategy is guided by a theory of action that prioritizes gathering reliable data on student learning from multiple sources throughout the academic year. Schools and districts are encouraged to choose the tools that best meet the needs of their teachers and students.

YOUR ASSESSMENT PROCUREMENT PROCESS:

1 Conduct an Assessment Inventory

2 Getting Clear on Use

3 Identifying Assessment Features

4 Evaluating Quality

The District Assessment Procurement Guidance was designed to help district and school teams select high-quality assessment products that address their specific information needs. This tool outlines four activities designed to inform the assessment procurement process: 1) conducting an assessment inventory, 2) clarifying information needs and the intended uses of results, 3) identifying desired assessment features, and 4) evaluating the technical quality of potential assessment options.

STEP 1. CONDUCT AN ASSESSMENT INVENTORY

An efficient assessment system collects the array of information needed to support teaching and learning with as few assessments as possible. With this goal in mind, we recommend that you document the current set of assessments operating within your district before engaging in a procurement process. This practice — often called an assessment inventory or audit —is an effective way to identify both gaps and redundancies in your existing system. Conducting an inventory allows you to identify and remove assessments that are no longer needed or appropriate and to critically evaluate the decision to introduce additional tools.

An assessment inventory should speak to the full assessment system in your district, particularly the range of information the current assessments provide and the intended purposes and uses of each. A comprehensive inventory should consider, but not be limited to:

1. Common assessments administered to all students in the district on a fixed schedule to monitor progress within a content area and inform decisions about supporting schools.
2. Curriculum-relevant assessments to evaluate student learning at different points within a cycle of instruction (e.g., unit, semester).
3. Instructionally-embedded assessment practices to inform and improve teaching and learning throughout the learning process.
4. Specialized assessments used with a subset of students to inform decisions about special education services or English language proficiency.

Several tools are available online that you can adopt or adapt to support the inventory process[1]. Regardless of the tool used, it is important to start with a general shared understanding of the different types of information needed to inform decision-making and support student learning, as discussed in Step 2.

Even with a coherent profile of assessment tools and practices in place, you may identify areas where a new or different type of assessment is needed to address a goal or priority (e.g., universal screening tools, early literacy fluency tests). This is where this structured procurement guidance can help!

[1] Options include: The Center's Audit Activity; Achieve's Assessment Inventory Tool

STEP 2. GETTING CLEAR ON USE

In a high-quality instructional materials adoption process, it is essential for the adoption committee to first establish an instructional vision. A similar process is necessary when selecting an assessment product that will address your local needs. Without this vision, there is no way to identify and address gaps, redundancies, and unneeded assessments within the existing system. Unfortunately, all too often commercial assessments are selected based on a perceived need, a referral, or a well-designed set of marketing materials, rather than a thoughtful needs analysis.

As a district leader, you must understand and articulate exactly what you are looking for and why, *before* pursuing a particular tool or vendor. This requires not only clarifying the intended uses and users of a desired assessment but also understanding how its use will positively impact teaching and learning. Identifying the intended uses and users can be done by district staff but should also be informed by feedback from school leaders and educators.

The following framework will support the process of gaining clarity on use before procurement:

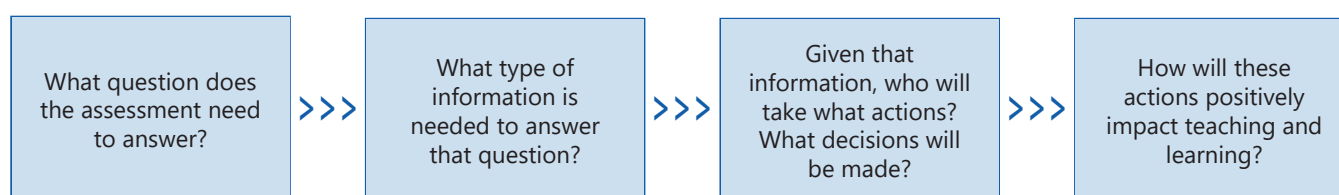| What question does the assessment need to answer? | >>> | What type of information is needed to answer that question? | >>> | Given that information, who will take what actions? What decisions will be made? | >>> | How will these actions positively impact teaching and learning? |

Table 1 below outlines a few examples of how you might use this framework to gain clarity on the information you need from assessment and its intended use. Before moving to Step 3, you should complete this table to reflect your own context and information needs.[2]

Table 1. Getting Clear on Use – Examples

	What question does the assessment need to answer?	What type of information is needed to answer that question?	Given that information, who will take what actions? What decisions will be made?	How will these actions positively impact teaching and learning?
Example 1	Are students making progress towards meeting end-of-year expectations in ELA and math?	Measures of student growth relative to the grade-level content standards	• District leaders will use this information to evaluate the effectiveness of educational programs or practices for supporting students in making progress toward end of year goals. • District leaders will use this information to make decisions about targeting resources.	• Identifying programs or practices that are not eliciting sufficient student progress will support district leaders in making improved curricular and instructional programming decisions to better support student learning. • Targeting resources where they are needed most will result in a greater positive impact on teaching and learning.
Example 2	Which K-1 students are at-risk for reading difficulties?	Student achievement of discrete skills associated with early literacy (e.g., phonological awareness, phoneme segmentation)	• Educator teams will use scores to make. decisions about additional academic interventions and supports for individual students. • School psychologists will use scores to make decisions about the need for additional evaluation of individual students for learning disabilities.	• Educators will provide students with support and interventions aligned to their needs • Additional testing will ensure students who need additional support will receive services.

[2] For additional support identifying information gaps and needs aligned to a vision of teaching and learning review Part 1 of the Interim Assessment Toolkit developed by the Center for Assessment (D'Brot & Landl, 2019).

Table 1. Getting Clear on Use – Examples (continued)

Example 3	How well did students master the standards covered in the first quarter of the academic year?	Student achievement relative to the subset of standards from the scope and sequence in the first quarter	• Educators will assign grades for students based on their assessment performance. • Educators will reflect on their instruction for the quarter in light of student performance on the assessment.	• Educator insights from their instructional reflections relative to assessment performance will improve instruction of the same content in future years.
Example 4	Which students require additional support across academic, behavioral, or social-emotional areas, aligned to multi-tiered systems of support (MTSS) framework?	Student referrals, screening and assessment data (academic, social-emotional)	• MTSS team (psychologists, social workers, teachers) will use data to identify students needing additional support, and determine appropriate levels of support and interventions (Tier 1, 2, or 3). • District leaders will use this information to make decisions about resources to enhance universal and tiered support.	• Early identification of students in need allows for timely, targeted interventions. • Tailored support helps students overcome barriers, leading to improved engagement and learning.

Reflection Questions:

- *What do you notice about the questions you generated? Do your questions tend to be at the group or individual student level?*
- *Who are the primary users of the assessment data in your table? How familiar are those users with leveraging assessment data for decision-making?*
- *Do the responses in the table above align with the greatest needs for improving teaching and learning in your district?*

Once you are clear on the intended assessment uses, you need to identify the assessment features (e.g., design, administration, reporting characteristics) necessary to support those uses.

STEP 3. IDENTIFYING ASSESSMENT FEATURES

Having a clear sense of the intended information, uses, and users of an assessment tool for supporting teaching and learning will lead to greater clarity in the features that are needed for the new assessment. For example, educators making decisions about additional instructional supports after teaching a unit, will need assessments that are closely connected to the standards for that unit. Some key questions to consider are outlined in Table 2 below[3].

Table 2. Identifying Assessment Features

In order to support the intended use and user…	Considerations/Options
1. What grade levels and content areas should the assessment product support?	• Multiple grade levels and domains • A single grade or domain
2. What type of information do you need?	• Current achievement in the assessed content domain (e.g., grade 3 math) • Predicted performance on a future assessment • Student growth throughout the year or between years in the assessed content domain • Relative areas of strength/weakness within the assessed content domain (e.g., sub-scores, performance profile)
3. What is the appropriate level of content granularity?	• A specific standard or skill • A specific sub-domain or set of standards (e.g., number sense) • Information that is relevant to a particular set of instructional materials
4. What inferences do you want/need to be able to make based on the results?	Criterion-referenced inferences describe performance relative to a defined expectation. For example: • Student is performing below/above/on grade-level in the assessed content domain. • Student is on-track to being proficient on the end of year summative assessment. • Student mastered/did not master the assessed content. Norm-referenced inferences describe performance relative to that of a defined norm group. For example: • Student scored better than X% of students in the nation/state/district/class.
5. What item types are necessary to provide the information desired? Does this differ by content area?	• Selected response • Open-ended (OE) or constructed response (CR) • Technology-enhanced items (TEIs) • Performance-based task (e.g. writing to a prompt; reading aloud; conducting an experiment) • Items that are responsive to student identity and culture
6. What accommodations should the assessment support/embed given the intended use and test-taking population?	• Four-function calculator, • Large print scientific calculator, graphing calculator • Screen reader compatibility • Color contrast capability • Sign language or video instructions • Illustration glossaries • Text-to-speech, captioning or read aloud • Native language-to-English translations or dual language options • Extended time • Customized administration time • Other

[3] For a more comprehensive table which includes additional questions, examples and considerations, see Part 2 of the Interim Assessment Toolkit developed by the Center for Assessment.

Table 2. Identifying Assessment Features (continued)

7. What level/degree of score comparability is necessary?	• Aggregate and compare the performance of students within a class • Aggregate and compare the performance of classrooms within a school • Aggregate and compare the performance schools within a district
8. Who should determine when the test is administered?	• District administrators • School leaders • Educators • Student input
9. How often should the assessment be administered?	• Once, prior to instruction • Once, at the end of a period of instruction • Pre- and post-instruction • At fixed points through a period of instruction (e.g., bi-weekly) • At multiple points throughout the school year
10. What level of test security is needed?	• Low • High
11. What resources/ supports do you need to ensure the results are used as intended?	• Integrative guides • Educator training materials • Score reports for multiple users • Instructional resources

Reflection Questions:

- *Which of the features identified above are most important for ensuring test users have the information they need?*

- *Do any of your identified features appear to conflict with one another (e.g., providing a score on the overall domain and providing information specific to a particular standard or instructional unit)?*

- *Are there any additional functionality considerations that should be added to this list (e.g., parent view for reports, transportability of data)?*

Once Step 3 has been completed, your responses can support the development of a request for proposal, identify assessment options/products that appear to address your needs (i.e., through a review of websites[4] and marketing materials), and discussions with assessment vendors about their products. The final step is to evaluate the evidence the vendors have provided to support the quality of their products and the validity of their claims.

---

[4] For example, the National Center on Intensive Intervention maintains a catalog of many commercially-available assessment tools used for screening, progress monitoring, and intervention.

STEP 4. EVALUATING QUALITY

As the final step of the District Assessment Procurement Guidance, you will engage with one or more assessment vendors that offer assessment products aligned with your needs. While not exhaustive, Table 3 outlines key documentation that vendors should be able to provide upon request and some broad criteria for evaluating the quality of that evidence. Please note that this table is only relevant for evaluating commercially available products. For more information on evaluating the quality of locally developed assessments, please refer to the *Classroom Assessment Standards*[5].

Table 3. Vendor Documentation and Quality Criteria

Claims to be Evaluated	Vendor-Provided Documentation	Quality Criteria
The assessment items and test forms align to the expectations of the targeted content standards.	**Alignment Documentation** • A study that evaluates the degree of alignment between the assessment product and the targeted content domain/standards.	• The alignment study is conducted by an independent organization with expertise in the content domain.
The assessment is fair and accessible for all students in the intended test-taking population.	**Fairness Documentation** • The assessment offers technology features (e.g., text to speech) and/or accommodated versions (e.g., Braille forms, Spanish language forms) to facilitate accessibility for students with disabilities and English learners. • Item content and item statistics are reviewed for potential sources of bias.	• The accessibility features and accommodations provided are appropriate for the intended test taking population. • There is a process in place for revising or removing items that show evidence of content or statistical bias.
There is a research agenda and validity evidence supporting the intended interpretations and uses of the assessment results.	**Technical Documentation** • A technical manual that articulates the claims and uses for the assessment is designed and validated to support. • A technical manual that identifies the intended test-taking population (i.e., grade level, language status, disability status). • A technical manual that includes information about: item and test development; test scoring, equating, and scaling; score reliability; and validity evidence supporting the interpretation and use of all reported scores/information. • Data privacy and security protocols are in place to ensure compliance with legal and ethical storage and use of student information. • An overview of the structures in place to obtain regular, independent technical feedback that supports the ongoing technical maintenance of the assessment program.	• Peer-reviewed or independently conducted research demonstrates the positive impact of using the assessment as intended. • Vendor communications about the supported uses of the assessment are consistent across marketing materials (e.g., brochures, website, sales presentations) and technical documentation. • Items are field-tested with a representative sample of students prior to operational use. • The design of the assessment reflects a clear, consistent learning theory. • Procedures used to calculate the reliability of scores/performance classifications and the associated acceptability criteria are provided for all types of reported information. • If cut scores are used to create performance levels (e.g., mastery, proficiency), the technical documentation includes information about the standard-setting procedures. • Peer-reviewed or independently conducted research demonstrates the positive impact of using the assessment as intended.

[5] For more information on these standards, visit the website for the Joint Committee on Standards for Educational Evaluation.

Table 3. Vendor Documentation and Quality Criteria (continued)

| Score reports and other resources support the interpretation and use of the assessment results. | **Score Reports & Supporting Materials**
• The functionality of the score reports supports the intended uses.
• Score reports provide guidance on how to interpret each of the different types of test results (e.g., performance levels, predicted scores, growth scores).
• Additional resources (e.g., interpretive guides, user manuals) provide guidance to support appropriate uses of test scores. | • Vendor provides evidence that the score reports have been refined based on feedback from the intended user groups.
• All reported scores include an indication of the degree of error or uncertainty.
• Interpretive guides provide clear, user-friendly guidance that specifies how each type of reported information can be used and factors that may threaten the use of results as intended. |

Reflection Questions:

- *Does the vendor's theory of action for how the assessment is designed to be used align with your district's intended use? If so, can the vendor supply independently verifiable evidence that when the assessment is used as intended, it leads to the desired impacts on teaching and learning?*

- *Does that vendor endorse a long list of potential assessment uses? If so, what technical evidence is the vendor able to share that is specific to the appropriateness of the assessment for each of those uses?*

- *Is the vendor transparent about which aspects of technical quality are supported by strong evidence and documentation and which aspects are less supported by existing evidence and documentation?*

We understand that reviewing and evaluating even the streamlined list of evidence in Table 3 can be difficult. How much evidence and of what type is good enough? Given the cost and effort of implementing a new assessment – especially something that will be administered broadly and have an impact on instruction – we recommended engaging a team of technical experts to review the information provided. This team may include your Chief Academic Officer, your district's data manager, as well as people with measurement expertise, such as representatives from your BOCES or local institution of higher education.

Additionally, a prospective vendor should be prepared to discuss the evidence provided and why they believe it appropriately supports your goals for assessment. These tools are built by measurement experts who understand how the results should and should not be used. Ask the sales representative to connect you with someone from their technical team to walk you through the evidence they have provided.

Once you've procured an assessment and have been using it, we highly recommend that you conduct periodic evaluations of how well the tool is meeting your needs and producing the intended outcomes. Even an assessment of high technical quality is worthless if it is not being used to support your district's vision for teaching and learning.

APPENDIX: DISTRICT ASSESSMENT PROCUREMENT WORKSHEET

Step 1: Getting Clear on Use

What question does the assessment need to answer?	What type of information is needed to answer that question?	Given that information, who will take what actions? What decisions will be made?	How will these actions positively impact teaching and learning?

Reflection Questions:

- *What do you notice about the questions you generated? Do your questions tend to be at the group or individual student level?*
- *Who are the primary users of the assessment data in your table? How familiar are those users with leveraging assessment data for decision-making?*
- *Do the responses in the table above align with the greatest needs for improving teaching and learning in your district?*

Step 2: Identifying Assessment Features

In order to support the intended use and user...	District Response
1. What grades and content areas does the assessment product need to support?	
2. What type of information do you need?	
3. What is the appropriate level of content granularity?	
4. What inferences do you want/need to be able to make based on the results?	
5. What item types are necessary to provide the information desired? Does this differ by content area?	
6. What accommodations should the assessment support/embed given the intended use and test-taking population?	
7. What level/degree of score comparability is necessary?	
8. Who should determine when the test is administered?	
9. How often should the assessment be administered?	
10. What level of test security is needed?	
11. What resources/supports do you need to ensure the results are used as intended?	

Reflection Questions:

- *Which of the features identified above are most important for ensuring test users have the information they need?*
- *Do any of your identified features appear to conflict with one another (e.g., providing a score on the overall domain and providing information specific to a particular standard or instructional unit)?*
- *Are there any additional functionality considerations that should be added to this list (e.g., parent view for reports, transportability of data)?*

Step 3: Evaluating Quality

Claims to be Evaluated	Vendor-Provided Documentation	Quality Criteria
The assessment items and test forms align to the expectations of the targeted content standards.	**Alignment Documentation** • A study that evaluates the degree of alignment between the assessment product and the targeted content domain/standards.	• The alignment study is conducted by an independent organization with expertise in the content.
The assessment is fair and accessible for all students in the intended test-taking population.	**Fairness Documentation** • The assessment offers technology features (e.g., text to speech) and/or accommodated versions (e.g., Braille forms, Spanish language forms) to facilitate accessibility for students with disabilities and English learners. • Item content and item statistics are reviewed for potential sources of bias.	• The accessibility features and accommodations provided are appropriate for the intended test taking population. • There is a process in place for revising or removing items that show evidence of content or statistical bias.
There is a research agenda and validity evidence supporting the intended interpretations and uses of the assessment results.	**Technical Documentation** • A technical manual that articulates the claims and uses for the assessment is designed and validated to support. • A technical manual that identifies the intended test-taking population (i.e., grade level, language status, disability status). • A technical manual that includes information about: item and test development; test scoring, equating, and scaling; score reliability; and validity evidence supporting the interpretation and use of all reported scores/information. • Data privacy and security protocols are in place to ensure compliance with legal and ethical storage and use of student information. • An overview of the structures in place to obtain regular, independent technical feedback that supports the ongoing technical maintenance of the assessment program.	• Peer-reviewed or independently conducted research demonstrates the positive impact of using the assessment as intended. • Vendor communications about the supported uses of the assessment are consistent across marketing materials (e.g., brochures, website, sales presentations) and technical documentation. • Items are field-tested with a representative sample of students prior to operational use. • The design of the assessment reflects a clear, consistent learning theory. • Procedures used to calculate the reliability of scores/performance classifications and the associated acceptability criteria are provided for all types of reported information. • If cut scores are used to create performance levels (e.g., mastery, proficiency), the technical documentation includes information about the standard-setting procedures. • Peer-reviewed or independently conducted research demonstrates the positive impact of using the assessment as intended.
Score reports and other resources support the interpretation and use of the assessment results.	**Score Reports & Supporting Materials** • The functionality of the score reports supports the intended uses. • Score reports provide guidance on how to interpret each of the different types of test results (e.g., performance levels, predicted scores, growth scores). • Additional resources (e.g., interpretive guides, user manuals) provide guidance to support appropriate uses of test scores.	• Vendor provides evidence that the score reports have been refined based on feedback from the intended user groups. • All reported scores include an indication of the degree of error or uncertainty. • Interpretive guides provide clear, user-friendly guidance that specifies how each type of reported information can be used and factors that may threaten the use of results as intended.

Reflection Questions:

• *Does the vendor's theory of action for how the assessment is designed to be used align with your district's intended use? If so, can the vendor supply independently verifiable evidence that when the assessment is used as intended, it leads to the desired impacts on teaching and learning?*

• *Does that vendor endorse a long list of potential assessment uses? If so, what technical evidence is the vendor able to share that is specific to the appropriateness of the assessment for each of those uses?*

• *Is the vendor transparent about which aspects of technical quality are supported by strong evidence and documentation and which aspects are less supported by existing evidence and documentation?*