Strengthening and Streamlining Local Comprehensive Assessment Systems: Guidelines and Support for Leadership Teams

Authors: Lori Dolezal, Megan Grube, Tracy Bettale Watterson

Contributors: Jenn Barone, Richard Boltax, Jess DeCarolis, Karin Edwards, Mary Fitzgerald, Pat Fitzsimmons, Gail Hall, Juliet Halladay, Lindy Johnson, Hilary Kissel, Katie McCarthy, Mary Perkins, Kathy Renfrew, Kathy Schaw, Ellen Thompson, Emily Titterton, Theresa Young

VERMONT

AGENCY OF EDUCATION

Table of Contents

Strengthening and Streamlining Local
Comprehensive Assessment Systems
(Revised: February 5, 2017)

Page **2** of **27**

VERMONT
AGENCY OF EDUCATION

Purpose of the Document

The purpose of this document is to provide guidelines and resources to help educators in supervisory unions and districts develop a streamlined, balanced local comprehensive assessment system for all students. The system should contain information that describes the types and purposes of assessments, the methods of ensuring validity, reliability and fidelity, the decision-making processes used to inform instruction, and the logistics of using assessments effectively.

This document was produced following a thorough review of literature and current practices in the field of student assessment by Agency of Education staff as well as educators in the field. It provides a synthesis of research and current policy, including the MTSS field guide and Education Quality Standards. Additional sources are cited and listed in the reference section. Links to external sources are also included.

Policy Connections

State and federal policies, such as the ones below, require a local comprehensive assessment system to support their implementation and efficacy.

Vermont Education Quality Standards (EQS)
Proficiency-Based Learning (PBL) (2120.8. Local Graduation Requirements)
Student achievement begins with highly effective instruction, informed by research and guided by state standards and proficiency-based requirements. Vermont's EQS require that locally-adopted school graduation requirements be rooted in demonstrations of student proficiency. A comprehensive and balanced assessment system should align with Vermont's proficiency-based learning policy (see Proficiency-Based Graduation Requirements) and the standards adopted by the State Board of Education.

Local Comprehensive Assessment Systems (2123 State and Local Comprehensive Assessment System)
Supervisory unions must implement a local comprehensive assessment system, including a balance of assessment types that provide information about student proficiency in a variety of ways; clearly communicated performance criteria; and methods for communicating student progress.

Multi-Tiered System of Supports (MTSS) (2121.5. Tiered System of Support)
A Multi-tiered System of Supports framework for instruction and intervention is comprehensive, systematic, and relies on a variety of assessment data to maximize the opportunities for all students to learn within any content area. In order to improve academic outcomes, schools also pair behavioral data with academic data as part of their local comprehensive assessment system.

Act 77: Flexible Pathways (2120.2. Flexible Pathways)
The Flexible Pathways Initiative (T.16, sec. 941), otherwise known as Act 77, was created to require school districts to develop and expand high quality educational experiences that acknowledge individual goals, learning styles, abilities and the pathways by which a student

can meet locally defined Proficiency-Based Graduation Requirements (PBGRs). This initiative describes the transition to personalized instruction and learning, a move documented in the expectation that all Vermont students in grades 7 through 12 participate in "an ongoing personalized learning planning process." A local comprehensive assessment system that is balanced and robust is the primary vehicle by which we can know and understand our students – a fundamental requirement in personalizing learning – and develop Personalized Learning Plans.

Consolidated Federal Programs

All schools receiving federal Title funds are required to use a comprehensive needs assessment to determine how they will use their funds and evaluate the impact of their investments on student learning and closing achievement gaps.

Strengthening and Streamlining a Local Comprehensive Balanced Assessment System: Essential Topics for Consideration

Components of a Local Comprehensive Balanced Assessment System

A local comprehensive assessment system includes multiple measures, for various purposes, and the data should be used to make instructional and programmatic decisions. These measures include screening, diagnostic, formative, progress monitoring, and summative assessments. As stated in the Vermont Multi-tiered System of Supports Response to Intervention and Instruction Field Guide (2014).

> *A successful comprehensive assessment system should invite action. The system must include not only what assessments will be used but also address where the data will be collected, who will review the results of assessment(s), when the data will be considered, and how they will be used (p. 35).*

> *School districts and/or supervisory unions can use the concept of a balanced assessment system to identify and organize the assessments they use by purpose…If the system appears out of balance, adjustments should be made. The specific tools and processes may differ across districts and schools, but a trustworthy system is comprehensive enough to address all purposes and to capture the full range of critical components within the academic or behavioral domain (p.27).*

Assessments within the system should be of high quality and used to inform instruction, and if not, students should not spend time taking it (CCSSO, 2015). A high-quality assessment system should include assessments of higher-order cognitive skills, and critical abilities, such as communication, collaboration, modeling, problem solving, reflection, and research; these assessments should be valid, reliable, fair, instructionally sensitive (i.e., be representative of content and concepts taught from curriculum and instruction), and have value for informing teaching (Darling-Hammond et al., 2013). The Criteria for Procuring and Evaluating High-Quality Assessments, published by the Council of Chief State School Officers, offers additional guidance for establishing or strengthening high-quality systems of assessment.

Different types of assessments can be used for more than one purpose and, generally, no one piece of assessment information can fulfill all purposes. Assessments should be inclusive of all students to the extent possible, which means that some students may require accommodations to access the assessment.

Types of Assessments

Performance-Based Assessment: Performance-based assessments allow for comprehensive demonstrations of learning for multiple standards, including the transferable skills. As the American Educational Research Association (1999) describes, students complete a process or produce a product in a context that *closely resembles real-life situations* (as cited in Reynolds, Livingston, & Willson, 2009). These assessments can take the form of constructed response items, on-demand tasks, curriculum embedded tasks, or complex projects. Rich performance assessments require application in authentic contexts, are rich learning experiences, and may focus on process and/or product.

Curriculum Embedded Formative Performance Assessments: These types of authentic learning and assessment opportunities are intended to be connected to standards and inform both educators and students about next steps in the learning process. They allow students to practice demonstration of essential skills and practices and to receive formative feedback from peers and/or instructors. These experiences can help prepare students to effectively transfer their learning and demonstrate mastery of skills and practices on summative performance assessments. Interdisciplinary professional learning and planning materials are available in the *Curriculum Embedded Performance Assessments* section of the Agency of Education ELA Resources web page. Additional information for building effective performance assessments is available from the Center for Collaborative Education.

Curriculum Embedded Summative Performance Assessments: These authentic assessment opportunities allow students to demonstrate skills and knowledge that may be difficult to capture on traditional forms of assessment, such as selected response formats. The data is usually used to determine student outcomes for a unit or course of study. Students may be asked to integrate a variety of transferable skills to demonstrate learning. For example, a student may construct a video in which he/she explains and demonstrates a particular concept or process, based on his/her previous research and investigation. This example incorporates a variety of transferable skills and standards.

REMINDER

The type and purpose of the assessment dictates the frequency of its administration. "Done well and thoughtfully, assessments are tools for learning and promoting equity...Done poorly, in excess, or without clear purpose, they take valuable time away from teaching and learning…" (USDOE, 2015). Additional guidance for strengthening and streamlining a local assessment system is found in the USDOE Principles for Fewer and Smarter Assessments and in the Achieve Student Assessment Inventory.

Categories of Assessments

In the following pages we explore various types of commonly used assessments, including: universal screening/benchmarks, formative, summative, diagnostic, progress monitoring, performance based, and early warning indicators.

Universal Screening/Interim Benchmarking

Universal assessments are used with **all students.** They can inform teachers about where students are performing relative to grade-level standards, which students need intervention, and how to adjust instruction and curriculum to affect students' success

Table 1: Universal Screening/Interim Benchmarking	
Purpose	● To identify students or groups of students who require diagnostic assessment ● To determine the efficacy of Tier 1/Universal Instruction
Frequency	● 3 times a year → Fall, Winter, Spring* ● When a new student enrolls in school outside of a screening window
What types of questions can this data answer?	● Who is performing above, at, or below grade level? ● Are at least 80% (or other local target) of my students demonstrating proficiency? ● What content should I emphasize in core instruction? ● What are areas of strengths for this cohort of students?
Tips & Considerations	● Identify benchmark scores for proficiency at the beginning, middle and end of the school year (some screeners provide national norms and benchmark scores) in advance of administering, share those targets broadly and provide usable reports for teachers and principals to track results. ● Screening will identify students with advanced or lagging skills so schools must be ready to support those students. ● It is essential that educators differentiate universal instruction; additionally, they must provide targeted or intensive supports.
Some Examples for Academic	● Smarter Balanced Digital Library ● i-Ready (ELA/Math) ● Ages and Stages (Kindergarten) ● Dynamic Indicators of Basic Early Literacy Skills (DIBELS-Literacy) ● AIMS Web ● Social, Academic, and Emotional Behavior Risk Screener (SAEBRS)
Some Examples for Behavior	● Universal Screening for Behavior (PBIS) ● Behavior Assessment System for Children (BASC™-3) Behavioral and Emotional Screening System ● Systematic Screener for Behavior Disorders ● Behavioral and Emotional Screening System ● Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) ● Early Warning Indicators for Secondary Students**

When a robust, comprehensive and balanced system is in place, and when all assessment administrators have the necessary assessment and data literacy competencies, scheduling may vary, as determined by students' needs and progress.

**Early Warning Indicators for Secondary Students: Research indicates that at the secondary level, Early Warning Indicators can help inform the selection of students in need of intervention and can be used proactively as a screening data source. According to Dr. Rebecca Sarlo's article on _Early Warning Systems_ (Allensworth, 2005), as a proactive measure, she reports that "By 9th grade, 85% of eventual dropouts can be accurately identified using readily available student data such as absenteeism, course failures, credits earned, and grade point average" (as cited in Sarlo, n.d.).*

Formative Assessment

Formative assessments provide information to both educators and students about what has been learned, which objectives have been addressed, and what techniques have been successful.

Table 2: Formative (Assessment *for* Learning)	
Purpose	• To inform the educator about possible facilitators and barriers to instruction for students • To identify appropriate focus for instruction • To inform future lessons and units • To determine the efficacy of Universal/Tier 1 Instruction
Frequency	• Several times throughout a learning opportunity
What types of questions can this data answer?	• Which students or groups of students need re-teaching and reinforcing of concepts? • Which students or groups of students need acceleration/enrichment? • In which areas are students proficient? Where are my students proficient? • What preconceptions/misconceptions do my students have about the content/concept? • How did my instruction impact student learning? What could be improved? • Where are my students on a continuum of proficiency?
Tips & Considerations	• Teachers and students need to share a common, clear and accurate understanding of learning targets. • Formative assessments are targeted at specific content and can be personalized for individual students. • Formative assessments occur as part of the learning process so adjustments can be made immediately or in the near future. • Formative assessments can take many forms such as verbal, written, virtual, or technology-based.
Some Examples for Academic	• Smarter Balanced Digital Library • On-going Assessment Project (Math) • Observations/Discourse • Interviews • Artifacts of Learning (hinge questions, exit tickets, response logs, ABC's of a concept, representations, models, etc.) • Interim Unit Check-Ins • Performance Tasks • NGSS / CCSS-M Sample Classroom Assessment Tasks
Some Examples for Behavior	• Teacher Acknowledgement of Demonstrating Behavioral Expectations • Teacher Observation • Office Discipline Data Focus Chart (PBIS) • Record Review • Recording Behaviors • Checklists • Student-to-Student Discourse • Teacher-to-Student Discourse

Summative Assessments

Summative assessments confirm what students can know and do, typically at the end of a year, semester, course, or instructional unit.

Table 3: Summative (Assessment _of_ Learning)	
Purpose	- To verify learning and confirm what students know, understand, and can do relative to state standards - To determine the efficacy of Universal/Tier1 Instruction - To develop and monitor student progress on a personalized learning plan to meet proficiencies
Frequency	End of unit, course, semester, year
What types of questions can this data answer?	- In which standards do students demonstrate proficiency? - Are at least 80% (or other local benchmark) of my students demonstrating proficiency? - Which students or groups of students need re-teaching and reinforcing of concepts? - Which students or groups of students need acceleration/enrichment? - Are there students who were not captured by Universal Screening that are in need of intervention or enrichment?
Tips & Considerations	- Strategies should be developed to extract the relevant information from the assessment. - Data from summative assessments should be used to inform programmatic decisions. - Assessments need to adhere to principles of validity and reliability.
Some Examples for Academic	- End of Unit/Year Tests - Smarter Balanced Assessment - New England Common Assessment Program (NECAP) Science
Some Examples for Behavior	- Minor/Major Discipline Referrals School-Wide Information System (SWIS) - Youth Risk Behavior Survey

Diagnostic & Progress Monitoring: Assessment Data for and from Targeted and Intensive Instruction

Diagnostic and progress monitoring assessments are conducted with only some students but are often necessary to plan instruction and/or intervention to meet the needs of students who need additional supports or new challenges. This may mean more formal and frequent assessment for some students in order to improve outcomes.

Table 4: Diagnostic	
Purpose	- To investigate and analyze a student's strengths and weaknesses - To identify appropriate focus for intervention - To explore the foundational skills/concepts of a content domain more comprehensively (math, behavior, literacy)
Frequency	One time per data cycle or as needed to determine student's academic/ behavioral need for intervention
What types of questions can this data answer?	- What foundational skill/concept is preventing the student from being successful in grade level content/classroom? - What prior knowledge does the student bring to the task/assessment?
Tips & Considerations	- Create a common understanding of foundational concepts and skills. - Use valid, research-based, norm-referenced tools to diagnose student need. - Calibrate the administration and scoring of diagnostic assessments. - Only use diagnostics with students demonstrating a need for closer examination.
Some Examples for Academic	- Key-Math 3 (Math) - Assessment of Lagging Skills (Academic) and Unsolved Problems
Some Examples for Behavior	- Functional Behavior Assessment - Achenbach System of Empirically Based Assessment - Assessment of Lagging Skills and Unsolved Problems (Behavior)

Table 5: Progress Monitoring	
Purpose	- To measure students' progress on a skill/concept in a predefined period of time - To determine if students have made sufficient progress to discontinue intervention - To determine if a student is in need of a different or more intensified intervention
Frequency	- One time every 2 weeks (Targeted/Tier 2) - One time per week (Intensive/Tier 3)
What types of questions can this data answer?	- What patterns do you notice in the data? - What are the obstacles to student learning? - Does this intervention meet the needs of this student? - What other strategies can be used to support the learner? - Is the student learning the skills/concepts needed to meet learning targets independently? - Does the student need more intervention?
Tips & Considerations	- Since diagnostic assessment is time intensive, students should be matched to targeted interventions using existing data sources. - Students should receive individualized diagnostic assessment to determine if more intensive supports are necessary. - If possible, use valid, research-based, norm-referenced tools to monitor progress; track student progress over time and across different contexts; and set a growth goal for a student based on their current level of performance. - The frequency of progress monitoring is determined by the student's needs and progress which means that monitoring could occur multiple times throughout the day, daily, weekly, or monthly.
Some Examples for Academic	- Dynamic Indicators of Basic Early Literacy (DIBELS-Literacy) - Pre/post tests - Northwest Evaluation Association (NWEA) Maps (ELA/Math)
Some Examples for Behavior	- School-wide Information Systems (SWIS Check-in/Check-out) (Behavior) - SWIS Suite including universal, targeted and intensive behavior data tracking

Reliability, Validity and Fidelity

Equitable methods of assessing require appropriate design, administration and interpretation. Whether designing or selecting assessments, reliability, validity, and fidelity are essential factors for making sound decisions. Additional considerations when selecting and using assessments can be found on the RTI Action Network website.

Reliability
Reliability refers to the consistency of measurements when assessments are repeated-- the degree to which results are free from measurement errors (AERA, APA, and NCME, 1999). Results or scores reported for individuals or schools must be accurate to support each intended interpretation (AERA, 2000). Reliability, in conjunction with validity and fidelity, increase confidence that the assessment will provide scores that consistently and accurately determine students' academic and behavioral abilities. Commercially available assessments will detail their reliability in their technical reports. Large scale and commercial assessments will also provide data regarding reliability.

Validity
Validity refers to the accuracy of the interpretation of assessment results (Reynolds, Livingston, and Willson, 2009). In other words, it is the degree to which evidence supports the interpretation of results, entailed by proposed uses of tests (AERA et al., 1999). Therefore, the validation process involves collecting evidence to support the proposed interpretation of results or test scores (Lane, 1999). For example, if assessment results are interpreted as determining reading comprehension, the assessment must actually measure degrees or levels of reading comprehension. Tests valid for one use may be invalid for another (AERA, 2000). Again, commercially available assessments will provide evidence of validity studies in their technical reports. Additional information about reliability and validity can be accessed on page 13 of the document Criteria for High-Quality Assessments.

Fidelity
Fidelity refers to the degree to which a practitioner follows prescription or protocol when delivering an assessment, program, or intervention (e.g., Sanetti & Kratochwill, 2009; Mowbray, Holter, Teague, and Bybee, 2003). Fidelity of administration ensures that assessments produce valid and reliable results. For further reading about the importance of fidelity, please access the National Center on Response to Intervention Brief #4: Ensuring Fidelity of Assessment and Data Entry Procedures, and Ikeda, Neessen, and Witt (2007) Best Practices in Universal Screening.

Data and Decision-Making
A strong, local comprehensive assessment system provides a rich collection of data. Educators and administrators must be prepared to collaboratively analyze, interpret, and make instructional and programmatic decisions based on these data. Problem-solving protocols and data-based decision making provide the ability to narrow broad statements about student performance to precise problem statements, making it more likely that continuous improvement plans are matched to need and tied to student outcomes. Data literacy, therefore,

is an essential competency for all decision-makers in this process. As stated by the Vermont Statewide Steering Committee on Response to Instruction and Intervention (2014):

> "*Assessment information alone is useless; it should invite action. Educators must be wise consumers of assessment data, understanding its properties and appropriate uses. The data and information provided by assessment must be examined, discussed, reflected upon, and used to make decisions…*"(p. 29).

Different data sources inform different decisions. Refer to the tables in the previous section for sample questions that can be answered by different types of data. Articulating the decision-making process will help to determine what professional learning will be needed in order for educators to engage in thoughtful data use throughout the tiers of instruction. Some brief, but informative articles on accurate decision making for universal instruction, targeted instruction and intensive instruction can be found at the RTI Action Network website. Additionally, Team-Initiated Problem Solving (TIPS) is a tool that can be used to structure the use of data efficiently by using a team approach to data analysis. Additionally, Harvard offers a self-paced course, Introduction to Data Wise: A Collaborative Process to Improve Learning & Teaching, which describes a process for analyzing and interpreting student data.

Systems and Structures

In order for a local comprehensive assessment system to effectively impact instruction and student outcomes, organizational systems, structures and culture must be built to support data collection, analysis and decision making.

School systems need to be places where ongoing learning happens at all levels. This requires administrators, teachers, and students to actively pursue deep learning through collaborative work and authentic learning opportunities. These systems include time during the school day to construct, analyze, and discuss evidence of learning. Clear, transparent learning expectations emphasize application and creation of knowledge, along with the development of important skills and dispositions Flexible pathways within the system can allow learners to advance upon demonstrating proficiency, so that learning is the constant and time is the variable. Technology serves as a platform for communicating virtually, organizing resources, housing personalized learning plans, and disseminating information. The goal is a system that fosters continuous improvement for each and every learner over time.

Additionally, some systems and structures to consider when implementing a local comprehensive assessment system are assessment calendars; data-based decision-making teams and articulated teaming structures; procedures for ensuring fidelity of assessment administration; and a data collection system which allows for the collection, manipulation and analyzation of various sources of data.

Taking Inventory and Conducting Analyses in Current Systems

Achieve's Student Assessment Inventory for School Districts may serve as a process for auditing your current system. This tool supports a process by which districts evaluate the assessments students are taking and determine the minimum testing necessary to serve essential diagnostic, instructional and accountability purposes. The goal is to ensure that every

supervisory union/district-mandated assessment is of high quality. Assessments need to provide the information necessary for specific school and supervisory union/district purposes as well as supported by structures and routines that data are actually used and action steps taken to help students.

Appendices
Appendix A: Resources
Assessment Resources
- [Assessment Evaluation Tool (AET)](#)
- [ELA/Literacy Resources Page](#)
- [EQuIP Science Rubric Facilitator's Guide](#)
- [Health Education Assessment Project](#)
- [New York City Department of Education Academics](#)
- [Professional Learning Materials for ELA Assessment Literacy and Design](#)
- [Schoolwide Integrated Framework for Innovation (SWIFT) Playbook](#)
- [Smarter Balanced Assessment Portal](#)
- [Smarter Balanced Claims and Assessment Targets (ELA)](#)
- [Smarter Balanced Claims and Assessment Targets (Mathematics)](#)
- [Teacher's College Reading and Writing Project Samples](#)

Education Quality Standards Explanation of Local Comprehensive Assessment System:

2123.2 Development and Implementation of Local Comprehensive Assessment System

Each supervisory union shall develop, and each school shall implement, a local comprehensive assessment system that:

a. assesses the standards approved by the State Board of Education;
b. employs a balance of assessment types, including but not limited to, teacher-or student-designed assessments, portfolios, performances, exhibitions and projects;
c. includes both formative and summative assessments;
d. enables decisions to be made about student progression and graduation, including measuring proficiency-based learning;
e. informs the development of Personalized Learning Plans and student support;
f. provides data that informs decisions regarding instruction, professional learning, and educational resources and curriculum; and
g. reflects strategies and goals outlined in the district's Continuous Improvement Plan.

The performance criteria of the assessment system shall be clear and be communicated to teachers, administrators, students, parents and other community members. Students and parents shall be informed at least annually regarding progress toward achieving the standards. This includes providing information in students' native languages or otherwise accessible formats.

Implementation and support by the Agency will be determined by the Secretary.

Information on Specific Screening Tools for Behavior
- [Systematic Screening for Behavior Disorders](#)
- BASC™-2 Behavioral and Emotional Screening System (BASC-2 BESS) May be a good fit if you have [AIMSWEB](#)
- [Student Risk Screening Scale](#)
- [Strengths and Difficulties Questionnaire](#) and [online scoring site](#)

Professional Learning Resources
- Implementer Series Screening Training Manual from Center on Response to Intervention
- Professional learning processes for data-based collaborative inquiry

Suggested Readings
- Beyond Basic Skills: The Role of Performance Assessment in Achieving 21st Century Standards of Learning, Darling-Hammond, L. and Adamson, F. (2014). *Beyond the Bubble Test: How Performance Assessments Support 21st Century Learning*, pp. 37-40. Jossey-Bass.
- Creating Systems of Assessment for Deeper Learning
- Developing Assessments for the Next Generation Science Standards (BOTA, BOSE of National Academies of Science 2014)
- Framework for Vermont's Early Childhood (Birth-3rd Grade) Comprehensive Assessment System
- Mathematics Assessment Project
- NGSS Appendix D: All Standards All Students
- NGSS Sample Assessment Tasks (The items in this publication integrate CCSS and NGSS content and practices in classroom tasks.)
- Ongoing Assessment Project
- Laud, Leslie. (2011). *Using Formative Assessment to Differentiate Mathematics Instruction: Seven Practices to Maximize Learning, Grades 4-10*. Corwin and NCTM.
- Universal Screening within an RTI Model

Appendix B: Assessment Summaries, Strengths, & Limitations

The charts below offer guidelines and sources for evaluating the intended purpose of specific assessments. Assessments and interventions that were selected have been evaluated by research institutions (e.g., IES, SEDL, etc.) These individual assessments should be used in conjunction with several other sources of evidence, as instructional decisions are best based on multiple sources of evidence. This list includes only stand-alone assessments that are not part of a larger program. Although some programs contain built-in assessments, they are not included in this document. This is not intended to be an all-inclusive list of assessments administered in a supervisory union's local comprehensive assessment system. Further, this list is ***not a recommended or endorsed list***, rather, it is a list of commonly used assessments in schools. When possible, links to the full databases are included. If there is an assessment that you would like content specialists to review, please contact the appropriate content person at the Agency.

For details about assessments in early education, please see the companion document: Framework for Vermont's Early Childhood (Birth-3rd Grade) Comprehensive Assessment System. *Vermont-Based Assessment*

Math Assessments
Sources used for Independent Research and Reviews for Assessments and Intervention Tools
- Center on Response Intervention
- National Center on Intensive Intervention

Title	Purpose	Summary	Strengths/Limitations
Primary Number and Operations Assessment (PNOA) *Vermont-Based Assessment*	• screening • formative • summative • multi-year summative	The PNOA (K-2) was created as a resource for schools, districts and supervisory unions and not as a state mandated assessment. Vermont educational organizations can choose from various implementation models for use of these tools or may choose to continue with their current local assessment techniques. If districts choose to implement the PNOA, training and professional development workshops will be offered throughout the state. It is imperative that specially designed professional learning and training precede the administration of this assessment by any professional.	**Strengths:** • one-on-one: 20-40 minutes to administer • considered CCSSM recommendations in 2014 revisions • scoring protocols included • procedural fluency • computation **Limitations:** • not norm-referenced • mathematical practices are not assessed

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
AIMSweb-Concepts and Application (M-CAP)	• screening • progress monitoring	M-CAP is a test of 8-10 minutes that assesses the general mathematics problem-solving skills expected in Grades 2-8. The mathematics domains assessed include number sense, operations, patterns and relationships, data and probability, measurement, data and statistics, geometry, and algebra. **Research:** AIMSweb National Norms Technical Documentation Center on Response to Intervention	**Strengths:** • the test may be administered in a large or small group setting or to individual students • based on the National Council of Teachers of Mathematics (NCTM) Principles and Standards • contains 33 probes per grade: three for benchmarking all students, and an additional 30 for monitoring the effectiveness of interventions **Limitation:** • not aligned to the Common Core State Standards
Connected Mathematics Project (CMP)	• progress monitoring	CMP is designed to collect data concerning the following three dimensions of student learning: content knowledge, mathematical disposition, and work habits. **Research:** Research Evaluation	**Strengths:** • reflects recommendations of the CCSSM • provides a variety of tools for student assessment: checkpoints, surveys of knowledge, and observations **Limitation:** • available for grades 6-8 only

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
KeyMath-3	• diagnostic	The KeyMath 3 assessment and instruction system gives you tools to assess and improve math skills of students ages 4½ - 21. This is a comprehensive system consisting of *three* linked components: the updated KeyMath 3 Diagnostic Assessment, ASSIST™ Scoring and Reporting Software, and the new KeyMath 3 Essential Resources instructional program, provides two levels of instruction: Level 1 (K- grade 2) and Level II (grades 3-5/6). **Research:** Sample Information pp. 4-6 Journal of Psychoeducational Assessment, v29 n1 p94-97 Feb 2011	**Strengths:** • individually administered, paper/pencil, and takes between 30-90 minutes/K-12 student • RtI Tiers 2 and 3 • measures mathematics concepts and skills • scoring Options: ASSIST software or manual scoring • publication date is 2007 (pre-CCSSM) **Limitations:** • this is often used by special educators to determine eligibility for special education services/disability which would take it out of use for classroom teachers
Measure of Academic Progress (MAP)	• screening • progress monitoring	MAP for Primary Grades (K-12) is computer administered and adaptive. Administration time is less than 60 minutes for most students or groups. **Research:** Center on Response to Intervention Regional Education Laboratory Program NWEA	**Strengths:** • Common Core State Standards • technically enhanced item types (e.g., hot spots, drop and drag) • measures instructional readiness and student growth • compares and predicts student achievement and growth over time via research-based normative and growth information • creates and reinforces data-informed instructional practices **Limitation:** • threshold scale scores and expected percentages represent a preliminary attempt at standard setting and Smarter Balanced may revise them after the spring 2015 testing term
Number Knowledge Test (NKT)	• screening	*The NKT was designed to measure the intuitive knowledge of number that the average child has available at the age-levels of 4, 6, 8 and 10 years.* **Research:** Research Articles *How the Brain Learns Mathematics* by David Sousa	**Strengths:** • oral response from child • administered individually; 5-10 minutes/child • developmental and formative **Limitation:** • only measures number sense

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
Number Sense Screener (NSS)	• screening	*The NSS is a quick number sense screening tool that identifies children at risk for mathematics difficulties and is administered at least three times: fall of kindergarten, spring of kindergarten, and fall of first grade. This test screens for: counting skills, number recognition, number, comparisons, nonverbal calculation, story problems, and number combinations.* **Research:** *Screen Children for Math Delays Early with Number Sense Screener* *Reliability and Validity of the Number Sense Screener*	**Strengths:** • can be completed and scored by a teacher, a learning specialist, or a school psychologist • has internal consistency of .85 and predictive validity of .70 (end of third grade math achievement) **Limitation:** • only measures number sense

Strengthening and Streamlining Local Comprehensive Assessment Systems (Revised: February 28, 2017)   Page **17** of **27**

VERMONT
AGENCY OF EDUCATION

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
Ongoing Assessment Project (OGAP)	• formative	OGAP is a professional development intervention that trains teachers to use single or multiple math items of high cognitive demand to gather information on student thinking and then analyze that information using frameworks based on research on student thinking in mathematics. **Research:** Research Underlying OGAP Marge Petit	**Strengths:** ● includes a strong professional development component focused on understanding math education research and using the research and learning progressions to understand evidence in student work (including classroom discussion) and make instructional decision ● provides tools and resources (e.g., item banks and learning progressions) to help teachers gather formative evidence of student thinking and make instructional decisions ● the professional development and item bank are aligned with the CCSSM ● helps participants understand their math instructional materials/math program ● supports a school wide formative assessment system in mathematics grades 3 – 7 that involves about 80% of the mathematics at these grade levels ● provides a structure for sustaining OGAP over time **Limitation:** ● available only for fractions, multiplicative reasoning, and proportionality
PAL-II Math	• diagnostic	PAL-II Math measures the development of cognitive processes that are critical to learning math skills and actual math performance. **Research:** Journal of Psychoeducational Assessment, v28 n1 p80-86 2010 Psychology Resource Centre-Hebblab	**Strengths:** ● identify weaknesses in targeted math skills for supplementary instruction to master the skill ● grades K-6; RtI Tiers 1, 2, and 3 **Limitations:** ● publication date is 2007 (pre-CCSSM) ● Level B qualifications required

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
Renaissance STAR Math	• progress monitoring	STAR Math is a computer-adaptive assessment of general mathematics achievement for students in grades 1 to 12. STAR Math provides information on student performance for hundreds of skills within 32 domains. The difficulty of items is adjusted automatically to reflect the skill level of all students. **Research:** National Center on Intensive Intervention Center on Response to Intervention National Center on Student Progress Monitoring	**Strengths:** • can be administered to individuals or groups of students • offers several accommodations for students with disabilities through the accessibility options built into a computer's operating system **Limitation:** • *Standards List with Aligned Product Skills* (2010) still reflects Vermont GEs from 2004.Only multiple choice question type
System to Enhance Educational Performance (iSTEEP)	• screening • progress monitoring	iSTEEP provides products for assessing, analyzing, progress monitoring and next generation technology. **Research:** Using Curriculum-Based Assessment and Curriculum-Based Measurement to Guide Elementary Mathematics Instruction: Effect on Individual and Group Accountability Scores	**Strengths:** • assesses Common Core skills at each grade level • paper/pencil or tablets/iPad/computer accessible • E-learning coursework for professional learning available **Limitation:** • for school or district wide use only
Test of Early Mathematics Ability-Third Edition (TEMA3)	• diagnostic	The TEMA-3 individually administered math performance items covering six skill domains; numbering skills, number-comparison facility, numeral literacy, mastery of number facts, calculation skills, and understanding of concepts. **Research:** Journal of Psychoeducational Assessment, v24 n1 p85-98 2006 RTI in Math: Practical Guidelines for Elementary Teachers Examining the Psychometrics of Number Sense Among Kindergarten Students	**Strengths:** • it has two parallel forms each containing 72 items • standard scores, percentile ranks, and age and grade equivalents **Limitations:** • 2003 NOTE: Publication date is prior to CCSS (2010) • Level B qualifications required

English Language Arts (ELA) Assessments

Sources used for independent research and reviews for assessments and intervention tools:

- SEDL
- Center on Response to Intervention
- Collaborative Center for Literacy Development
- National Center on Intensive Intervention

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
Aimsweb	• screening • progress monitoring	**Summary:** *Aimsweb Reading* The aimsweb Reading subscription contains assessments for early literacy and reading – as well as aimsweb's robust online data management and reporting system. *Aimsweb Writing* The **aims**web Language Arts subscription includes assessments for early literacy, reading, spelling, and writing – as well as **aims**web's robust online data management and reporting system. **Skills Assessed:** **Letter Naming Fluency**: Student names printed upper and lower case letters (presented in random order) for 1 minute. **Letter Sound Fluency**: Student produces the sound for upper and lower letters presented in print (and in random order for 1 minute. **Phoneme Segmentation Fluency**: Student segments orally presented words into phonemes for 1 minute. **Nonsense Word Fluency**: Student reads non-sense words for 1 minute. See more on National Center on Intensive Intervention	**Strengths:** ● time efficient measures ● robust data reporting system ● multiple alternate forms available ● compatible with all curriculum and standards, including the Common Core State Standards ● moderate to high reliability and validity reports for most measures (exception: letter naming fluency) L**imitation:** ● comprehension is measured using Timed Oral Reading Accuracy report, unless the supplemental Maze measure is also used
Clay's Observation Survey (POA)	• progress • monitoring	**Summary:** Clay's Observation Survey was created to assess children's reading abilities for determination of Reading Recovery services. The subtests of this system were incorporated into a Vermont	**Strengths:** ● intense diagnostic measures for the foundational reading skills

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
		assessment that was created and tested in the mid-1990's to assess and monitor students in grades K-2. Benchmarks were set based on field testing and schools who use this tool use those benchmarks for monitoring progress over the three years. **Skills Assessed:** Subtests include: Letter identification, word test, word writing test, dictation (hearing and recording sounds in words), Concepts About Text. **Research:** Clay's Observation Survey	**Limitations:** ● administration and analysis requires an extended period of time ● this tool does not measure comprehension
DRA-2	● progress monitoring	**Summary:** The Developmental Reading Assessment is a criterion-referenced, formative assessment that measures reading engagement, oral reading fluency and comprehension for grades K-8. In grades K-3, students are evaluated on an oral retelling in which they recall key concepts and vocabulary, main characters, and text events. In grades 4-8 students are evaluated on their Prediction, Summary, Literal Comprehension, Interpretation, Reflection and Metacognitive Awareness responses. **Skills Assessed:** Pearson DRA-2	**Strengths:** ● the assessment has a thorough comprehension component (including metacognitive considerations), an oral reading component, a fluency component, and a reading engagement/reading behaviors component ● the assessment is vertically aligned grades K-8 ● both fiction and nonfiction texts are included in the kits-one kit is K-3 and the other is 4-8 ● The DRA2 does a thorough job of assessing comprehension. From Level A through Level 3 students engage in a picture walk before reading. At levels 4-16 there is a connection component added to the comprehension portion of the assessment. A nonfiction text choice is added beginning at level 16, and there are specific nonfiction comprehension questions relating to nonfiction text features included. **Limitations:** ● beginning at level 18, the comprehension portion of the assessment for fiction text is broken into several components: a prediction, a retelling/summary, an interpretation/inference question, and

VERMONT
AGENCY OF EDUCATION

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
			a reflection/main idea question. At level 20 there are no longer any connection questions ● starting at level 24 up through level 80 students write out their answers to the comprehension questions, but the teacher does have the opportunity to scribe for the student if necessary ● nonfiction selections are included at benchmark levels 16 (typically end of grade 1), 28 (typically end of grade 2), 38 (typically end of grade 3), and at every level beginning at level 40
Dynamic Indicators of Basic Early Literacy Skills [DIBELS](#)	● Benchmark screening ● progress monitoring	**Summary:** Information for DIBELS Next: DIBELS are a set of short, fluency measures for assessing the acquisition of early literacy skills from kindergarten through sixth grade. Measurement areas include alphabetic principle, phonological awareness, phonics, fluency, and comprehension. DIBELS Next comprises six measures: [First Sound Fluency (FSF), Letter Naming Fluency (LNF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), DIBELS Oral Reading Fluency (DORF) and a maze comprehension measure](#).	**Strengths:** ● reported reliability and validity scores are well within acceptable range **Limitation:** ● individual administration
[**Fountas and Pinnell Benchmark Assessment System**](#)	● progress monitoring	**Summary:** The Benchmark Assessment System helps determine instructional and independent reading levels and places students on the Fountas and Pinnell Text Level Gradient. **Skills Assessed:** -Reading Accuracy, Fluency and Comprehension -Writing about Reading **Research:** [Fountas and Pinnell: Benchmark Assessment System](#) [The Guided and Reading Approach](#)	**Strengths:** ● coordinating Intervention program ● at each level (A to Z) texts are analyzed using ten characteristics: (1) genre/form; (2) text structure; (3) content; (4) themes and ideas; (5) language and literary features; (6) sentence complexity; (7) vocabulary; (8) word difficulty; (9) illustrations/graphics; and (10) book and print features ● over 35,000 books are levelled on their website for a resource ● texts were demonstrated to be both reliable and valid measures for assessing students' reading levels **Limitations:** ● limited set of texts

Strengthening and Streamlining Local Comprehensive Assessment Systems (Revised: February 28, 2017)   Page **22** of **27**

VERMONT
AGENCY OF EDUCATION

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
			● limited set of skills measured without optional assessments ● limited independent research on reliability and validity
Independent Reading Assessment by Jennifer Seravallo (Non-Fiction Kit)	● progress monitoring	**Summary and Skills Assessed:** Scholastic: Independent Reading Assessment	**Strengths:** ● assessment has a thorough comprehension component with questions divided into four strands: main idea, key details, vocabulary, and text features ● kits are available for both fiction and nonfiction and multiple titles are available at every level ● texts included are trade books-not texts written for the assessment ● assessment includes a teacher planning form with scoring rubrics, an If/Then-Teaching Suggestions Guide and a Goal-Setting Form ● students complete the assessment independently, so teachers do not sit with the student for any portion of the test **Limitations:** ● kit only addresses grades 3-5 ● fiction and nonfiction kits must be purchased separately ● there is no oral reading/decoding component to the assessment
Phonological Awareness Literacy Screening (PALS)	● screening ● diagnostic	**Summary:** PALS-K is a criterion-referenced, phonological awareness literacy screening and diagnostic tool that measures developing knowledge of foundational skill, including decoding, cipher knowledge, letter knowledge, concepts about print and phonological awareness. PALS 1-3 and PALS Plus are used to screen and identify students in need of additional instruction and diagnose specific skill deficits. Cognitive elements are measured on a leveled system. Initially, students are administered a spelling inventory and *word recognition in isolation* assessment: Level A,, which measures oral reading in context (accuracy, fluency, rate and	**Strengths:** ● PALS assesses most CCSS reading standards, including diagnostics for foundational skills ● comprehension questions are text-based and include literal and inferential levels examples for each selection ● the leveled system is allows for diagnostic information to be obtained when students are not meeting benchmarks ● entry level spelling inventory measures correct spelling and phonics features, including vowel, syllable, and affix patterns ● oral reading selections are connected to student results on the entry level word recognition task

Strengthening and Streamlining Local Comprehensive Assessment Systems (Revised: February 28, 2017)   Page **23** of **27**

VERMONT AGENCY OF EDUCATION

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
		comprehension), Level B, which measures alphabetics for students who do not meet level A expectations, and Level C, which measures phonological skills in depth for those students not meeting Level B benchmarks. **Research:** Collaborative Center for Literacy Development SEDL Reading Assessment Database	● reported reliability and validity measures for the Pre-K-3rd grade packages are within expected ranges **Limitation:** ● with limited comprehension selections, accurate progress monitoring may be difficult for the oral reading in context measure
Renaissance STAR *Early Literacy Diagnostic Assessment*	● progress monitoring	**Summary:** This is a computer adaptive test, self-administered by computer (even for children who are not yet readers) by means of digitized audio directions for the test itself and for every test item. The assessment has attained recognition as a scientifically research-based progress monitoring instrument by the federally-funded National Center for Student Progress Monitoring (NCSPM). **Skills Assessed:** **General Readiness (GR)** — Student must demonstrate understanding of written word length, position words, words vs. letters, basic numeracy, word matching, word boundaries, shapes and sequences. **Graphophonemic Knowledge (GK)** — Student must demonstrate understanding of letter names and sounds, alphabetic letter sequence, and alphabetical order. **Phonemic Awareness (PA)** — Student must demonstrate understanding of rhyming words, ability to blend word parts, and phonemes, sound discrimination, oral word length, and ability to identify missing sounds. **Phonics (PH)** — Student must demonstrate understanding of long vowels, short vowels, beginning and ending consonants, consonant and	**Strengths:** ● there is convincing evidence that STAR Early Literacy is aligned with state and national reading standards ● there is convincing evidence for the reliability and validity of this assessment tool **Limitation:** ● no predictive statistics are reported ● no statistics are reported comparing results of children of different demographic characteristics

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
		vowel replacement, word families (onset and rime), consonant blends, clusters and digraphs. **Comprehension (CO)** — Student must demonstrate ability to read and derive meaning from words, sentences, and paragraphs. **Structural Analysis (SA)** — Student must demonstrate ability to find words within other words, build words and compound words. **Vocabulary (VO)** — Student must demonstrate knowledge of high frequency words, synonyms, and antonyms. **Research:** National Center on Intensive Intervention Alignment to CCSS	
Renaissance STAR *Reading Assessment*	• progress monitoring	**Summary:** The STAR Reading test yields a variety of test scores, some of which (instructional reading level - IRL) support criterion-referenced interpretations, and others of which (percentile ranks, grade equivalents, NCR scores) support norm-referenced interpretations. This assessment requires students to rely on background information, apply vocabulary knowledge, and use active strategies to construct meaning from the assessment text. The assessment is designed for repeated administration throughout the school year. All record-keeping and report preparation functions are completely automated. The assessment has attained recognition as a scientifically research-based progress monitoring instrument by the federally-funded National Center for Student Progress Monitoring (NCSPM).	**Strengths:** • yields a variety of data • automated record keeping • time efficient administration **Limitations:** • selected response format only • relies on student background information

Assessment Instrument	Stated Purpose	Summary/Skills Assessed/Research	Strengths and Limitations
		Skills Assessed: -Foundational Skills -Reading Informational Text -Reading Literature -Language **Research:** National Center on Intensive Intervention Center on Response to Intervention Southwest Educational Development Laboratory Collaborative Center for Literacy Development Alignment to CCSS Research Links	

References

Allensworth, E. (2005). *Graduation and dropout trends in Chicago: A look at cohorts of students from 1991 to 2004.* Chicago, IL: Consortium on Chicago School Research.

AERA, APA, & NCME (1999). Standards for Psychological Testing. Washington, DC: American Psychological Association.

CCSSO (2013). *Toolkit for evaluating alignment of instructional and assessment materials to the common core state standards.* Washington, DC: Author.

CCSSO (2015). *Comprehensive statewide assessment systems: A framework for the role of the state education agency in improving quality and reducing burden.* Washington, CD: Author.

Darling-Hammond, L., Herman, J., Pellegrino, J., et al. (2013). Criteria for high-quality assessment. Stanford, CA: Stanford Center for Opportunity Policy in Education.

Ikeda, M.J., & Neesen, E. (2007). Best practices in universal screening. *Best Practices in School Psychology* V.

Lane, S. (1999). *Validity Evidence for Assessment*, Reidy Interactive Lecture Series.

Linn, R. L. (1993). *Educational Assessment: Expanded Expectations and Challenges* CSE Technical Report 351, CRESST/University of Colorado at Boulder.

Mowbry, C.T., Holter, M.C., Teague, G.B., & Bybee, D. (2003). Fidelity Criteria: Development, Measurement, and Validation. *American Journal of Evaluation. 24*(3), 315-340. doi: 10.1177/109821400302400303.

Reynolds, CR., Livingston, R.G., & Willson, V. (2009). Measurement and Evaluation in Education. Upper Saddle River, NJ: Pearson.

Ryan, J. (2010). Envisioning a state educational system: Improving learning through a comprehensive assessments system. Office of Superintendent of Public Instruction. Olympia, Washington.

Sarlo, R. (n.d.) Early Warning Systems: Moving from Reaction to Prevention. RtI Action Network.

Sanetti, L. M., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to the special series. School Psychology Review, 38(4), 445–459.

U.S. Department of Education (2015). Fact Sheet: Testing Action Plan.

Vermont State Board of Education Rule 2000 (2014). *Education quality standards*. Montpelier, VT: Vermont Agency of Education.

Vermont Statewide Steering Committee on Response to Instruction and Intervention (2014). *Vermont multi-tiered system of supports response to intervention and instruction field guide.* Burlington, VT: Author.