

# CHOOSING THE RIGHT TESTS:

## *The District Assessment Procurement Protocol (DAPP)*

JUNE 2023



Preferred Citation: Landl, E., & Lyons, S. (2023). *District Assessment Procurement Protocol (DAPP)*.  
EdReports.org, Center for Assessment, Lyons Assessment Consulting.

The District Assessment Procurement Protocol (DAPP) is for you, district leaders. It supports you in selecting high-quality assessment products that serve the specific information needs of your district. This tool walks you through three activities designed to inform your assessment procurement process: 1) getting clear on use, 2) identifying desired assessment features, and 3) evaluating the technical quality of your assessment options.

Before engaging in a procurement process, we recommend that you document the current set of assessments operating within your district. This practice is commonly referred to as an assessment inventory or audit. An assessment inventory should speak to the full assessment system in your district, particularly the range of information the current assessments provide and the way each assessment is used. 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.

Even with a coherent profile of assessment tools and practices in place, you may identify other purposes for assessment that could be served by additional assessments (e.g., universal screening tools, early literacy fluency tests). This is where the DAPP can help!

## YOUR ASSESSMENT PROCUREMENT PROCESS:

- 1 getting clear on use
- 2 identifying desired assessment features, and
- 3 evaluating the technical quality of your assessment options.

### STEP 1. 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 use and user of a desired assessment, but also how that use will positively impact teaching and learning. Identifying the intended use and user can be done by district staff, but should be informed by feedback from school leaders and educators.

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

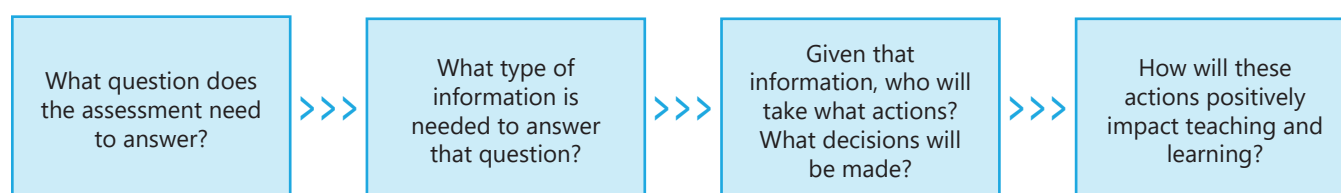


Table 1 below outlines a few examples of how you might use this framework to gain clarity on the information you need from assessment and its intended use. Before moving to Step 2, you should complete this table to reflect your own context and information needs.<sup>1</sup>

**Table 1. Clarifying Use with the DAPP Framework – 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?
<b>Example 1</b>	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	<ul style="list-style-type: none"> <li>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.</li> <li>District leaders will use this information to make decisions about targeting resources.</li> </ul>	<ul style="list-style-type: none"> <li>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.</li> <li>Targeting resources where they are needed most will result in a greater positive impact on teaching and learning.</li> </ul>
<b>Example 2</b>	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)	<ul style="list-style-type: none"> <li>Educator teams will use scores to make decisions about additional academic interventions and supports for individual students.</li> <li>School psychologists will use scores to make decisions about the need for additional evaluation of individual students for learning disabilities.</li> </ul>	<ul style="list-style-type: none"> <li>Educators will provide students with support and interventions aligned to their needs</li> <li>Additional testing will ensure students who need additional support will receive services.</li> </ul>
<b>Example 3</b>	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	<ul style="list-style-type: none"> <li>Educators will assign grades for students based on their assessment performance.</li> <li>Educators will reflect on their instruction for the quarter in light of student performance on the assessment.</li> </ul>	<ul style="list-style-type: none"> <li>Educator insights from their instructional reflections relative to assessment performance will improve instruction of the same content in future years.</li> </ul>

## STEP 2. IDENTIFYING ASSESSMENT FEATURES

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.

It is important to note that assessment characteristics are not independent of one another, and some uses dictate that certain features hold. For example, an assessment that is intended to be used to identify specific student misconceptions requires a narrow focus on the targeted content of interest. Further, many assessment characteristics conflict with one another and are difficult to maintain simultaneously (e.g., comparability of assessment results and flexibility of administration timing/conditions). A careful consideration of these constraints and trade-offs will be necessary when evaluating the features afforded by different commercial assessment options.

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



Some of the key questions to consider are outlined in Table 2 below.<sup>2</sup> Recall that the assessment features should be identified in consideration of the questions you are trying to answer and the intended use and users of the results.

**Table 2. Identifying Assessment Features**

In order to support the intended use and user...	Considerations/Options		
1. What grades and content areas does the assessment product need to support?	<ul style="list-style-type: none"> <li>• Are you looking for assessments that addresses ELA and math across multiple grades?</li> <li>• Are you focused on a particular grade and domain (e.g., grade 2, reading)?</li> </ul>		
2. What type of information do you need?	<ul style="list-style-type: none"> <li>• Current achievement in the assessed content domain (e.g., grade 3 math);</li> <li>• Predicted performance on a future assessment</li> <li>• Student growth throughout the year or between years in the assessed content domain</li> <li>• Relative areas of strength/weakness within the assessed content domain (e.g. sub-scores, performance profile)</li> </ul>		
3. What is the appropriate level of content granularity?	<ul style="list-style-type: none"> <li>• A specific standard or skill</li> <li>• A specific sub-domain or set of standards (e.g. number sense)</li> </ul>		
4. What inferences do you want/need to be able to make based on the results?	<p>Criterion-referenced inferences describe performance relative to a defined expectation.  <u>For example:</u></p> <ul style="list-style-type: none"> <li>• Student is performing below/above/on grade-level in the assessed content domain.</li> <li>• Student is on-track to being proficient on the end of year summative assessment.</li> <li>• Student mastered/did not master the assessed content.</li> </ul> <p>Norm-referenced inferences describe performance relative to that of a defined norm group.  <u>For example:</u></p> <ul style="list-style-type: none"> <li>• Student scored better than X% of students in the nation/state/district/class.</li> </ul>		
5. What item types are necessary to provide the information desired? Does this differ by content area?	<ul style="list-style-type: none"> <li>• Selected response</li> <li>• Open-ended (OE) or constructed response (CR)</li> <li>• Technology-enhanced items (TEIs)</li> <li>• Performance-based task (e.g. writing to a prompt; reading aloud; conducting an experiment)</li> <li>• Other</li> </ul>		
6. What accommodations should the assessment support/embed given the intended use and test-taking population?	<table border="0"> <tr> <td> <ul style="list-style-type: none"> <li>• Four-function calculator, scientific calculator, graphing calculator</li> <li>• Braille/refreshable Braille</li> <li>• Video sign language</li> <li>• Text-to-speech</li> <li>• Native language-to-English translations</li> <li>• Customized administration time</li> </ul> </td> <td> <ul style="list-style-type: none"> <li>• Large print</li> <li>• Captioning</li> <li>• Color contrast capability</li> <li>• Illustration glossaries</li> <li>• Read aloud</li> <li>• Extended time</li> <li>• Other</li> </ul> </td> </tr> </table>	<ul style="list-style-type: none"> <li>• Four-function calculator, scientific calculator, graphing calculator</li> <li>• Braille/refreshable Braille</li> <li>• Video sign language</li> <li>• Text-to-speech</li> <li>• Native language-to-English translations</li> <li>• Customized administration time</li> </ul>	<ul style="list-style-type: none"> <li>• Large print</li> <li>• Captioning</li> <li>• Color contrast capability</li> <li>• Illustration glossaries</li> <li>• Read aloud</li> <li>• Extended time</li> <li>• Other</li> </ul>
<ul style="list-style-type: none"> <li>• Four-function calculator, scientific calculator, graphing calculator</li> <li>• Braille/refreshable Braille</li> <li>• Video sign language</li> <li>• Text-to-speech</li> <li>• Native language-to-English translations</li> <li>• Customized administration time</li> </ul>	<ul style="list-style-type: none"> <li>• Large print</li> <li>• Captioning</li> <li>• Color contrast capability</li> <li>• Illustration glossaries</li> <li>• Read aloud</li> <li>• Extended time</li> <li>• Other</li> </ul>		
7. What level/degree of score comparability is necessary?	<p><u>Within a classroom</u></p> <ul style="list-style-type: none"> <li>• compare and aggregate the performance of students within a class</li> </ul> <p><u>Within a school</u></p> <ul style="list-style-type: none"> <li>• compare the performance of students within a school</li> <li>• aggregate and compare the performance of classrooms within a school</li> </ul> <p><u>Within a district</u></p> <ul style="list-style-type: none"> <li>• compare the performance of students within the district</li> <li>• aggregate and compare the performance schools within a district</li> </ul>		
8. Who should determine when the test is administered?	<ul style="list-style-type: none"> <li>• District administration</li> <li>• School leaders</li> <li>• Educators</li> <li>• Students</li> </ul>		

<sup>2</sup> 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)**

9. How often should the assessment be administered?	<ul style="list-style-type: none"> <li>• Once, prior to instruction</li> <li>• Once, at the end of a period of instruction</li> <li>• At the beginning and end of a period of instruction</li> <li>• At fixed points throughout a period of instruction (e.g., weekly, monthly, every 9 weeks)</li> <li>• At multiple points throughout the school year</li> </ul>
10. What level of test security is needed?	<ul style="list-style-type: none"> <li>• Low</li> <li>• High</li> </ul>
11. What resources/supports do you need to ensure the results are used as intended?	<ul style="list-style-type: none"> <li>• Interpretive guides</li> <li>• Educator training materials</li> <li>• Score reports for different users</li> <li>• Instructional resources</li> </ul>

Once each of these questions has been considered, 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 and marketing materials), and discussions with assessment vendors about their products. The next step is to evaluate the evidence they have to support the quality of their product or the validity of their claims.

### STEP 3. EVALUATING QUALITY

As the final step of the District Assessment Procurement Protocol, you will engage with one or more assessment vendors who offer assessment products that align with your needs as identified in Steps 1 and 2. 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.

To aid you in a more comprehensive review, we have created a set of ELA and Mathematics interim assessment [review tools and evidence guides](#) for 3-8 and high school that can inform your evaluation of a broad range of quality claims and associated research in each of the categories reflected below.

**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.	<b>Alignment Documentation</b> ✓ A study that evaluates the degree of alignment between the assessment product and the targeted content domain/standards. <i>Note: <a href="#">Explore Gateway 1</a> within the tools linked above for EdReports' criteria and process for evaluating alignment to college- and career-ready standards.</i>	<input type="checkbox"/> The alignment study is conducted by an independent organization with expertise in the content domain. <input type="checkbox"/> The alignment study is a two-way evaluation of alignment (i.e., content standards to items, items to content standards).
The assessment is fair and accessible for all students in the intended test-taking population.	<b>Fairness Documentation</b> ✓ 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.	<input type="checkbox"/> The accessibility features and accommodations provided are appropriate for the intended test taking population. <input type="checkbox"/> There is a process in place for revising or removing items that show evidence of content or statistical bias.

**Table 3. Vendor Documentation and Quality Criteria (continued)**

There is a research agenda and validity evidence supporting the intended interpretations and uses of the assessment results.	<b>Technical Documentation</b> <ul style="list-style-type: none"> <li>✓ A technical manual that articulates the claims and uses the assessment is designed and validated to support.</li> <li>✓ A technical manual that identifies the intended test-taking population (i.e., grade level, language status, disability status)</li> <li>✓ 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. <i>Note: <a href="#">Explore Gateway 2</a> for expectations specific to each type of reported score/information--i.e., overall achievement, predicted performance, growth/progress, or performance in sub-domain/skill areas.</i></li> <li>✓ An overview of the structures in place to obtain regular, independent technical feedback that supports the ongoing technical maintenance of the assessment program.</li> </ul>	<ul style="list-style-type: none"> <li>❑ Vendor communications about the supported uses of the assessment are consistent across marketing materials (e.g., brochures, website, sales presentations) and technical documentation.</li> <li>❑ Items are field-tested with a representative sample of students prior to operational use.</li> <li>❑ The design of the assessment reflects a clear, consistent learning theory.</li> <li>❑ Procedures used to calculate the reliability of scores/performance classifications and the associated acceptability criteria are provided for all types of reported information.</li> <li>❑ If cut scores are used to create performance levels (e.g., mastery, proficiency), the technical documentation includes information about the standard-setting procedures.</li> <li>❑ Peer-reviewed or independently conducted research demonstrates the positive impact of using the assessment as intended.</li> </ul>
Score reports and other resources support the interpretation and use of the assessment results.	<b>Score Reports &amp; Supporting Materials</b> <ul style="list-style-type: none"> <li>✓ Score reports are appropriate for the intended users (e.g., students, parents, educators, district leaders).</li> <li>✓ Score reports provide guidance on how to interpret each of the different types of test results (e.g., performance levels, predicted scores, growth scores).</li> <li>✓ Additional resources (e.g., interpretive guides, user manuals) provide guidance to support appropriate uses of test scores. <i>Note: <a href="#">Explore Gateway 3</a> for detailed expectations specific to each type of reported score/information.</i></li> </ul>	<ul style="list-style-type: none"> <li>❑ Vendor provides evidence that the score reports have been refined based on feedback from the intended user groups.</li> <li>❑ All reported scores include an indication of the degree of error or uncertainty.</li> <li>❑ 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.</li> </ul>

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 measurement expert to review the information provided.

If you do not have a psychometrician on staff, there are several other ways you can find this type of support: reaching out to your state department of education for guidance on obtaining technical assistance, connecting with regional support centers, or contacting a local institution of higher education. Engaging this expertise is worth the cost/effort to ensure you choose a high-quality assessment that will meet your needs.

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?

## 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?	



### 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.	<b>Alignment Documentation</b> ✓ A study that evaluates the degree of alignment between the assessment product and the targeted content domain/standards. <i>Note: <a href="#">Explore Gateway 1</a> within the tools linked above for EdReports' criteria and process for evaluating alignment to college- and career-ready standards.</i>	<input type="checkbox"/> The alignment study is conducted by an independent organization with expertise in the content domain. <input type="checkbox"/> The alignment study is a two-way evaluation of alignment (i.e., content standards to items, items to content standards).
The assessment is fair and accessible for all students in the intended test-taking population.	<b>Fairness Documentation</b> ✓ 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.	<input type="checkbox"/> The accessibility features and accommodations provided are appropriate for the intended test taking population. <input type="checkbox"/> 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.	<b>Technical Documentation</b> ✓ A technical manual that articulates the claims and uses 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. <i>Note: <a href="#">Explore Gateway 2</a> for expectations specific to each type of reported score/information--i.e., overall achievement, predicted performance, growth/progress, or performance in sub-domain/skill areas.</i> ✓ An overview of the structures in place to obtain regular, independent technical feedback that supports the ongoing technical maintenance of the assessment program.	<input type="checkbox"/> Vendor communications about the supported uses of the assessment are consistent across marketing materials (e.g., brochures, website, sales presentations) and technical documentation. <input type="checkbox"/> Items are field-tested with a representative sample of students prior to operational use. <input type="checkbox"/> The design of the assessment reflects a clear, consistent learning theory. <input type="checkbox"/> Procedures used to calculate the reliability of scores/performance classifications and the associated acceptability criteria are provided for all types of reported information. <input type="checkbox"/> If cut scores are used to create performance levels (e.g., mastery, proficiency), the technical documentation includes information about the standard-setting procedures. <input type="checkbox"/> 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.	<b>Score Reports &amp; Supporting Materials</b> ✓ Score reports are appropriate for the intended users (e.g., students, parents, educators, district leaders). ✓ 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. <i>Note: <a href="#">Explore Gateway 3</a> for detailed expectations specific to each type of reported score/information.</i>	<input type="checkbox"/> Vendor provides evidence that the score reports have been refined based on feedback from the intended user groups. <input type="checkbox"/> All reported scores include an indication of the degree of error or uncertainty. <input type="checkbox"/> 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.