SECTION III-4

PURPOSES, USERS, AND TECHNICAL ADEQUACY OF ASSESSMENTS: Features of early literacy assessment that reflect what we know

This chapter provides information to help districts address the needs of multiple users of assessment, who often have different purposes for assessment, within one integrated early literacy assessment system (ELAS). It describes the function of various assessment tools and practices (and their desirable inferential properties) and considers the specific components of literacy that can and should be assessed. The content provides some of the relevant explanation and backing for **Principle #4** and associated **Phase II Implementation Recommendations.**

Phase II RECOMMENDATIONS (Principle #4)

Principle #4: The ELAS must reflect what we know about the PURPOSES, USERS, AND TECHNICAL ADEQUACY OF EARLY LITERACY ASSESSMENT.

2.1: The **ELAS LEADERSHIP TEAM** should use the logic model and theory of action (called for in Phase I) to guide the selection and implementation of assessment tools and resources for inclusion in the system.

The **ELAS LEADERSHIP TEAM**, in collaboration with **PRINCIPALS AND TEACHERS**, should:

- 2.4: Select individual assessment resources on the basis of evidence of appropriate levels of technical quality with respect to validity, reliability, and fairness given the intended interpretive use(s) and the potential consequences for students:
 High-stakes judgments call for high levels of technical quality.
 - Lower stakes decisions require sufficient technical quality.
- **2.5:** Provide technical assistance and guidance to the system's various assessment users to help ensure that they can select assessment tools and practices that best meet their information needs and then use the results from assessment in appropriate and technically defensible ways.

Introduction

There are several challenges in developing a cohesive assessment system where multiple users of assessment (e.g., teachers, students, families, administrators, policymakers) use different types of assessment data for various purposes. In this section, we address these challenges and make specific recommendations for building a cohesive system, where each user understands the important decisions that other users make as well as the types and desirable properties of the assessments to make those decisions.





The science on literacy development is vast and rapidly expanding. Districts need someone in their district or consulting with their district (e.g., ISD) who has time devoted to continuing education specifically in the area of reading, and/ or writing.

In starting this discussion, we distinguish between two basic functions of educational assessment. **Assessment for learning** describes the *processes* that occur daily during instruction that help teachers plan instruction and adjust it as needed, based on student learning as it is occurring. Teachers use observational data, interviews with students, questioning, and probing to determine students' levels of understanding and to adjust instruction "in the moment" in order to nudge student learning forward. This assessment process is formative in nature. **Assessment of learning** occurs at the *end of instruction*, which may be at the end of a unit of instruction, a marking period or semester, or at the end of a school year. This assessment is summative in nature. It might consist of a state assessment, or formal tests developed and used by a teacher. The goal of summative assessment is to see where students are in the trajectory of their learning so that, if necessary, an intervention can be determined and implemented.

Both assessment functions are important but are different from each other. Used together, they provide a more balanced approach to assessment. For example, a second-grade teacher assesses students on which syllable types they can read. She uses the results of this assessment for learning to flexibly group and re-group her students for small-group, targeted instruction—a powerful tool for moving students' learning forward (e.g., Foorman, Beyler, Borradaile, Coyne, Denton, Dimino,... Wissel, 2016). Although such data serves a very important instructional purpose, it may not provide accurate information about the likelihood that a student will meet grade-level standards. On the other hand, an administrator needs information from assessment of learning about how many students might not meet grade level standards so that she can allocate resources (personnel) for supplemental intervention for those students.

This implies that we need a *system* of assessment to meet the needs of multiple stakeholders. Each stakeholder also needs to be aware of the instrumental function of assessment tools and processes used by other professionals to improve learning outcomes for students. Therefore, increasing the assessment literacy among all stakeholders is beneficial for promoting learning for all students.

Four essential factors to consider

We identify four essential factors of an assessment system that stakeholders need to consider, particularly for literacy decisions in kindergarten through grade 3:

- Users Stakeholders need data from assessment to answer the questions that are relevant to their roles and responsibilities for moving student learning forward.
- **2. Decisions** Each user must first identify the question that they are seeking to answer, before choosing an assessment or interpreting the assessment data.
- **3. Technical adequacy** In order to appropriately answer the question identified, an assessment needs to demonstrate the level of technical rigor necessary for that particular decision.
- **4. Content** Each user must know the specific domains of literacy that an assessment measures and how that domain relates to overall achievement in reading.

Users

Children and families. Young learners can actively participate in assessment for learning activities in the classroom in order to have agency in their own learning. Families typically use results from various classroom assessment activities, both formative and summative, to understand how their child is progressing in their literacy development, how much progress their child is making toward grade-level expectations, and whether or not their child needs additional support in order to meet grade-level expectations.

Teachers. Starting with an instructional plan, teachers use the formative assessment process to determine targets for student learning, the instructional strategies to be used, when (and how) levels of student understanding will be checked as the lesson unfolds, and most importantly, what changes to instruction within the lesson might be needed, depending on what students indicate they know and can do, in order to nudge learning forward. They make these decisions in the moment of teaching to provide individualized feedback, for the next week when they need to re-teach certain concepts or re-assign students to different small groups. Because teachers have so many different decisions to make, including requesting further intervention for their students, they need a large range of assessment tools and practices depending on each specific purpose.

Literacy specialists or intervention teachers. Typically, students work with literacy specialists or intervention teachers after they have already been identified as needing support based on data from a district-selected assessment. Specialists need to ascertain the instructional needs for individual students via diagnostic assessment and then ensure that students receive differentiated intervention based on the diagnostic information. Specialists also use assessment to determine whether students have learned what is taught.

Other specialists in schools bring a wealth of expertise to the school to promote students' literacy development. School psychologists have a deep understanding of the uses and limitations of assessment for identifying which students need additional intervention (at Tier 2, Tier 3, or within special education). Social workers and school psychologists can also assess other factors that may impact student literacy development and recommend individualized adjustments to literacy instruction for students. Special education teachers and speech and language pathologists have extensive literacy backgrounds that can be very useful for guiding school-level curricular and instructional decisions as well as problem solving for individual students.

Administrators and school leadership teams. An important role for leaders in the school (we include decision making teams in this definition of leaders) is to make sure that students in their buildings/districts are making progress towards meeting state and district standards and that resources are allocated appropriately to best meet the building's or the district's goals. Historically, leaders interpret state assessment data and other data in grades 3 through 12 to understand student progress more broadly (i.e., compared to students in prior years, to students in other schools or districts, to classrooms that are making more or less growth, and/or to other students in the state). It is important to provide the instructional resources to the areas identified

through these data systems; however, it is more important to prevent those issues through increasing resources to building literacy in kindergarten through grade 3. Assessment processes described in this Guide (e.g., initial, extensive, progress monitoring, formative assessment process) can guide instructional decisions that have three times the impact on student literacy outcomes in kindergarten through grade 2 as the impact of instruction in later grades (Scammacca, Fall, & Roberts, 2015).

Additionally, leaders have the responsibility to use assessment to determine whether the systems-level decisions they make for their school or district are working. These leaders must also assess the implementation of their systems before they can determine which practices at their school worked or did not work.

Policymakers. The development of literacy has long been a public health initiative. When students are provided with high quality early literacy experiences (i.e., ages 4 through 8), society benefits from higher rates of high school graduation, lower incarceration rates, higher levels of employability, and improved life satisfaction (Allensworth & Easton, 2005; Balfanz, Bridgeland, Moore, & Fox, 2010; Hernández, 2012). Furthermore, when reading difficulties are identified early enough and appropriate instruction is provided in the early elementary grades, the impact of those difficulties later in life is greatly reduced and the higher costs of more intensive intervention later is prevented (e.g., Blachman, Schatschneider, Fletcher, Murray, Munger, & M. Vaughn, 2014).

Local and state policymakers play an important role in assisting educators to work successfully to provide students with needed literacy resources and instruction. Policymakers can provide human, financial, or technology resources to educators; they also can adopt policies that will support systematic administrator, teacher, or parent activities. For example, the State of Florida provided a free high-quality reading screening and diagnostic assessment statewide along with highly qualified reading coaches in every elementary school, extended training for all elementary teachers in reading instruction, and statewide technical support. Following those efforts, the percentage of students reading at grade level increased and the percentage of students at high risk of failing to meet standards decreased (Foorman, Petscher, Lefsky, & Toste, 2010).

Decisions

Another significant challenge with developing a cohesive assessment system stems from the competing demands of collecting enough information to make the informed decisions needed to support student literacy development while at the same time minimizing the time spent in assessment that could potentially reduce valuable instructional time. In well-meaning efforts to reduce assessment time, some assessment scores are used for purposes for which they were not designed, resulting in equally undesirable outcomes.

For example, teachers are often given data reports that are designed to indicate which students have made progress in their overall reading abilities and are told to make decisions from the data. However, this type of data provides limited information for the types of decisions that teachers need to make. When used for the purposes for which

they were designed, assessment practices lead to effective instruction that improves student outcomes (Graham et al., 2012; Hamilton, Halverson, Jackson, Mandinach, Supovitz, & Wayman, 2009). However, administering assessments without first making plans for how assessment information will inform instructional decisions can lead to wasted time and other resources, as well as inappropriate decisions.

Before collecting assessment data, educators need to know 1) what **decision(s)** will be made based on the data, and 2) which specific **score type(s)** from the assessment is validated for that decision. It can be difficult to articulate the questions/decisions that are being made based on data, and many times educators ask questions of the data that cannot be answered.

To help with identifying this information, the most relevant instructional decisions are provided in **Table III-4.1**. Those listed are also supported by research as having a positive impact on student outcomes. Illustrations of assessment to answer these questions are embedded in the **Portraits** under the headings "Assessment" and "Using Data to Inform Instruction."

Identifying the right question

Questions answered through the **formative assessment process** for information used daily by teachers during instruction include:

- Are students learning what is being taught?
- What instructional adjustments are needed? For which students?
- What instruction is needed next for each student?

Questions answered through **student assessment collected periodically** and used by school teams to make instructional changes:

- Which (and how many) students achieved and did not achieve grade-level proficiency standards?
- Which students (and how many) are at risk for not meeting the grade-level proficiency standards; thus, need additional instructional support?
- What do the students in the school know and what are the ongoing learning needs and interests of students in the school?
- For which specific literacy skills do students need support through smallgroup instruction or supplemental/Tier 2 intervention or intensive/Tier 3 intervention?
- Are students making progress toward meeting end-of-year expectations?
 Who needs more intensive intervention?
- Does this student have a learning disability or other disability that impedes learning?

Questions answered through **periodic assessment of the school's processes** by school teams:

- Are the assessment and intervention systems at our school effective for most students?
- Is instruction being implemented as intended or do we need to provide more support to implement effective practices for students?

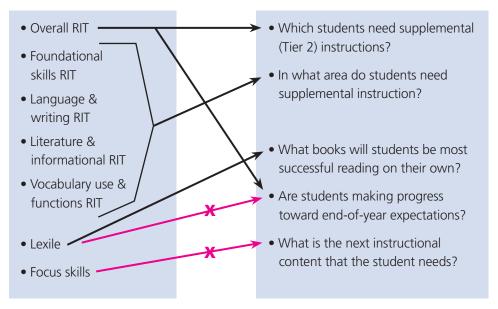


Instead of posting student data on the wall, write the decision to be made/question to be answered in a prominent location for a reference point.

Matching scores to decisions

The next step is to identify which scores match each decision and which user needs each type of score. Commercial assessment tools are constantly evolving and striving to address more of the decision points listed above. It can be a challenge for schools to stay current with the research indicating which instructional decisions can be accurately associated with each type of score. As an example, the NWEA MAP assessment reports multiple scores (Overall RIT, Foundational skills RIT, Language and Writing RIT, Literature and Information RIT, Vocabulary Use and Functions RIT, Lexile, and Focus Skills), each of which is designed and validated for different purposes for different users. However, many common uses of some scores do not have research support. Examples are provided in **Figure III.4.1** of appropriate uses of scores (marked by arrows) and misuses of scores (marked by **X**).

● FIGURE III.4.1 Example Matching Score Types from the NWEA MAP to Decisions



Technical adequacy

In any educational assessment, there is some degree of error that affects the obtained score a student receives on a test. Assessments cannot be 100% accurate at capturing a student's true learning or knowledge level because assessment results represent only a sampling of the student's behavior, knowledge, or skill. That is, the score the child obtains is an estimate of their true skills in the area assessed plus error resulting from various sources.

Error is introduced from two primary sources: random and systematic. **Random error** is introduced when an assessment results in inconsistent scores across time, across different forms of the test, or across items within a test. **Systematic error** often results from the test design itself. If there is a certain feature of a test that systematically and consistently under- or over- estimates a student's true ability, that test feature leads to systematic error in the obtained score. For example, if a vocabulary test designed to measure students' vocabulary knowledge (breadth of vocabulary) includes items that are culturally-dependent (e.g., Hanukah), the obtained score may represent a different construct (e.g., cultural knowledge) than what was

intended (e.g., vocabulary knowledge). Both random error and systematic error can be estimated in carefully designed studies of assessment. In these studies, the degree to which random error is controlled in an assessment is called reliability. The term validity is used to describe the degree to which systematic error in the interpretation of a test is controlled.

Reliability refers to the consistency with which an assessment provides the same information about the same student, regardless of the time the student is assessed or if different forms of the assessment are given. An assessment cannot be valid without being consistent; therefore, reliability is necessary before validity can be evaluated. Reliability is reported on a scale of 0.00 to 1.00. A reliability of 0.50 means that the assessment is about as reliable as the flip of a coin.

There are different types of reliability reported for different assessment tools. These different types of reliability are included in the glossary of this Guide. Reliability information can be found in technical documentation for an assessment and at the National Center for Intensive Intervention (NCII) (https://intensiveintervention.org/). The NCII provides an independent evaluation of the reliability, validity, and fairness (i.e., bias) for many commercial screening (initial) and progress monitoring assessment tools.

Validity describes the degree to which theory and evidence support the suggested interpretation of assessment data. Validity is not a property of a test, per se; rather, it is the human interpretation of the assessment data that is valid or not. Thus, it is important to understand for which uses an assessment was validated (that is, for which uses is supportive information available)? This should be clearly stated in technical documentation. In such technical documents, assessment authors describe

COMMON MISPERCEPTION: GRADE LEVEL

One of the most common pieces of information that users want from literacy assessment is the student's estimated grade level of reading. If a user is asking this question, it is absolutely critical at this juncture to determine what decision the user wants to make based on that information. Often, users want to use grade level to do one of the following:

- Describe how far above or below a student is from their current grade level
- Measure growth
- Group students for instruction

Although assessment would be much more intuitive to use if grade-level information worked this way, grade-level information has NOT been validated for any of the three purposes listed above (e.g., Parker, Zaslofsky, Burns, Kanive, Hodgson, Scholin, & Slingbeil, 2015). It is important to keep in mind that the grade level reported has one purpose: to match students to the level of text they will likely read successfully when they are reading independently. However, when choosing texts for students to read independently, users should also keep in mind that information about a student's interest in the topic area of the text is more important for helping students choose books to read independently (Renninger & Hidi, 2011).

the construct that the assessment is designed to measure (i.e. theory) and then report the correlation between their assessment and another well-established gold standard assessment, such as a state achievement test (i.e., the supportive information). Just like reliability, validity is on a continuous scale of 0.00 to 1.00, with estimates between 0.50 and 0.70 being most common. It is important to note that different types of validity correlations are needed for different types of decisions. (See the glossary for more information on each type of validity estimate.) Different levels of reliability and validity evidence are required for different decisions.

A very important aspect of validity is associated with the consequences for students or others of using the results—consequential validity. Users must investigate both positive/negative and intended/unintended consequences of the inferences made based on an assessment result. If the assessment result is used to design instruction and leads to improved literacy development, the assessment has high positive consequential validity. On the other hand, assessment has little or negative consequential validity if the assessment results cannot be used to adjust instruction, were not used to inform further assessment, or had an adverse impact on other outcomes. Again, note here that validity is not a property of the test, but is associated with the decision made based on the results. Therefore, it is incumbent upon users to evaluate whether decisions result in positive or negative outcomes, intended or otherwise.

Fairness. Relatedly, use of assessment data may not be fair to a certain group of students if used without validity evidence. Assessment can be misused if it is systematically biased toward certain groups of students, or if the assessment data is not used as intended. This brief list demonstrates a few concrete steps that schools can take to increase fairness in their use of assessment.

- Carefully evaluate if the decisions that will be made based on this assessment align with the intended purpose of the assessment.
- Select tools for which there is documentation of the steps taken to assure fairness (that is, assure that the assessment is not biased towards any group of students). For example, it is critical to ensure that a broad range of students and educators from a wide variety of backgrounds are part of the development, review, and field testing of the assessment. This can include formal fairness reviews by experts in detecting bias and the use of statistical procedures for detecting bias¹.
- Request results of the steps taken by assessment vendors to assure the fairness of their assessment tools.
- Check the demographics of the norm groups from the technical manual.
 The norm group or comparison sample should contain a significant and roughly proportional number of students in each demographic category of the students found in the school (e.g., racial-ethnic, socio-economic status, English learner population, and special education status category).
- Higher stakes decisions should be based on the triangulation of several data points. This usually means integrating results from two or more assessment tools in addition to data from teacher observations or examination of students' work.

¹ One common statistical procedure for detecting bias is differential item functioning. It should be noted that very few assessment tools have conducted and publicly published the results of DIF studies (as well as the steps taken to review items where DIF is detected), a shortcoming of many assessment tools.

Related to considerations of reliability, validity, and fairness, the most important implication is that **high-stakes decisions**, such as retention in grade, should never be made based on only a single test score (AERA/APA/NCME, 2014; Snow, Griffin, & Burns, 2005). A decision is high stakes when the consequences of an inaccurate decision are very high for the students involved.

- High stakes decisions require the highest levels of reliability, validity, and fairness, as well as multiple assessment data that support the same conclusion.
- Moderate stakes decisions, such as determining which students need small-group intervention outside the typical classroom, require slightly lower reliability, because errors in placement can be readily observed and corrected without consequence to the student.
- **Lower stakes decisions,** such day-to-day instructional decisions, may not require formal evidence of reliability, validity, or fairness.

As demonstrated in the **Portraits**, multiple data points were used with increasing stakes of assessment. Furthermore, the primary data points used for decisions were commensurate with the level of technical adequacy of the data point. For example, Mr. Ahmed used learning checks to create small groups in his class and then regrouped his students after Emmanuel mastered the text features component. Flexible grouping of students during instruction does not require high levels of technical adequacy in assessment, and learning checks are powerful tools for moving learning forward when used in this way. For the decision to have Ayesha receive more individualized instruction, Ms. Robins used assessments that had higher levels of technical adequacy (e.g., the benchmark assessment) and were based on several data points (across grades 1 and 2).

Information on the reliability, demonstrated validity for specific decisions, and fairness of assessment tools should be provided by assessment vendors to educators via technical manuals and literature that describe these technical characteristics in accessible language. Standards for the levels of reliability, validity, and fairness for many moderate stakes decisions have been set by the National Center on Intensive Instruction (https://intensiveintervention.org/) and are the standards adopted for this Guide.

Table III.4.1 provides a crosswalk between the educational decision, typical type(s) of assessment used, the level of technical adequacy required from the assessment to make the decision, and the users who are likely to make those decisions. These specific questions/decisions were chosen based on research studies indicating their utility for moving learning forward. Many of these decisions are also included in federal and Michigan policy. In the second column we match the decisions with the assessment type that educators typically use. Many educators will name the decisions by assessment type (e.g., initial, extensive, benchmarking, progress monitoring). We encourage educators to use the "decision/question" instead of naming the general type of the assessment to increase clarity and reduce confusion in data meetings.

"Standards for the levels of reliability, validity, and fairness for many moderate stakes decisions have been set by the National Center on Intensive Instruction."

■ TABLE III.4.1 — Decision/Question, Assessment Users, Assessment Types, and Required Level of Technical Adequacy

Decision/Question	Assessment Type	Required Level of Technical Adequacy	Assessment Users
Tier 1 (All Students)			
Determine students' ongoing learning needs, interests, and strengths and facilitate learning	Criterion-referenced measures, which can serve to inform areas for re-teaching or investigating pre-requisite knowledge.	Content validity at Level 4 in Figure III.4.2	Primary user: Teachers make day-to-day instructional decisions about literacy skills on which to focus instruction
How much are students learning from instruction?	Information collected by teachers during instruction using the		for individual or groups of students.
Where is instruction going and how do we close the gap between where the student is and their learning target?	formative assessment process. Illustrated by the observations portion of the Portraits.		Other users: Students use feedback from the formative assessment process to adjust their learning strategies.
Determine proficiency	Summative assessment	Reliability > .90	Primary users: Administrators
Who achieved the content?	State assessment	Content Validity	– for accountability & resource allocation
Who is proficient?	National criterion or norm- referenced tests		Policymakers Other users: Parents/guardians
Determine student achievement	Summative assessment embedded in the curriculum such as quizzes	Content validity with	Primary user: Teachers and coaches making day-to-day
Can the students perform the curriculum/grade-level standards?	& unit tests Free and Very Low Cost Assessment list	overall reading achievement	instructional decisions about what students learned from the instruction.
Are there areas that need to be reviewed or are there areas that need to be further explored?	Illustrated by the lesson checks in the Portrait.		Other users: Students, Parents
How should students be grouped for the language and literacy block?			

Decision/Question	Assessment Type	Required Level of Technical Adequacy	Assessment Users
Determine risk status for meeting end-of-year expectations Which students do and do not need additional support to meet end-of-year expectations? Which students need increased intensity of intervention (Tier 2, Tier 3, special education services)?	Interim or benchmark assessment (also referred to as screening or initial assessment) that provide a "risk score" that is standardized, norm-referenced at national, state, or local level and predicts reading comprehension (Figure III.4.2 level 1) Illustrated in the Portraits by the computer adaptive benchmark assessment.	Reliability > .80 Predictive validity > .60	Primary user: Administrators – for resource allocation School leadership team & teachers – determining placement in standard protocol interventions

Tier 2 decisions (assessment used with students scoring below a cut-point on an initial assessment)

Her 2 decisions (assessment	used with students scoring below	a cut-point on a	an initial assessment)
Determine primary areas for instruction for students who need more support What primary components of literacy do the students who need support to meet end-of-year expectations need to be taught?	Assessments that provide information about students' strengths and weaknesses in their knowledge relative to the subcomponents of literacy (Level 2 content in Figure III.4.2). Some interim assessment tools provide this extensive information alongside the initial information. Illustrated in the Portraits by the early reading, word reading, and language comprehension results, as well as the assessment conducted by the literacy specialist.	Reliability > .60 Concurrent Validity > .60	Primary users: School leadership team and teachers to determine placement in standard protocol interventions Teachers to make instructional grouping decisions.
Determine learning progress Are students in supplemental (Tier 2) intervention making progress toward meeting expectations?	Interim or benchmark assessment occurring in winter and spring. Either initial information (risk score indicating Level 1 in Figure III.4.2) or extensive information (Level 2 content in Figure III.4.2) can be used. Illustrated in the Portraits by the computer adaptive benchmark assessment administered in January and May.	Reliability > .60 Concurrent Validity > .60	Primary users: School leadership team & teachers – are interventions generally effective & which students need more intensive instruction?

Decision/Question	Assessment Type	Required Level of Technical Adequacy	Assessment Users
Tier 3 decision (assessment f	or students receiving intensive in	struction)	
Determine learning progress of students who need more support Are students in intensive (Tier 3) intervention and special education making progress toward their individualized learning goals? Does a student have learning disabilities in areas related to literacy?	Summative assessment and progress monitoring assessment with equivalent, alternate assessment forms. These types of assessment provide information regarding students' progress toward mastering the component skills in level 3 or 4 in Figure III.4.2. Illustrated in the assessment that Ms. Robins administers to Ayesha during What I Need time. Assessment tools used to identify students for a learning disability need to be standardized and norm-referenced. Should occur at least monthly.	Alternate-form Reliability > .70 Slope reliability > .40 Slope predictive validity > .40	Primary users: School leadership team, reading specialists, school psychologists, and special education teachers – are interventions generally effective & which students need more intensive instruction?
	For special education eligibility decisions, more frequent brief assessment may need to occur to reach the requisite minimum of 12 data points.		

Content

When designing an assessment system, users need a depth of knowledge about the development of literacy and which components of literacy need to be measured at specific times in order to maximize the impact of instruction. Reliable and valid assessment of all components of literacy is not feasible due to time and resource constraints. Therefore, assessments must reflect a small sample of the target literacy domain. As a result, the interpretation of why students perform the way they do on an assessment can be dangerous, either by leading the teachers to concentrate on the inevitably limited definition of the domain reflected in the test, or by leading to misattributions based on insufficient information. Therefore, users should understand which content domains an assessment does and does not measure.

Assessment tools must be designed to measure the domains that predict success on later literacy outcomes *and* are malleable (can be changed by instruction occurring in schools). As demonstrated earlier, reading is a complex process, and having a deep

understanding of the components of reading is needed to promote students' learning. The domains of reading included in assessment depend completely on the decisions that will be made based on the assessment. Some decisions require more specific information than do others. For example, when a grade 2 teacher needs to make decisions about grouping students and determining day-to-day instruction for teaching vowel teams, she needs more detailed information about the vowel teams that her students have mastered. This specific information about vowel teams may or may not represent how students are achieving in their overall reading. Therefore, an assessment of a larger grain size that is an indicator of overall reading achievement in grade 2 (for example, oral reading fluency) would be needed to answer that particular question.

In **Figure III.4.2** we demonstrate how each of these domains of reading fit together and list some example measures of those domains. This is not a comprehensive list of constructs that impact reading. Note that the domains in this figure are all domains that (a) can be assessed, (b) have been shown to predict important outcomes in K through grade 3, and (c) are malleable in K through grade 3 (Connor, Spencer, Day, Giuliani, Ingebrand, McLean, & Morrison, 2014; Foorman, Herrera, Petscher, Mitchell, & Truckenmiller, 2015; Foorman, Petscher, Stanley, & Truckenmiller, 2017).

Larger-Grained Domains to Finer-Grained Domains

Reading Comprehension

(e.g., M-STEP, ITBS, SAT10,TERRA-NOVA, Composite scores of NWEA MAP, iReady, Lexia RAPID, STAR)

Oral Reading Fluency

(e.g., ORF or Maze from AIMSweb, Acadience, DIBELS, EasyCBM, FastBridge)

Word Reading/ Decoding

(e.g., Nonsense Word Fluency, Word Identification Fluency, Decoding subtest scores from A2i, iReady, Lexia RAPID)

Oral & Written Language Comprehension

(e.g., Language subtest scores from A2i, Lexia RAPID)

Phonological Awareness Orthographic Knowledge

Morphological Awareness Vocabulary

Knowledge of Sentence/Text Structure & Features Inference Making & Strategy Use

(Subtest scores from Map Growth, iReady, and RAPID; Assessments on the Free or Very-Low Cost Assessment List; Subtest scores from achievement batteries (e.g., Woodcock Johnsonn))

• FIGURE III.4.2

Example content & measures for the decisions described in Table III.4.1

Note: This graphic represents only a partial list of all constructs that impact reading. The assessment tools listed are examples; they do not represent the full range of options districts have available.

In **Table III.4.2,** we demonstrate how each specific reading domain content is aligned with specific decisions/questions and provide example assessments of those domains.

■ TABLE III.4.2 How different grain sizes of reading domain information are needed to meet different purposes

Reading domains (larger to smaller grain sizes)	Decision/questions	Example assessments
General reading achievement	Question: Were students supported enough to achieve expectations? Decision: Where to devote more school resources.	M-STEP (or standards-based assessment) For grades K-2: ITBS, TERRA-NOVA, SAT10 Each example test is a standardized, nationally normed test of reading achievement with internal reliability > .90
General reading achievement	Question: Which students do and do not need additional support to meet end-of-year expectations? Decision: To whom to provide Tier 2 instruction	The composite score of some computer adaptive screening assessments are standardized, nationally normed assessments that have internal reliability > .80 and predict one of the assessments listed in the row above > .60. They also have slope reliability > .40 for measuring growth across 3 times per year or monthly. Examples include NWEA MAP, iReady, Lexia RAPID, STAR The fluency rate of some Curriculum-Based Measurement screening assessments in standardized, nationally normed assessments that have parallel form reliability > .80 and predict one of the assessments listed in the row above > .60 and most have slope reliability > .40 for measuring growth weekly. Examples include AIMSweb, DIBELS Next, EasyCBM, FastBridge

Reading domains (larger to smaller grain sizes)	Decision/questions	Example assessments
Decoding and language comprehension	Question: In which main area do students need supplemental instructional time? Decision: selecting Tier 2 interventions for groups of students.	Many computer adaptive assessments measure decoding and language comprehension. A2i, iReady & Lexia RAPID provide subtest scores for decoding and language comprehension. Curriculum-Based Measurement assessment systems measure decoding and need supplemental assessment to determine language comprehension.
Phonological awareness, orthographic knowledge, fluency, vocabulary, sentence structure, text structure, comprehension	Question: Why is a student struggling with reading? Decision: selecting Tier 3 intervention or individualized education plan (IEP) goals for individual students.	The subtest scores on computer-adaptive assessments like MAP Growth, iReady, and RAPID provide information about several, but not all of the domains. Assessments on the Free or Very-Low Cost Assessment list Subtest scores from various academic achievement batteries (e.g., Clinical Evaluation of Language Fundamentals, Woodcock Johnson Test of Achievement)
Each of the reading domain areas listed in the section above	Question: Where are the specific opportunities for learning progress day to day for individuals? Decision: content to re-teach and provide more practice; grouping students for instruction	Quizzes, unit tests, curriculum-embedded assessment, spelling inventories, informal reading inventories, assessments on the Free or Very-Low Cost Assessment list



Tools/Resources for PHASE II, Principle #4:

Human resource recommendation

The science on literacy development is vast and rapidly expanding. Districts need someone in their district or consulting with their district (e.g., ISD) who has time devoted to continuing education specifically in the area of reading, and/or writing.

Formative Assessment for Michigan Educators (FAME)

FAME is a professional learning initiative sponsored by the Michigan Department of Education (MDE) that promotes teacher collaboration and planning for effective formative assessment practice. A cadre of Michigan educators serves as coaches for site-based learning teams of teachers and administrators in Michigan schools.

Learn more at www.FAMEMichigan.org.

National Center on Intensive Intervention (NCII)

The NCII provides an independent evaluation of the reliability, validity, and fairness (i.e., bias) for many commercial screening (initial) and progress monitoring assessment tools. Their six <u>Tools Charts</u> assist educators and families in becoming informed consumers who can select academic and behavioral assessment tools and interventions that meet standards for technical rigor and address their specific needs.

Learn more and explore the resources at https://intensiveintervention.org.

The Standards for Educational and Psychological Testing (AERA/APA/NCME)

This complete set of professional standards for assessment should be met in the design, development, implementation, use, reporting, and analyses of assessments used for all purposes. They are a product of the American Educational Research Association, the American Psychological Association. and the National Council on Measurement in Education. They have been published collaboratively since 1966 and represent the gold standard in guidance on testing in the United States and in many other countries.

Available for purchase online at https://www.apa.org/science/programs/testing/standards.

Understanding Screening: What Do the Technical Standards Mean? (NCII, 2019)

The National Center for Intensive Intervention (NCII offers five one-page documents that provide a brief overview of each standard (validity, reliability, classification accuracy, statistical bias, and sample representativeness) used on the NCII Screening Tools Charts. The one-pagers include a definition, examples, and information on why each particular standard is important for understanding the quality of screening tools.

Available at https://intensiveintervention.org/resource/screening-standards-overviews.

Using Student Achievement Data to Support Instructional Decision Making (IES Practice Guide/What Works Clearinghouse, 2009)

This resource is designed to help schools understand the role of assessment in instructional improvement.

Available at https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/dddm_pg_092909.pdf.

Find all Tools and Resources at www.MichiganAssessmentConsortium.org/ELAS.

Notes

Formative Assessment Process: Assessment for Learning

The Michigan Department of Education has noted the importance of the formative assessment process in teaching and learning and has adopted the following definition developed by the Council of Chief State School Officers (CCSSO).

"Formative assessment is a planned, ongoing process used by all students and teachers during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended disciplinary learning outcomes and support students to become more self-directed learners."

(CCSSO FAST SCASS, 2017)

Effective use of the formative assessment process requires students and teachers to integrate and embed the following practices in a collaborative and respectful classroom environment:

- Clarifying learning goals and success criteria within a broader progression of learning;
- Eliciting and analyzing evidence of student thinking;
- Engaging students in self-assessment and peer feedback;
- Providing actionable feedback to students; and
- Using evidence and feedback to move learning forward by adjusting learning strategies, goals or next instructional steps.

Table III.4.3 shows Sadler's three questions as well as the components and elements of formative assessment used in Michigan's Formative Assessment for Michigan Educators (FAME) professional learning program which helps educators learn about, learn to use, and reflect and improve their use of the formative assessment process (Sadler, 1989).

The formative assessment process, often referred to as assessment for learning, is "...embedded in the ongoing flow of activity and interactions in the classroom"

(Heritage, 2019)

■ TABLE III.4.3

Michigan Formative Assessment for Michigan Educators (FAME) Components and Elements

Guiding Questions	FAME Components and Elements
Where are we (teacher and students) going?	Planning 1.1—Instructional Planning: planning based on knowledge of the content, standards, pedagogy, formative assessment process, and students. Learning Target Use 2.1—Designing Learning Targets: the use and communication of daily instructional aims with the students 2.2—Learning Progressions: connection of the learning target to past and future learning 2.3—Models of Proficient Achievement: examples of successful work for students to use as a guide.
What does the student understand now?	Eliciting Evidence of Student Understanding 3.1—Activating Prior Knowledge: the opportunity for students to self-assess or connect new ideas to their prior knowledge 3.2—Gathering Evidence of Student Understanding: use of a variety of tools and strategies to gather information about student thinking and understanding regarding the learning targets from all students 3.3—Teacher Questioning Strategies: the intentional use of questions for students to explain their thinking or to connect their idea to another student's response 3.4—Skillful Use of Questions: a focus on the purpose, timing, and audience for questions to deliver content and to check students' understanding
How do we (teacher and students) get to the learning target?	Formative Feedback 4.1—Feedback from the Teacher: verbal or written feedback to a student to improve his or her achievement of the learning target 4.2—Feedback from Peers: feedback from one student to another student about his or her learning in relation to a learning target 4.3—Student Self-Assessment: the process in which students gather information and reflect on their own learning in relation to the learning goal. Instructional and Learning Decisions 5.1—Adjustments to Teaching: teachers' daily decisions about changes to instruction 5.2—Adjustments to Learning: students' use of feedback for improvement.

Attachments

Attachments A and B illustrate the application of the formative assessment process by the teacher with the students in the second-grade classroom depicted in the **Portraits.**

Attachment A is the planning template that the teacher completed before teaching the lesson. It provides information on how the teacher planned the lesson, when the formative assessment process elements would be applied, and how the teacher planned to collect information on student understanding during the lesson so as to move instruction and student learning forward.

Attachment B shows how the formative assessment process was implemented in the lesson. It indicates when both the FAME components and elements and the *Essential Instructional Practices in Early Literacy: Grades K to 3* (MAISA/GELN/ELTF, 2016) were used during the lesson illustrated in the vignette.



Attachment A: Formative Assessment Planning Template

Feedback Planning



DATE

What am I teaching? [State Standard(s)

- RL.2.3. Describe how characters in a story respond to major events and challenges.
- SL.2.1b. Build on others' talk in conversations by linking their comments to the remarks of others.
- SL.2.1c. Ask for clarification and further explanation as needed about the topics and texts under discussion.

How can I make this clear to students? [Student-Friendly Learning Target(s)

RL.2.3

Use evidence from the text to prove what I know about my character.

Provide a brief description of how students know that they've met the learning targets.

Using evidence from the text, students will list on three sticky notes what they already know about their character. The first sticky note will be labeled Always (3 or more items), the second Sometimes 2 or more items) and the third OMG (1-2 items). I will model this for students

How will I know if they understand the learning target? (Mode of Assessment & Student Evidence)

⊠ Product

⊠Conference

⊠ Observation

(Check all that apply.)

What strategies will be used to gather evidence of student understanding?

I will use self- assessment and goal setting through the use of conferring and student reading bookmarks. I'll use activating prior knowledge through strategic questioning and student turn and talks.

How will I teach students? (Instruction)

I will start with activating prior knowledge of common text. I'll model the new learning target with lots of student input.

What curricular resources will I need?

Common text for whole class model; sticky notes for my model to display on doc camera; reading goal bookmarks; book club books

How will they practice before the assessment?

During my whole class model, students will turn and talk with a partner and add ideas to our sticky notes. They will also check in with their book club partners and share two things they all know about their character that will go on their own sticky notes

How much time should I plan for instruction <u>and</u> practice?

Whole class with embedded practice: 15 min

Individual work time: 20 minutes Small group book clubs: 15 minutes



Attachment A (side 2)

Feedback Planning



How will I involve my students in the process of assessment? (Formative Strategy)					
What tool(s) will I use? Student book marks, sticky notes, o	bservation				
What feedback will I give a are learning and being ass			idents have the o use the feedback?		
☑ Verbal □	Written		feedback right after the conferring an use feedback from their book the moment.		
Possible Misconception Students might describe their character Students might summarize the whole	cter's physical appea				
Ů	ght I begin thinking				
Idea #1 After my first whole group demonstration, I will invite "confused" students to stay and work with me until they are ready to work independently.	I will point out and r demonstration mod focusing on specific actions. Then, durir small group time, I'l with correct models work and thinking w needing more assis	el that it is about c character ng individual and Il invite students to share their vith students	Idea #3 I'll use the bookmark tool at the conclusion of the lesson along with the sticky note to gather evidence of student understanding. Then, I'll use that to inform and adjust my teaching.		



Attachment B: FAME Formative Assessment Process Applied in the Grade 2 Portrait

What follows is a sample vignette showing the formative assessment process in a second-grade classroom. The left column addresses the Formative Assessment for Michigan Educators (FAME) Components and Elements of the lesson, and the right column addresses the relevant "essential instructional practice" developed by the Michigan Association of Intermediate School Administrators (MAISA) General Education Leadership Network (GELN) Early Literacy Task Force (ELTF). Michigan K-3 educators are charged with using these *Essential Instructional Practices in Early Literacy: Grades K to 3* (MAISA/GELN/ELTF, 2016) and are supported in their use by a program of professional learning.

In the vignette, the teacher engages in instruction that aligns with the formative assessment process as well as the *Essential Instructional Practices in Early Literacy: Grades K to 3* (MAISA/GELN/ELTF, 2016). In particular, during this lesson, the teacher engages in ongoing assessment and observation of children's literacy development that informs their education (Essential 9). The teacher is attentive to goal setting and other approaches to foster children's literacy motivation and engagement (Essential 1). In addition, during this lesson, the teacher engages students in a read-aloud (Essential 2), and the teacher provides small-group and individual literacy instruction (Essential 3). It is also clear that there are abundant reading opportunities for children in the classroom (Essential 8).

Grade 2 Formative Assessment Process Vignette

KEYE= Essential B= bullet list item

Fame Components & Elements	Narrative	Literacy Essentials Practices
	It is mid-January and the second-grade team in Mr. Ahmed's school is teaching a reading unit that makes use of book clubs.	
1.1	Planning	
3.2 4.1	Along with the posted learning target from the lesson, Mr. Ahmed also considers the foundational reading skills his second-grade students are acquiring and how he can support these on a minute-to-minute instructional basis. While these skills may not live in the posted learning target, Mr. Ahmed is constantly observing and eliciting evidence of these skills in his data binder and in the students' reading-goal bookmarks. Additionally, Mr. Ahmed offers in-the-minute actionable feedback for his students in the teaching and learning cycle	E3; B2 E9; B4

Fame Components & Elements	Narrative			Literacy Essentials Practices
1.1	As Mr. Ahmed plans his upcoming lesson and considers his students' needs, he makes decisions for both his direct instruction and small-group book clubs. The main comprehension focus in this lesson is for all students to use their growing knowledge of how characters act and how these actions influence the plot of the story. A common text has served as the model for his direct instruction time. This lesson has three main segments: • Whole group instruction with a common class text • Independent reading and work time using book club books matched to students' reading skill and interest • Small-group time with book club peers			
1.1			gathered and are seated close to their their book club text and a pencil.	
3.1 3.2 3.3 3.4	Using the whole group common text, visible to all students, Mr. Ahmed activates prior knowledge by reviewing what students already know about the main character. This allows him to briefly revisit and assess former learning targets. Following his read-aloud of the text, Mr. Ahmed uses questioning strategies to encourage students to explain their thinking and to reinforce student-self directedness. As Mr. Ahmed listens in on partner conversations, he is able to gather evidence of students' understanding of the previous learning progressions .			E2; B1
	Learning Progression	ıs		
2.1	Building Block	Learning Target	Success Criteria	
2.2	Last Week Readers think about how a series flows; seeing patterns and predicting what will happen.	By reading and studying patterns, I can explain how these books fit together in a series.	With my book club, I can share at least 3 ideas from my jot notes to help explain how these books are similar.	
	Today's Lesson Readers expand their ideas and understanding of their main character in a series.	Use evidence from the text to prove what I know about my character. Then, share and learn more about this with my book club group.	I can use sticky notes labeled "Always," "Sometimes," and "Oh my goodness! (OMG)" to show my understanding of my character 3 or more items for Always 2 or more items for Sometimes 1 or more item for OMG	

Fame Components & Elements	Narrative	Literacy Essentials Practices
3.3 3.4	"What are three things you know about the main character, and what is your evidence from our text?" Mr. Ahmed listens in to partner responses and then shares a few themes with the whole group.	E9; B1 E2; B4
	"Sara and Cassie realized" "Emma and Sam thought about" "A question I heard a few of you asking"	EZ, 04
2.1	Mr. Ahmed then introduces today's learning target. "Today, in your individual reading and then later in your book clubs, the focus will be on what you already know about your characters, and on showing your evidence from the text for that knowing."	E1; B5
	The target is posted on the screen. Mr. Ahmed reads the target out loud to the students.	
	"Target: Use evidence from the text to prove what I know about my character. Then, share and learn more about this with my book club group.	
	Success Criteria: I can use sticky notes labeled "Always," "Sometimes," and "Oh my goodness! (OMG)" to show my understanding of my character	
	3 or more items for Always 2 or more items for Sometimes	
	1 or more item for OMG	

Fame Components & Elements	Narrative	Literacy Essentials Practices
2.3	Mr. Ahmed uses the class common text to model a "think-aloud" of what this looks like. He ensures all students can see and read his sticky notes by placing them on the document camera.	E2; B2
2.1	Included in his think-aloud is the "why": "Why is it important for readers to know about characters?" Mr. Ahmed also reinforces what proficient achievement looks like by listing three items on the Always sticky note, two items on the Sometimes sticky note, and one item on the OMG sticky note. To engage participation and practice during this model, he has students turn and talk about items that might go on the sticky notes. He uses some of their ideas for his own models.	
3.3	"Please think to yourself about your own character. What do you already know that you want to add to the Always sticky note?" 30-second pause. "Now, please turn to your book club friends and each share just one item you'll add to your Always sticky note."	E2; B5
4.1	Mr. Ahmed again shares the whole group target and checks for clarity.	
4.2	"Please turn and talk to your partner about what you understand about the target and also what questions you or your partner might have about the target."	
	Mr. Ahmed listens in to the student talk and jots a few notes to address with the whole class. He then briefly offers feedback to clarify the target. Mr. Ahmed also takes a few notes about which students he'll want to check in with first, based on their confusion/understanding.	
	Example: "Emma and her partner want to review what 'evidence from the text' means, while a few other partnerships are curious about what they will do with their three sticky notes."	
5.1	Mr. Ahmed adjusts his teaching to provide support about the sticky notes.	
	With the whole class, he reviews the success criteria regarding how the sticky notes will be composed and organized while addressing the needs of students needing a bit more support.	
2.1	Again, communication and interaction with the learning target continues.	
4.2	"Please check in with your book club group and answer this question: 'How will we know we have met the learning target?'"	

Fame Components & Elements	Narrative	Literacy Essentials Practices				
	"As you ad					
3.2 4.3	The student and add too group. These their under elements of Ayesha's Re	E3; B4				
	Date	My Reading Goal	Self-Assessment Reflection	Book Club Target	Self-Assessment Reflection	
	1-15-19	Notice linking words and add them to my word list	Copied the words also and together and used them in my story	Use evidence from the text to prove what I know about my character. Then, share and learn more about this with my book club group		
5.1 3.2	Mr. Ahmed gathering wanting to a few minu please reco	E3; B4				

Fame Components & Elements	Narrative	Literacy Essentials Practices
	Individual Learning Time	
3.2	Students read for fifteen minutes and then work on the learning target as they jot	E1; B2
	on their sticky notes. Students are grouped near their book clubs during this time.	E1; B3
	Mr. Ahmed confers with students on both book club and individual reading goals. He names what he notices the students doing, asks questions, and gives actionable	E2; B4
4.1	feedback. He ensures students use the feedback to adjust their own learning . He also encourages collaboration amongst students:	E3; B2
5.2	How did you know to do that?	E9; B2
	 Look, you've used a word-wall word. Where might you look to make sure it's spelled correctly? Oh, please check in with Brian. He had the same question. So, next time, you can try How will you know you've? 	
	Small-Group Book Clubs	E1; B3
	Mr. Ahmed invites students to meet with book clubs and share what they are	E3, B2
4.2	learning. The sticky notes are used to help focus their conversations. Students know they are to "read their evidence from the text" out loud during their book	E3; B3
	club time. This helps to practice reading fluency. In previous lessons, students have learned how to have substantive conversations and offer peer feedback . Sentence	E3; B4
	and question stems and samples are posted in all the book club meeting areas.	E9; B2
	How did you figure?Thank you for sharing	
	Could you please say more?	
4.1	As you think about today's targetHere is another idea	
	Mr. Ahmed visits the small groups, listening in and offering instruction and feedback as needed.	
1.1	During the initial planning for this unit, Mr. Ahmed organized the small-group book	E1; B1
	clubs based on students' interest and instructional needs.	E1; B2
	Malcolm's group of four includes more advanced readers. Each student is reading a different book from the same, advanced series.	E8; B2
	Emma's three group members have copies of the same book. It is from the same series as the common class text. The students in this group have a specific goal of noticing and recording linking words.	
	Cassie's three group members have copies of the same book. It is also from the same series as the common class text. Two of the group members are Spanish speakers.	

Fame Components & Elements	Narrative	Literacy Essentials Practices						
3.2	Mr. Ahmed	E1; B3						
4.3	circle and, \	E9; B4						
4.2	1. W							
	2. W							
	3. W							
	Again, this understand offer peer							
3.2	Mr. Ahmed listening in has offered students kr evidenced	E9; B1						
	Emma's Rea							
	Date	My Reading Goal	Self-Assessment Reflection	Book Club Target	Self-Assessment Reflection			
	1-15-19	Notice linking words and add them to my word list	Copied the words also and together and used them in my story	the text to prove what I know about my page character. Then, share on the		that Clara d, and I read where she sits ground and used all my		
5.1 1.1	As Mr. Ahn to adjust h							
1.1	In this daily, minute-to-minute formative assessment process, Mr. Ahmed is continually making changes to instruction in order to support his students' needs.							

Notes

Early Literacy Assessment Systems that Support Learning